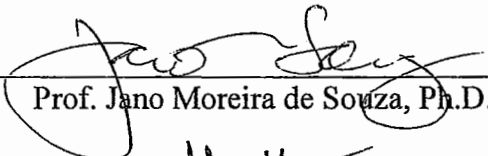


UMA ABORDAGEM PARA DISSEMINAÇÃO DE INFORMAÇÃO EM
CONTEXTOS COLABORATIVOS

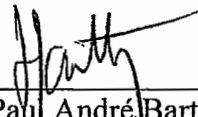
Adriana Santarosa Vivacqua

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS
EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

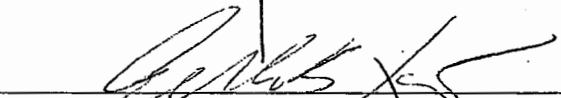
Aprovada por:



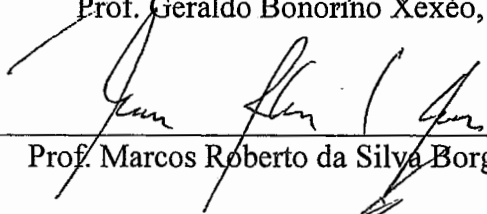
Prof. Jano Moreira de Souza, Ph.D.



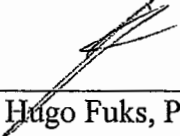
Prof. Jean-Paul André Barthès, Ph.D.



Prof. Geraldo Bonorino Xexéo, D.Sc.



Prof. Marcos Roberto da Silva Borges, Ph.D.



Prof. Hugo Fuks, Ph.D.



Prof. Patrick Brézillon, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO DE 2007

VIVACQUA, ADRIANA SANTAROSA

Uma Abordagem para Disseminação de
Informação em Contextos Colaborativos

[Rio de Janeiro] 2007

XIII, 265 p. 29,7 cm (COPPE/UFRJ, D.Sc.,
Engenharia de Sistemas e Computação, 2007)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Trabalho Cooperativo Apoiado por Computador
2. Gestão do Conhecimento
3. Sistemas de Apoio a Percepção

I. COPPE/UFRJ II. Título (série)

Be regular and orderly in your
life so that you may be violent
and original in your work

-Flaubert

AGRADECIMENTOS

Every work is undertaken with the help and support of many people, more so when it is something of the magnitude and length of a doctoral thesis.

First and foremost, I must thank my advisors Jano M de Souza and Jean-Paul Barthès, for the guidance, inspiration, patience and the vote of confidence. I am sure this is but the beginning of a long and fruitful relationship in research and academia. The many ideas we developed during these five years will not be forgotten and even the more eccentric ones (and there were many of those) will surely see the light of day in due time.

To my thesis committee, Patrick Brézillon, Hugo Fuks, Marcos Borges and Geraldo Xexéo, for coming from near and far for the defense, for taking the time to read the thesis document (probably the only ones who ever will) and for their invaluable suggestions that make this research better.

To the Navigator team, who were a part of this work: Leandro, Diogo, Marcos, João and Cadu.

To my colleagues in DocBD and other venues, for their support and fruitful discussions: Jonice, Juliana, Zeh, Barros, Beбето, Wallace, Mutaleci, Wladimir, Vaninha, Gilson, Emerson, RafaLeo. You're next!

My students, from whom I also have learned a lot: Melfry, Carla, Pinho, Leandro, Marcos.

For lack of a better category, for inspiration, those individuals with whom I have had the pleasure of working in one capacity or another and who made me want to do research (even if I disagree with them from time to time): Cris Garcia, GJ de Vreede, Claudinha, Palma, Ronnie, Henry Lieberman.

To my fun-n-games crew, for keeping me sane and providing much needed diversion from research: Nilo, Aki, Vicente, Cristina, Ronnie, Prof. X, Eddy and the rest of the AFKKK crew, LC, Monclar, Claudia, Monica et Gilson (a Compiègne alors!)

Last but not least, to my family, without whom none of this would be possible: Alba, Angelo and my (not so) little brother, Eduardo.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

UMA ABORDAGEM PARA DISSEMINAÇÃO DE INFORMAÇÃO
EM CONTEXTOS COLABORATIVOS

Adriana Santarosa Vivacqua

Setembro/2007

Orientadores: Jano Moreira de Souza

Jean-Paul André Barthès

Programa: Engenharia de Sistemas e Computação

Este trabalho apresenta um modelo e um método para distribuição de informações em ambientes colaborativos fracamente estruturados. Estes ambientes se caracterizam pela fluidez e facilidade com a qual os usuários alteram regras e procedimentos, se ajustando a eventos externos. No entanto, esta reorganização só é possível se os participantes tiverem informações que levem à adequada tomada de decisão e ajustes entre os membros. Baseados em teorias sociológicas, apresentamos uma arquitetura ponto-a-ponto baseada em agentes para distribuição de informações, e um método para analisar interações entre usuários e descobrir colaborações em andamento. Uma análise deste método mostra ser uma avenida promissora, apesar de melhorias serem necessárias para seu perfeito funcionamento.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

AN APPROACH FOR INFORMATION DISSEMINATION
IN COLLABORATIVE ENVIRONMENTS

Adriana Santarosa Vivacqua

September/2007

Advisors: Jano Moreira de Souza

Jean-Paul André Barthès

Department: Computer Science and Engineering

This thesis presents a model and a method for information distribution in loosely structured collaborative environments. These environments are characterized by their flexibility and easy with which participants alter rules and procedures, adjusting to external events. However, this reorganization is only possible if participants have information leading to adequate decision making and adjustments between group members. Based on sociological theories, we present an agent based peer-to-peer architecture for information distribution, and a method to analyze interactions between users and discover ongoing collaboration. An analysis of this method shows it is a promising avenue, even though improvements are necessary for it to function perfectly.

Resumé de Thèse présentée à COPPE/UFRJ comme réalisation partielle des exigences
pour le degré de Docteur en Sciences (D.Sc.)

UNE APPROCHE POUR LA DISSEMINATION D'INFORMATIONS
DANS ENVIRONNEMENTS COOPERATIFS

Adriana Santarosa Vivacqua

Septembre/2007

Directeurs: Jano Moreira de Souza
Jean-Paul André Barthès

Programme: Génie de Systèmes et Informatique

Cette thèse présente un modèle et une méthode pour la distribution d'informations dans environnements coopératifs avec une structure faible. Ces environnements sont fluides et ses membres d'un groupe peuvent facilement changer les règles et processus pour s'adapter à événements externes. Pour que cette réorganisation soit possible, il faut que les membres aient informations pour permettre la prise de décisions et l'ajustement entre les membres. Nous présentons une architecture basée sur des agents pour la distribution d'informations, et une méthode pour l'analyse des interactions entre utilisateurs et découvrir les interactions en cours. Une analyse de cette méthode montre qu'il s'agit d'une bonne stratégie, même s'il a besoin d'être amélioré pour fonctionner parfaitement.

ÍNDICE

1	Introdução.....	1
1.1	Base Filosófica.....	4
1.2	Problema e Objetivos.....	5
1.2.1	Aplicabilidade e Limitações.....	8
1.2.2	Abordagem Técnica.....	9
2	Fundamentação Teórica.....	12
2.1	Embasamento Sociológico e Organizacional.....	12
2.2	Redes Pessoais e Intensionais.....	15
2.3	Distancia e Colaboração.....	18
2.4	Consciência em Trabalho Colaborativo.....	20
2.5	Atenção e Foco do Usuário.....	27
2.6	Estrutura de Locais (Locales Framework).....	29
2.7	Rede Social e Análise de Interação.....	31
2.8	Busca e Recuperação de Informação.....	34
2.9	Esferas de Colaboração como Pontes entre Membros de Grupos.....	34
3	Trabalhos Correlatos.....	37
3.1	Organização do Trabalho.....	37
3.2	Sistemas de Recomendação.....	39
3.3	Sistemas de Consciência.....	40
3.4	Sistemas baseados em agentes.....	44
3.5	Arquiteturas para Disseminação de Informação.....	45
3.6	Considerações.....	46
4	Estudos de Caso.....	48
4.1	Caso 1: Caso Europeu.....	48
4.1.1	Organização da Equipe e Sistemas de Apoio.....	49
4.1.2	Interações Assíncronas: Fórum.....	51
4.1.3	Reunião Face a Face.....	54
4.1.4	Reuniões via IRC.....	55
4.1.5	Considerações.....	56
4.2	Caso 2: Caso Artigo.....	59

5	Uma Arquitetura Baseada em Agentes para Captura e Organização de Informações	66
5.1	Uma Arquitetura Multiagente para Gestão de Conhecimento.....	66
5.2	CUMBIA: uma Estrutura para Apoiar Colaboração Oportunista.....	69
5.2.1	Uma Estrutura para Apoio a Colaboração.....	69
5.2.2	Serviços de Gestão de Conhecimento	71
5.2.3	Awareness e Serviços de Matchmaking	73
5.2.4	Serviços de Colaboração	73
5.2.5	Serviços de Interface com o Usuário.....	74
5.3	Estudo Piloto: Navigator – um Sistema para Promover Encontros entre Colaboradores Potenciais	74
5.3.1	Perfis de Usuário	75
5.3.2	Awareness e Matchmaking.....	78
5.3.3	Colaboração e Interface	81
5.3.4	Detalhes técnicos	83
5.4	Lições Aprendidas com CUMBIA e Navigator.....	86
6	Uma Abordagem Computacional para a Descoberta de Esferas de Colaboração..	88
6.1	Constituindo Esferas de Colaboração	93
6.2	Interações Baseadas em E-Mail	96
6.2.1	Descoberta de Colaboradores Ativos	97
6.2.2	Determinação de Recursos e Temas.....	101
7	Análise.....	103
7.1	Análise de Interações para Determinar Esferas de Colaboração Ativas.....	103
7.1.1	Análise I: Padrões de Interação no Dataset do Projeto UE	104
7.1.2	Análise II: Padrões de Interação no Dataset de Emails	110
7.2	Análise de Conteúdo Extraído de Recursos.....	114
7.2.1	Análise III: Análise de Conteúdo em Dataset do Projeto EU	114
8	Discussão e Trabalhos Futuros	122
8.1	Limitações.....	124
8.2	Comparação com outros métodos/sistemas	125
8.3	Trabalhos futuros	127
8.3.1	A Questão da Privacidade.....	129
9	Referencias Bibliográficas.....	131
	Anexo I.....	143

Journals.....	143
Post-Proceedings	143
Conferências	144
Relatórios Técnicos	147
Anexo II: Versão da Tese em Inglês	148

ÍNDICE DE FIGURAS

Figura 1-1: Diferentes tipos de esforço (NUNAMAKER et al., 2001).....	9
Figura 2-1: Trabalhadores autônomos como uma porcentagem da mão-de-obra (de Castells, 1996, p332)	13
Figura 2-2: Funcionamento de consciência de situação (de Endsley, 2000).....	21
Figura 2-3: A consciência como o sobreposição entre foco e nimbo.....	22
Figura 2-4: Relacionamento entre consciência, atenção e ação (DAVENPORT, BECK, 2001, p. 21).....	28
Figura 2-5: Exemplo de sociograma e tabela correspondente de ligações entre elementos.....	32
Figura 4-1:Uma série de sociogramas correspondente a momentos diferentes nas atividades do grupo.....	63
Figura 5-1: Arquitetura multiagente OMAS (adaptado de TACLA, BARTHES, 2003).....	67
Figura 5-2: Arquitetura interna de agente OMAS (de TACLA, BARTHÈS, 2002).....	68
Figura 5-3: Arquitetura de Serviços CUMBIA	69
Figura 5-4: Esferas de trabalho.....	71
Figura 5-5: Perfil de Usuário usado em Navigator.....	77
Figura 5-6: Lista de contatos	81
Figura 5-7: Visão de radar	82
Figura 5-8: A visão geral da interface Navigator: janela de lista de contatos, janela de chat e radar (parcialmente mostrado)	83
Figura 5-9: Arquitetura de Agentes de Navigator e sua Comunicação	85
Figura 6-1: Esferas de Trabalho e Esfera de Colaboração	89
Figura 6-2: Ontologia de conceitos da esfera de colaboração.....	91
Figura 6-3: Ontologia de conceitos envolvidos na descoberta de esferas de colaboração	95
Figura 6-4: Busca em dois níveis para distribuição de informação.....	96
Figura 6-5: Modelo UML de interações de usuário, aplicado ao envio de mensagens de email	97
Figura 7-1: Mensagens totais por dia	105
Figura 7-2: Mensagens enviadas por B por dia	105

Figura 7-3: Mensagens trocadas entre membro B e outros membros do grupo	106
Figura 7-4 (A-F): Número de mensagens enviadas por dia de B a outros membros (A, C, D, F). O último gráfico mostra uma sobreposição dos anteriores	109
Figura 7-5: Screenshot de visualização, vários grupos estão visíveis.	112
Figura 7-6: Gráfico de interação baseado no tempo.....	113
Figura 7-7: Distribuição de semelhança para 10 faixas de valores, para mensagens e threads, como uma porcentagem do total.	116
Figura 7-8: Distribuição de avaliação de semelhança para 10 faixas de valores, para mensagens e threads (com anexos incluídos), como uma porcentagem do total.	117
Figura 7-9: Distribuição de avaliação de semelhança para 10 faixas de valores, para comparações entre recursos (arquivos) e threads (com e sem anexos), como uma porcentagem do total.	118

ÍNDICE DE TABELAS

Tabela 1-1: Matriz Espaço x Tempo (adaptado de DIX et al., 1988 e GRUDIN, 1994). Tecnologia não computacional em itálico.....	8
Tabela 2-1: Características de interações síncronas co-localizadas e seu apoio pelas tecnologias de hoje (adaptado de Olson e Olson, 2000).....	20
Tabela 2-2: Os elementos de consciência de área de trabalho e as perguntas que eles endereçam (adaptou de GUTWIN, GREENBERG, 2002).....	23
Tabela 2-3: Elementos assíncronos de consciência de mudança de acordo com diferentes perspectivas, e perguntas que devem ser consideradas (adaptado de Tam e Greenberg, 2006)	24
Tabela 4-1: Distribuição de mensagens no fórum	54
Tabela 4-2: Log parcial de mensagem para o projeto	61
Tabela 5-1: Tipos de oportunidades gerados pela combinação de dimensões distintas de perfis	80
Tabela 5-2: Semelhança de trabalho recente	81
Tabela 5-3: Semelhança de habilidade	82
Tabela 7-1: Palavras-chave para threads	115
Tabela 7-2: Resultados para um thread (10 primeiros mostrados).....	115
Tabela 8-1: Análise comparativa entre esferas de colaboração e outros sistemas	126

1 Introdução

L'adoption des technologies de communication et réseau a connecté les individus, qui utilisent des médias électroniques pour faire des connaissances et travailler. Donc, individus qui normalement seraient dans un même endroit aujourd'hui peuvent être dispersés partout. Ça a résulté en une adoption des groupes de travail virtuels, qui sont dispersés géographiquement et travaillent principalement utilisant des logiciels de communication électronique (HERTEL, 1995). En plus, une décentralisation du travail s'est passée : les gens travaillent sur plusieurs projets coopératifs à la fois et peuvent être associés avec différents individus, qui, eux mêmes, sont associés avec autres groupes et projets (GONZALÉS, MARK, 2005). De cette façon, l'organisation devient un réseau distribué, et les individus doivent gérer plusieurs projets coopératifs à la fois.

A disseminação e adoção de tecnologias computacionais e da internet tem levado a um crescimento nas relações pessoais e de trabalho mediadas pelo computador (WELLMAN, 2005), como pode ser evidenciado pelo sucesso de sites de redes sociais como Orkut (www.orkut.com) ou Friendster (www.friendster.com). Estudos mostram um aumento na adoção de equipes virtuais por parte de empresas, nas quais membros estão dispersos geograficamente e se comunicam e coordenam principalmente através de tecnologias de comunicação (HERTEL et al., 1995). A adoção de ferramentas de comunicação e redes interconecta indivíduos, que utilizam email, messengers e outras mídias eletrônicas para se conhecer, interagir e trabalhar. Como resultado, trabalhadores que costumavam estar no mesmo local agora podem estar dispersos pelo mundo e proximidade não é mais um fator excludente no estabelecimento de relacionamentos ou formação de comunidades.

Adoção de tecnologias de rede, distribuição de organizações e subcontratação também levam a uma descentralização do trabalho: trabalhadores agora participam em diversos projetos colaborativos ao mesmo tempo, e podem estar envolvidos com

diferentes indivíduos que, por sua vez, também estão envolvidos em outros projetos (GONZALÉS, MARK, 2005). Assim como organizações juntam pessoas para trabalhar em diferentes projetos, as pessoas participam em múltiplos projetos, dividindo seu tempo e atenção da maneira mais apropriada (MORAN, 2005).

Dados o aumento de complexidade do ambiente e a crescente especialização dos trabalhadores, organizações frequentemente constituem equipes multidisciplinares para resolução de problemas. Deste modo, trabalhadores de informação e conhecimento muitas vezes se encontram engajados em situações colaborativas. Projetos envolvem múltiplas pessoas, que precisam trabalhar juntas para chegar a objetivos conjuntos. Mesmo em ambientes aparentemente solitários, como pesquisa e academia, existem fortes elementos sociais que influenciam o resultado do trabalho em andamento (LATOUR, 1987). Outros estudos enfatizam a importância de redes sociais e de como estas contribuem para a criação de um ambiente dinâmico e criativo (SAXENIAN, 1994; CASTILLA et al., 2000).

O crescimento das tecnologias de rede mudou a maneira como as pessoas trabalham, e as redes se tornaram lugar comum (CASTELLS, 1996), com indivíduos navegando redes sociais pessoais (WELLMAN, GULIA, 1999) e criando redes “intencionais” para trabalhar (NARDI et al., 2002b). Para funcionar nesta estrutura em rede, trabalhadores precisam gerenciar suas relações com outros, e seus papéis e tarefas nos projetos nos quais estão ativos. Eles precisam organizar a si próprios e ao seu trabalho para realizar diferentes tarefas, frequentemente com diferentes colaboradores. Participação em diferentes grupos significa que, dependendo da situação, um ator pode ter diferentes obrigações e papéis, realizar tarefas diferentes e trabalhar para atingir objetivos diferentes, que devem ser gerenciados para que não conflitem uns com os outros.

Mesmo em situações colaborativas, muito do trabalho ainda é realizado individualmente, e indivíduos alternam trabalho sozinho com colaboração direta e troca de informações (GUTWIN, GREENBERG, 2002; GONZALÉS, MARK, 2005). A interdependência entre atividades de grupos leva a uma necessidade de articulação de esforços (SCHMIDT, BANNON, 1992): atores precisam dividir, alocar, coordenar, agendar e inter-relacionar suas atividades. Para conseguir articular suas atividades e realizar seu trabalho, membros do grupo precisam trocar informações sobre aspectos do trabalho em andamento, informar outros de suas ações individuais, discutir o projeto ou

consultar com os outros (SCHMIDT, BANNON, 1992, p.13). Conhecimentos sobre as capacidades, atividades e disponibilidade dos outros, bem como sobre o progresso do trabalho em geral, competências requeridas, alterações ou correções imprevistas auxiliam o trabalho em conjunto, uma vez que permitem que os membros controlem o projeto conjunto e as atividades uns dos outros, gerenciem níveis de acoplamento (decidir quando passar para cooperação direta ou trabalhar individualmente) e se reorganizem quando necessário. Este conhecimento é comumente chamado de *awareness* (consciência) (GUTWIN, GREENBERG, 2002).

Grupos auto-governados são grupos nos quais atores têm controle sobre alocação de tarefas e operações diárias. Eles emergem de uma necessidade de lidar com eventos imprevistos ou contingências, e por vezes desviam de normas e regras pré-estabelecidas (CARSTENSEN, SCHMIDT, 2002). Mesmo em ambientes controlados como chão de fábrica, equipes têm alto grau de independência e controle sobre seu próprio trabalho (CARSTENSEN, SCHMIDT, 2002, p.50). Informação é necessária para viabilizar estes ajustes e estruturas auto-organizadas.

Consciência é a compreensão de atividades dos outros que provê um contexto para as atividades de uma pessoa (DOURISH, BELLOTTI, 1992). Isto envolve informação sobre os outros, suas tarefas e o ambiente de trabalho como um todo. Consciência é obtida através da percepção de informações do ambiente, colhidas e transmitidas através de interações com o ambiente, e deve ser mantida quando o ambiente se altera (GUTWIN, GREENBERG, 2002). Esta informação é importante para que o grupo possa realizar suas tarefas mais facilmente, pois permite que os indivíduos ajustem seus comportamentos de acordo com a situação e geram sinergia entre o grupo (GUTWIN, GREENBERG, 2004). Quando não existem papéis ou hierarquias estritamente definidas (ou seguidas), este tipo de informação se torna ainda mais importante, visto que atores independentes precisam negociar uns com os outros sobre seu trabalho para chegar ao objetivo final. Decisões são frequentemente o resultado de acertos entre membros do grupo, bem como é o trabalho que é finalmente realizado (BERNOUX, 1995).

Em ambientes co-localizados, muitas destas informações podem ser facilmente disseminadas e percebidas com pouco esforço adicional (KRAUT et al., 1990a). Este quadro muda quando o grupo passa para um ambiente distribuído. Manter consciência dos outros é um problema em ambientes computacionais, pois os recursos de

comunicação e interação são mais pobres que os encontrados em situações face-a-face (GUTWIN, GREENBERG, 2002). Além disso, a distância tem um efeito negativo na manutenção de consciência, uma vez que os indivíduos têm dificuldade de se manterem informados uns dos outros quando não estão no mesmo ambiente físico (KRAUT et al., 1990a). Como a consciência é obtida através de um processo de obtenção e processamento de informações, o processo de manutenção de consciência necessariamente envolve provisão de informações para os usuários. A consciência de um indivíduo sobre o trabalho do grupo será tão boa quanto as informações às quais ele têm acesso.

Devido à grande quantidade de informações disponíveis, usuários devem escolher entre os múltiplos itens que requerem sua atenção, uma vez que informação sobre as atividades dos outros é irrelevante e pode atrapalhar o trabalho (SOHLENKAMP, 1998). Decidir o que deve receber a atenção tornou-se uma tarefa freqüente nos ambiente de trabalho com informação (DAVENPORT, BECK, 2001). Isso gera uma necessidade de controle cuidadoso do fluxo de informações, para prover informações relevantes e minimizar distúrbios. Informação provida por sistemas automáticos deve ser reduzida para apenas os itens de interesse. A questão de que informação é pertinente para o usuário é importante e deve ser trabalhada. Este é o principal foco desta pesquisa, como apresentado nas seções seguintes.

1.1 Base Filosófica

Sociologia organizacional interpreta organizações modernas como redes de atores que interagem, onde as relações entre estes atores estão sujeitas à renegociação constante (BERNOUX, 1995). Estes laços vão além das estruturas, protocolos e hierarquias organizacionais, muitas vezes se sobrepondo a eles. Frequentemente, conhecimento é trocado e trabalho é realizado através destas relações informais entre trabalhadores (para um exemplo, ver GRANOVETTER, 1983). Elas transpassam fronteiras departamentais, funcionais e organizacionais, e constituem uma nova forma de trabalhar à qual a organização deve se adaptar. As redes influenciam o comportamento dos atores ao organizar e conduzir seu trabalho, e suas relações uns com os outros também são influenciadas pela rede como um todo (SIMMELL, 1955).

Quando as pessoas colaboram, elas freqüentemente alternam entre trabalho conjunto e individual, e entre colaboração fortemente e fracamente acoplada,

especialmente quando tarefas individuais têm baixa interdependência (GUTWIN, GREENBERG, 2002). Gutwin e colegas (2005) consideram que uma abordagem sensata neste caso é a de apoiar trabalho individual e prover ferramentas para viabilizar a alternância para colaboração quando necessário. À medida que o indivíduo se torna o centro da sua rede de colaboração, é razoável adotar uma abordagem centrada no usuário, para auxiliar na gerência dos múltiplos arranjos e interdependências estabelecidas com outras pessoas. Schmidt e Bannon (1992) conceitualizam o campo de trabalho cooperativo como uma disciplina que deveria estar preocupada com arranjos relativos à cooperação, e defendem sistemas que apoiem articulação pelos próprios trabalhadores, de modo que possam adaptar sua divisão de trabalho e procedimentos para compensar eventos imprevistos.

Seguindo estas considerações, deixamos de lado perspectivas que analisam grupos como um todo e seguimos uma abordagem que vai do indivíduo para o grupo, usando relações individuais como elemento básico de análise. Esta abordagem é consistente com a posição do indivíduo como um nó em uma rede de relacionamentos (e o nó central em uma rede pessoal), e fornece uma maneira direta de gerenciar as múltiplas relações e envolvimento em diferentes grupos. Gerenciar múltiplos papéis e tarefas tornou-se uma constante nas vidas de trabalhadores da informação: estes indivíduos devem alternar entre múltiplas *esferas de trabalho* (unidades de trabalho ligadas a estruturas colaborativas), alternando sua atenção de um projeto para outro (GONZALÉS, MARK, 2004). Nosso objetivo é projetar sistemas que auxiliem o usuário a gerenciar seus múltiplos compromissos e relacionamentos com outros indivíduos ou grupos, e manter-se a par das múltiplas atividades que ocorrem nas diferentes esferas de trabalho. Damos um passo na direção de criar ligações entre o trabalho individual e conjunto, ajudando cada pessoa a entender como seu trabalho corrente se relaciona com o trabalho dos outros e manter consciência do projeto colaborativo.

1.2 Problema e Objetivos

Alternar entre múltiplas esferas de trabalho envolve lembrar das pessoas e estado do trabalho na da rede e comunicar com estas para criar e manter ligações ou trabalhar (NARDI et al., 2002). Isso implica em uma necessidade de manter-se a par do

estado da rede e de que contatos estão ativos e em que contextos (realizando que atividades ou projetos).

O tema desta pesquisa é a provisão de apoio computacional para distribuição de informações de consciência para o indivíduo posicionado dentro de múltiplas esferas de trabalho. Abordamos esta questão através da análise de atividade na rede, o que requer a determinação de que partes da rede estão ativas. Saindo de um ponto de vista individual, introduzimos o conceito de esfera de colaboração, para unir diversas esferas de trabalho. Constituir uma esfera de colaboração envolve a determinação de que indivíduos estão trabalhando juntos, que recursos estão envolvidos, e a manutenção de uma visão geral da esfera de trabalho envolve determinar quando trabalho individual cai dentro destas esferas de colaboração. Este tipo de informação poderia facilitar a coordenação do grupo, viabilizando ajustes espontâneos quando necessário (SCHMIDT, BANNON, 1992).

Aumentar a consciência dos membros do grupo em relação às atividades do grupo como um todo implica em coletar e disseminar dados da atividade. Distribuição de dados da atividade envolve a criação de um sistema que possa inferir quem está relacionado a determinadas esferas de colaboração e que atividades se relacionam a um projeto colaborativo em andamento. Nossos objetos de estudo são indivíduos que compõem grupos de trabalho, executando as tarefas individuais e cooperativas relacionadas a um projeto colaborativo; as interações que os unem e os artefatos manipulados.

Em um ambiente computacional, grande quantidade de dados está disponível para a coleta, processamento e distribuição. Dados tais como o tempo gasto em determinadas tarefas, arquivos utilizados, tamanhos de arquivo e “timestamps”, registros de interação, sites visitados, as contribuições para fóruns ou participação em listas podem ser automaticamente coletados e disponibilizados aos usuários sem sua intervenção. Entretanto, mais dados não significam mais informação, uma vez que nem todos os dados são úteis a um indivíduo em um momento dado (ENDSLEY, 2000). Lidar com a grande quantidade de informação à qual indivíduos são expostos diariamente torna-se uma tarefa difícil. Informação requer atenção, e o aumento na informação disponível cria uma demanda de atenção que é difícil de suprir (VERTEGAAL, 2003).

Apresentamos uma arquitetura baseada em agentes para apoiar a coleta, e distribuição automáticas de dados referentes a espaços de trabalho individuais dos usuários. Os agentes criam modelos de usuários baseados nas interações de cada um, onde os laços com outros são estabelecidos e mudam de acordo com a situação. Estes modelos são dinâmicos e são ajustados automaticamente com tempo. Como nem todos os dados são relevantes, apresentamos um método para filtrar informação relevante. Esta combinação deve viabilizar um sistema computacional de análise e disseminação proativa de informação na qual um usuário pode estar interessado sem sobrecarregar o usuário com demasiada informação.

Nossa abordagem é aplicável no contexto de atividades que utilizam conhecimento e envolvem a manipulação de informação em formato digital. Foi concebida para grupos distribuídos, que utilizam meios interativos de comunicação baseados em texto, assim fornecendo subsídios para a análise. Para alcançar nossos objetivos, nós:

- Revisamos bibliografia sobre trabalho colaborativo e auto-organização em grupos distribuídos, relacionando casos de estudo com observações filosóficas e sociológicas;
- A partir dos pontos de vista teóricos, desenhamos uma abordagem técnica para apoiar colaboração em rede;
- Realizamos observações de grupos de trabalho, tanto em ambiente distribuído como em interações co-localizadas, para descobrir necessidades e requisitos para suporte computacional;
- Desenvolvemos uma arquitetura para apoiar a captura e distribuição automática de informação, como parte de uma solução técnica;
- Desenvolvemos um método para detectar ligações ativas em uma rede intensional de um usuário;
- Desenvolvemos um método para relacionar atividades e pessoas às suas esferas colaborativas;
- Construimos protótipos para verificar e refinar nossas idéias.

Estas são explicadas ao longo deste documento.

1.2.1 Aplicabilidade e Limitações

Esta abordagem foi projetada com determinadas características em mente. As atividades de grupo são freqüentemente classificadas de acordo com localização física dos participantes e participação temporal no grupo (DIX et al., 1998). Nossa pesquisa se foca nos grupos distribuídos, onde os indivíduos não estão necessariamente no mesmo local ao trabalhar, e independente do tempo (isto é, podem trabalhar simultaneamente ou não). Em nossas investigações, verificamos que a abordagem funciona melhor quando há poucas oportunidades para interações face a face.

Tabela 1-1: Matriz Espaço x Tempo (adaptado de DIX et al., 1988 e GRUDIN, 1994). Tecnologia não computacional em itálico.

	Mesmo Local	Locais Diferentes
Mesmo Tempo	<i>Conversas</i> <i>Face a face</i> Facilitação de reuniões	<i>Telefone</i> Chats
Tempos Diferentes	<i>Post-its</i> Salas compartilhadas	<i>Carta</i> Fóruns

Outra categorização define três tipos de esforço colaborativo que podem ser apoiados através de sistemas de groupware (NUNAMAKER, et al. 2001). Estes são:

- coletivo (onde cada indivíduo trabalha sozinho e as partes produzidas são juntadas);
- coordenado (onde membros do grupo realizam esforços individuais e precisam se coordenar para atingir seu objetivo);
- em concerto, onde todos os membros devem fazer um esforço síncrono para atingir o objetivo comum.

Nos focamos em esforços coordenados, especialmente em grupos onde indivíduos empreendem determinadas tarefas individualmente e devem então coordenar uns com os outros a fim atingir objetivos comuns. Neste caso, a produtividade é dependente do nível do esforço individual e da coordenação entre indivíduos.

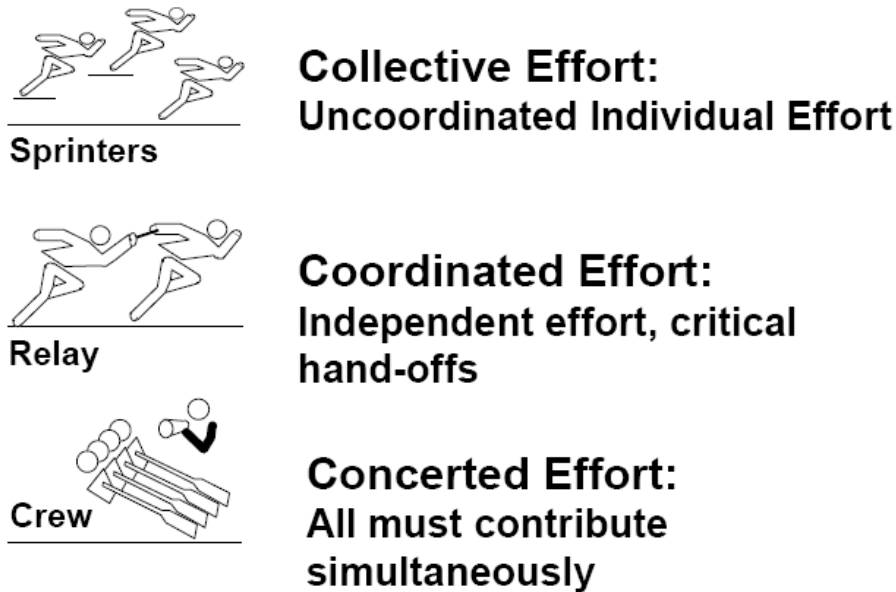


Figura 1-1: Diferentes tipos de esforço (NUNAMAKER et al., 2001).

1.2.2 Abordagem Técnica

Uma das técnicas usadas por sociólogos para analisar o grupo e as interrelações individuais, a análise de redes sociais analisa grupos de pessoas enquanto redes, onde membros estão interligados por relacionamentos que representam aspectos sob estudo (SCOTT, 1991; WASSERMAN, FAUST, 1994). Esta técnica emprega operações de grafo e cálculos específicos para detectar a importância de um nó, a força do laço e a relevância dentro do grupo. As análises podem ser conduzidas baseadas na rede inteira (análise sociocêntrica) ou em um ponto de vista individual (análise egocêntrica) (GARTON et al., 1999).

Dado que o indivíduo é o nó central em uma rede pessoal e mantém colaborações diferentes, adotamos a análise de rede social egocêntrica e análise de interação como base de trabalho. Isto nos permite enfatizar o papel central do indivíduo na gerência de seus laços ou sua alocação de atenção ao trabalhar. Nossa abordagem é de analisar laços individuais a outros para descobrir padrões de colaboração e atividades que pertencem a estes. Esta análise leva em consideração a alocação da atenção de cada indivíduo para determinar o interesse e urgência: os pares que recebem mais atenção são provavelmente aqueles com os quais o usuário está atualmente colaborando. Adicionalmente, a análise é usada para relacionar atividades individuais aos laços interpessoais, a fim determinar o contexto social e os indivíduos envolvidos na manipulação de um artefato.

Para nosso protótipo de sistema, um projeto baseado em agentes foi usado para implementar uma arquitetura peer-to-peer (P2P). Desenvolvemos um cenário de consciência, onde os agentes recolhem e distribuem a informação a seus usuários, de acordo com sua análise das ligações uns com os outros e da manipulação da informação. Nossos estudos e análises, que serão apresentados em umas seções mais atrasadas, indicam que esta é uma técnica viável.

Consciência de grupo já foi objeto de muitos estudos, e diferentes sistemas de apoio a grupos foram propostos fornecer consciência – transmitindo informação a seus usuários. Para ser útil a um usuário, entretanto, a informação necessita ser fornecida no tempo correto, dependendo das atividades e do contexto global. Neste trabalho, consideramos a distribuição da informação de consciência dentro de um contexto colaborativo. Deste modo, a manutenção da consciência em uma situação de rede intensional envolve descobrir as ligações a outras pessoas e a que estas dizem respeito. Nosso método proposto de filtragem da informação aplica análise de interação e de rede social: as subredes às quais o usuário destina sua atenção são consideradas ativas, e exigem coordenação e sincronização. Atividades executadas por membros destas subredes que se relacionam a tópicos da discussão entre eles são aquelas sobre as quais o usuário pode necessitar saber.

Este texto está organizado da seguinte forma: começamos fornecendo motivação e uma introdução geral, apresentando o problema e nossos objetivos (seção 1). Isto é seguido por uma exploração teórica dos conceitos e teorias envolvidos, para ajudar ao leitor a compreender nossa linha de pensamento, conceitos e idéias sociológicas e organizacionais envolvidas, e como estes se tornaram uma abordagem e solução técnica para o problema em estudo (seção 2). Esta estrutura conceitual é seguida na seção 3 por uma apresentação de soluções técnicas relacionadas. A discussão é deslocada então para alguns estudos de caso empreendidos para compreender melhor o trabalho fracamente estruturado na seção 4. Uma arquitetura proposta é apresentada na seção 5 e seguida por um método para a determinação de contextos colaborativos na seção 6. Isto é seguido então por análises na seção 7 e por uma discussão da aplicabilidade das vias propostas de solução e de trabalhos futuros na seção 8.

Ce qui nous intéresse est la distribution d'informations pour créer la conscience des multiples environnements de travail avec lesquels un travailleur est concerné, pour qu'il puisse ajuster et coordonner ses actions avec les autres. Cette thèse concerne le travail distribué, où les individus interagissent à travers des logiciels, de façon synchrone et asynchrone. Les groupes de travail doivent avoir une hiérarchie plus égalitaire et une structure en réseau. Ces groupes normalement ont plus de liberté pour ajuster ses directions et activités quand nécessaire, et ont besoin d'informations pour le faire. Dans ces groupes, les efforts individuels sont coordonnés par des échanges d'informations et ajustements et négociations individuelles.

2 Fundamentação Teórica

L'interdépendance entre tâches crée une nécessité d'articulation d'efforts (SCHMIDT, BANNON, 1992): les acteurs doivent diviser, partager, coordonner et trouver les corrélations entre leurs activités. Connaissance des activités et disponibilités des autres et du progrès du travail en général permet que les individus maintiennent le contrôle du projet, décident quand changer entre coopération plus proche ou plus distante et réorganisent si nécessaire, et est essentielle pour que le groupe puisse travailler ensemble. Cette connaissance est normalement appelé awareness (ou conscience), et a été étudiée en détail dans environnements synchrones.

O embasamento teórico é apresentado neste capítulo, para pavimentar a apresentação da nossa abordagem conceitual e solução nos capítulos seguintes. Apresentamos conceitos e técnicas de sociologia e teoria das organizações que discutem indivíduos e organizações, e como as pessoas se associam para conduzir seu trabalho. Também apresentamos um arcabouço desenvolvido para o design e análise de grupos de apoio à colaboração e análise de redes sociais como uma técnica e efeitos da distância e informação que influenciaram os nossos.

2.1 Embasamento Sociológico e Organizacional

A disseminação da tecnologia de redes trouxe mudanças para a organização que conduzem novas estruturas, em forma de rede. A forma de rede, na qual elementos estão interconectados e a inteligência está na rede, é agora aplicável a qualquer tipo de atividade ou segmento que possa ser conectado eletronicamente, como extensivamente descrita em (CASTELLS, 1996). Este formato de rede é bem adaptado à crescente complexidade do ambiente, pois fornece adicionada flexibilidade às organizações (BERNOUX, 1985).

Redes existem em um nível global, conectando empresas, com outras organizações, imprensa e governo. Na economia em rede, atores são fortemente

independentes, uma vez que organizações (especialmente as pequenas e médias) formam redes para atingir seus objetivos (CASTELLS, 1996, p.218), subcontratando e colaborando em partes diferentes do ciclo de produção para ganhar penetração de mercado. Alianças estratégicas tornaram-se uma maneira comum de funcionamento de organizações no mundo ligado em rede.

Redes também são encontradas dentro de organizações, com a adoção de novas estruturas: empresas passaram de estruturas horizontais a burocracias verticais, com o objetivo de ganhar performance, agilizar a produção e cortar custos (CASTELLS, 1996, p.221). A tecnologia e o aumento da competitividade no ambiente mudaram a maneira de trabalhar das pessoas, levando à reestruturação de empresas e reforçando a individualização do trabalho dentro dos processos de trabalho. Na nova organização, a administração é descentralizada, o trabalho é individualizado e mercados são personalizados (CASTELLS, 1996, p.330). A tecnologia da informação simultaneamente viabiliza a descentralização de tarefas e coordenação em redes interativas de comunicação em tempo real, locais ou globais. Como regra, as formas tradicionais de trabalho estão aos poucos sendo substituídas por contratos mais flexíveis e individualizado (CASTELLS, 1996, p.337).

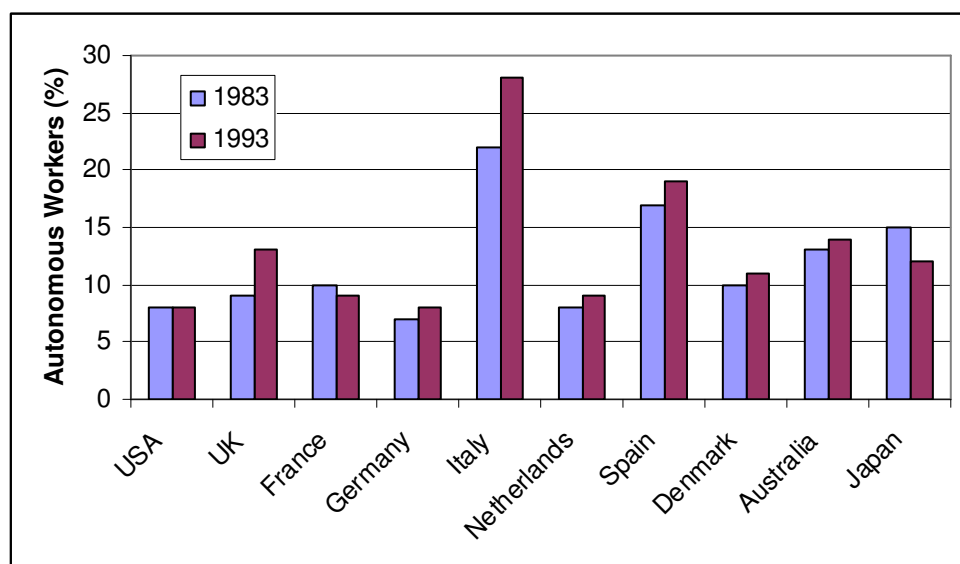


Figura 2-1: Trabalhadores autônomos como uma porcentagem da mão-de-obra (de Castells, 1996, p332)

Consistente com esta visão, Bernoux (1995) descreve organizações como redes de atores em interação. Estes atores são autônomos e cada um tem uma visão pessoal da organização, de acordo com seus ou seus objetivos e necessidades. As associações entre empresas são governadas por contratos e por regras de seu próprio desígnio, sujeitas à

negociação constante e redefinição quando a situação exige. A organização transforma-se um sistema dos laços sociais, onde os atores criam o sistema, as regras, a rede e suas próprias interpretações dela (BERNOUX, 1995, p.173, BERNOUX, 1985). Não somente os indivíduos definem o grupo, ajustando regras, métodos e protocolos, mas também o grupo influencia o indivíduo, com estas mesmas regras e protocolos (SIMMEL, 1955). Assim, um indivíduo muda, define e é restringido pelos grupos em que participa.

Freqüentemente, o conhecimento é trocado e o trabalho é empreendido através de relações informais entre os trabalhadores, que se organizam em comunidades de prática (WENGER et al., 2002, WENGER, 1998) ou se valem de suas redes pessoais levantar informação ou trabalhar (GRANOVETTER, 1973). Assim, uma organização fornece estruturas e regulamentos, mas funciona tanto em consequência das interações dos atores e relações quanto da estrutura organizacional fornecida (BERNOUX, 1995).

Uma organização pode assim ser vista como uma coalizão de indivíduos motivada por interesses individuais (BERNOUX, 1985). A organização formal transforma-se uma estrutura contratual de governança que influencia o comportamento dos indivíduos e dos coletivos de acordo com os interesses dos outros membros da organização e de partes interessadas externas. Trabalho é conduzido através do estabelecimento de arranjos cooperativos de trabalho, que emergem para assegurar exigências específicas e se dissolvem quando são mais necessários. Estes arranjos são organizados de acordo com necessidades específicas, mas seguindo padrões estabelecidos (SCHMIDT, BANNON, 1992).

São estas redes que permitem a organização de se adaptar rapidamente às mudanças no ambiente (BERNOUX, 1995). Elas atravessam departamentos, funções e às vezes até mesmo limites organizacionais, e constituem uma maneira nova de trabalhar à qual a organização deve se adaptar. Conseqüentemente, organizações modernas requerem coordenação e integração de atividades além dos limites funcionais e profissionais, e os indivíduos dentro destas organizações necessitam de ferramentas para melhor compreender e trabalhar nesta configuração em rede. Os sistemas de suporte a grupo tradicionais se focam no desenvolvimento de modelos eficientes das estruturas e processos relevantes ao trabalho e em maneiras adequadas à organização para apresentá-los, mas não se aprofundam muito em como sistemas podem apoiar estas

redes informais ou levá-las a funcionar mais eficazmente (SCHMIDT, BANNON, 1992).

Mesmo em ambientes rígidos como os de chão de fábrica, houve um aumento na autonomia do trabalhador, que levou a um crescimento no desempenho em grandes organizações (CASTELLS, 1996, p.310). Carstensen e Schmidt (2002) descrevem grupos auto governados como aqueles em que os atores têm o controle sobre a alocação do trabalho, o planejamento e o controle do dia a dia. Estes grupos emergem fora de uma necessidade segurar eventos ou contingências imprevisíveis, e permitem uma organização de adaptar-se rapidamente às demandas novas geradas pelo ambiente, às vezes fugindo das normas e regras pré-estabelecidas. Em muitos casos, os grupos são compostos de pares, onde não há nenhuma estrutura hierárquica formal. Neste cenário, muitas das decisões advêm de arranjos entre pares, bem como o trabalho que é efetivamente realizado (BERNOUX, 1995).

Devido à interdependência entre suas tarefas, trabalhadores têm que articular (isto é, dividir, alocar, coordenar, programar, relacionar, etc..) suas atividades (SCHMIDT, BANNON, 1992). A organização fornece um padrão relativamente estável de arranjos cooperativos e de uma estrutura de governança para regular os interesses diversos dentro do grupo cooperativo, mas novos grupos de trabalho colaborativo emergem porque os atores não poderiam realizar determinadas tarefas tão bem se o fizessem individualmente (SCHMIDT, BANNON, 1992). Na maioria de situações, os indivíduos têm algum nível do controle sobre seu trabalho, e são capazes de se ajustar para compensar erros ou circunstâncias incomuns, lidando com eventos imprevisíveis ou contingências.

A informação é importante para o funcionamento e a organização dos elementos na rede. Castells (1996, p.119) ressalta a importância da informação na nova economia, uma vez que a produtividade e a competitividade das unidades ou dos agentes nesta economia são fortemente dependentes de sua capacidade gerar, processar e aplicar eficientemente informação baseada no conhecimento.

2.2 Redes Pessoais e Intensionais

Os indivíduos podem ocupar posições diferentes na estrutura de organização, mas tem três papéis principais dentro das redes: Trabalhadores ativos, que estabelecem conexões por conta própria e navegam pela rede de acordo com suas necessidades;

trabalhadores passivos, que são ligados a outros mas não decidem quando, como, a quem nem por que e trabalhadores não relacionados estão presos às suas tarefas específicas, definido por instruções unilaterais, não-interativos (CASTELLS, 1996). Isto apresenta novos desafios para trabalhadores ativos, que devem administrar laços com outros, estabelecendo novas conexões quando necessário.

Através de uma série de estudos, Wellman detectou a existência de comunidades pessoais em sociedades avançadas: redes sociais individuais que existem tanto *online* quando *offline*. Neste caso, comunidades são substituídas por redes pessoais, em que laços são especializados e diversificados, como as pessoas criam seus “portfólios pessoais” (WELLMAN, GULIA, 1999).

Em estudos recentes, Nardi e colegas (2002) documentaram o crescimento de redes pessoais no local de trabalho. Eles chamam estas redes de *redes intensionais* e afirmam que, neste tipo de ambiente, a unidade mais importante de análise para trabalho cooperativo apoiado por computador não está no nível de grupo, mas no nível individual. Eles também ressaltam que relacionamentos fora da organização (e.g., governo e imprensa) são importantes para muitos negócios, e que muitas corporações agora operam num meio crescentemente distribuído, onde trabalhadores, empreiteiros, consultores e outros contatos estão em situações diferentes. Dentro de organizações, reorganizações constantes significam que responsabilidades, colegas e relacionamentos mudam frequentemente.

As *redes intensionais* são redes de trabalhadores sociais pessoais. As redes intensionais são emergentes (podem ser ficar ativas para realizar um trabalho em particular) e históricas (o trabalho é baseado em relacionamentos e experiências compartilhadas conhecidas, e cada novo projeto em conjunto se adiciona à história do relacionamento, alterando-o). Seções diferentes da rede tornam-se ativas dependendo do trabalho em andamento. Quando um conjunto de contatos está ativo, as relações são constantemente renovadas através de atos de comunicação (NARDI, 2002, p.221). A sub-estrutura ativa é um resultado da interação humana, que acontece devido ao trabalho continuado do grupo. As redes de intensionais pulsam como declínios de atividade e fluxos, e partes diferentes da rede animam. Trabalhar em redes intensionais traz complicações adicionais para os participantes, na medida em que precisam se lembrar de outros participantes, papéis e responsabilidades (NARDI, 2002, p213). Sub-redes ativas são as partes de uma rede (contatos) que mantêm um trabalhador mais

ocupado, mas mesmo as relações dormentes necessitam ser lembradas para trabalho conjunto futuro.

Criar e manter relacionamentos confiáveis que possam ser utilizados para propósitos de trabalho exige trabalho cuidadoso e ponderado de proprietários de rede (NARDI, ENGESTROM, 1999; SCHWARZ et al., 1999). Os grupos de trabalho são constituídos por conjuntos de pessoas encontrados via redes intensionais, a invés de serem equipes criadas por planejamento e estruturação organizacional. Muito do trabalho é empreendido via ferramentas interpessoais de comunicação, tal como email, messengers ou o telefone, para coordenar, administrar interdependências e se manterem a par de trabalho continuado.

Outras formas semelhantes de associação foram descritas em trabalho recente: *Coalizões* (Coalitions) são grupos de pessoas trabalhando num mesmo problema de fins diferentes, normalmente ignorante um do outro (ZAGER, 2000). As coalizões estão fracamente acopladas, temporárias e fluidas. Os indivíduos são limitados por interesses compartilhados, e normalmente não comunicam uns com os outros, são parte de sub-estruturas de organização diferente. Como indivíduos normalmente não são em contato, trabalho às vezes é duplicado. *Nós* (Knots), por outro lado, são grupos pequenos montados pela organização com um propósito explícito (ENGESTROM, 2005). Em nós, combinações de pessoas, recursos e tarefas são únicos e o grupo não é longo. Nós ocorrem em situações guiadas, em que indivíduos se juntam para resolver um problema específico, e normalmente se preocupam principalmente com a designação de papéis, sendo projetado pela organização.

Comunidades de Prática são grupos de pessoas com interesses compartilhados que se juntam para trocar informação e experiências (Wenger, 1998). Estes grupos se formam espontaneamente dentro de organizações, e congregam pessoas que freqüentemente devem encarar problemas semelhantes. As comunidades de prática são centrais para o aprendizado e disseminação de conhecimento dentro de organizações, já que fornecem um ambiente favorável para trocar lições aprendidas. Estruturalmente, há normalmente um grupo de membros centrais, que dedica uma quantia significativa de tempo à comunidade, um conjunto maior de membros ativos, que participa ativamente, e um conjunto periférico, de membros que participam muito pouco, entrando como observadores e aprendendo como resultado. A participação é fluida, com indivíduos tornando-se mais ou menos envolvidos de acordo com tempo e interesses.

Zacklad (2003) propõe o estudo de indivíduos em situações estruturalmente abertas, onde é possível mudar as estruturas sociais que ligam estes indivíduos e as configurações nas quais eles trabalham. Nestas ditas "*Comunidades de Ação*", a atividade coletiva é orientada a metas, mas os meios para atingir estas metas não são nem completamente formalizados nem uniformizados. Atores têm significativa autonomia e são livres para definir suas modalidades de coordenação e adaptar se a situações emergentes.

Em comunidades de ação, a coordenação é alcançada através de transações intelectuais, ou comunicações lingüísticas em que atores trocam conhecimento e estabelecem compromissos. O nível básico de coordenação é alcançado via percepção mútua, em que membros se coordenam através da coleta e troca informação sobre as atividades de sócios e seus campos de operação (ZACKLAD, 2003).

Um grande número de teorias de organização descreve grupos onde atores têm oportunidade para transformar a estrutura da organização, mudar a estrutura do grupo, alterar papéis de membros ou características funcionais de ferramentas disponíveis (ZACKLAD, 2003). As teorias supracitadas realçam situações mais soltas, onde indivíduos têm liberdade alterar configurações e improvisar quando necessário. O fio unificador é a possibilidade negociar mudanças com outros membros da equipe. As comunidades de ação são o foco de nossa pesquisa: grupos onde indivíduos têm a liberdade de se reorganizar para se ajustar a mudanças, e se coordenam através da percepção mútua, reunindo e processando informação sobre as atividades dos outros. Nossa principal questão de pesquisa é se é possível fornecer tal informação automaticamente e, neste caso, como descobri-la.

2.3 Distância e Colaboração

Mesmo em equipes fortemente acopladas, indivíduos ainda executam a maioria das atividades individualmente (ANDRIESSEN, 2003). Quando indivíduos colaboram, eles freqüentemente alternam entre trabalho individual e trabalho compartilhado, e entre colaboração fracamente e fortemente acoplada (GUTWIN, GREENBERG, 2002). Isto é especialmente verdadeiro quando há baixa interdependência entre eles. Nestes casos é recomendável adicionar apoio à colaboração a ferramentas individuais de trabalho, viabilizando a alternância para colaboração próxima quando necessário (GUTWIN et al, 2005). A consciência de esforços correntes e passados torna-se necessária, uma vez que

um indivíduo pode trabalhar temporariamente num artefato compartilhado e outro pode utilizá-lo mais tarde (EDWARDS, MYNATT, 1997).

A globalização criou uma tendência de criação de equipes geograficamente difundidas (virtuais), onde o trabalho dos indivíduos é mediado pela tecnologia de computador (HERTEL, 2005). As principais razões pelas quais equipes virtuais surgem são: redução esperada de custos, aumento de flexibilidade, tecnologias facilmente acessíveis, que também facilitam captura e disseminação de informação e adoção de redes de organização (ANDRIESSEN, 2003). Trabalho distribuído, no entanto, dificulta a comunicação. Proximidade física reduz um número de problemas que surgem quando se colabora, tal como encontrar parceiros, definir problemas, planejar, supervisionar colegas e subordinados, etc. Os estudos mostraram que essa facilidade de contato tem um papel importante na escolha de parceiros, e que a distância cria uma barreira entre indivíduos, como descrito por Kraut e colegas (1990a), que mostrou que a frequência de comunicação e colaboração diminui com o aumento na distância.

Interações síncronas co-localizadas têm algumas características que são difíceis de reproduzir em ambientes computacionais, tais como realimentação rápida, múltiplo canais, nuances de informação, contexto local compartilhado, encontros informais “no corredor” e comunicação implícita (OLSON, OLSON, 2000), e argumentam que a maioria destas características são fracamente apoiadas pelas tecnologias de hoje. A lista é mostrada na Tabela 2-1.

Uma estrutura mais solta e a distância entre os membros podem levar a uma diminuição em envolvimento e interação. Como uma consequência, indivíduos perdem oportunidades para colaboração, e às vezes terminam executando tarefas que se sobrepõem ou duplicando trabalho porque ignoram estas atividades. Grupos distribuídos exibem uma tendência a mudar estrutura organizacional e a distribuição de tarefas para reduzir o acoplamento de membros de não co-localizados, reduzindo assim as necessidades de coordenação (OLSON, TEASLEY, 1996). Isto leva à fragmentação, uma vez que membros de equipes diferentes podem não estar a par das últimas evoluções em sub-projetos das outras equipes, e o grupo como um todo torna-se fragmentado.

Tabela 2-1: Características de interações síncronas co-localizadas e seu apoio pelas tecnologias de hoje (adaptado de Olson e Olson, 2000).

Característica	Descrição	Implicação	Apoiado por hoje as tecnologias?
Realimentação rápida	Como interações fluem, realimentação é tão rápida quanto pode ser	Correções rápidas, possíveis quando são notados equívocos ou discórdias	Não apoiado
Múltiplos canais	A informação entre fluxos de participantes em muitos canais. Expressões faciais, gesto, postura de corpo, e assim por diante	Há muitos meios de transportar uma mensagem sutil ou complexa; também fornece redundância	Fracamente apoiado
Informação pessoal	A identidade de contribuintes a conversa normalmente é sabida	As características da fonte podem ser levadas em conta	Fracamente apoiado
Informação com Nuances	O tipo de informação que flui é freqüentemente análogo ou contínuo, com muitas dimensões sutis (e.g., gestos)	Diferenças pequenas de significado podem ser transmitidas; informação pode ser facilmente modulada	Fracamente apoiado
Contexto local compartilhado	Os participantes têm uma situação semelhante (tempo de dia, acontecimentos locais)	Uma armação compartilhada nas atividades; permite para socializar fácil assim como entendimento mútuo sobre ele que está na mentes de cada um	Não apoiado
Tempo informal de "corredor" antes de e depois	Interações improvisadas acontecem entre os subconjuntos de participantes em chegada e partida	Trocas oportunistas de informação acontecem, e ligações sociais importantes ocorrem	Fracamente apoiado
Co-referencias	A facilidade de estabelecer referência em conjunto a objetos	O olhar e gesto facilmente podem identificar o referencial de termos de dêiticos	Não apoiado
Controle individual	Cada participante livremente pode escolher o que assistir a e mudar o foco de atenção facilmente	Controle rico e flexível de como os participantes reagem ao que acontece	Não apoiado
Indicações implícitas	Uma variedade de dicas normalmente está disponível na periferia	Operações naturais de atenção humana fornecem acesso à informação contextual importante	Não apoiado
Referências espaciais	As pessoas e objetos de trabalho são localizados em espaço	Tanto pessoas como idéias podem ser referidas a espacialmente; "quadros no ar"	Não apoiado

2.4 Consciência em Trabalho Colaborativo

A pesquisa de consciência de situação (situation awareness - SA) focaliza na capacidade de cada indivíduo perceber elementos e nos processos cognitivos envolvidos em manter consciência do ambiente. Endsley (2000) define consciência de situação

como o processo de perceber elementos do ambiente, interpretar seu significado e projetar seu estado no futuro próximo. Esta informação é usada como uma base para decisão tomada durante o trabalho individual. Um número de fatores influencia esta percepção, tanto pessoais (ex, capacidades ou treinamento) como inerentes à tarefa ou sistema (ex, complexidade ou projeto de interface). A consciência de situação influencia a tomada de decisão, que por sua vez influencia o desempenho de ações que afetam o ambiente (estes efeitos também serão percebidos recomeçando o ciclo). O modelo de SA de Endsley é mostrado na Figura 2-2.

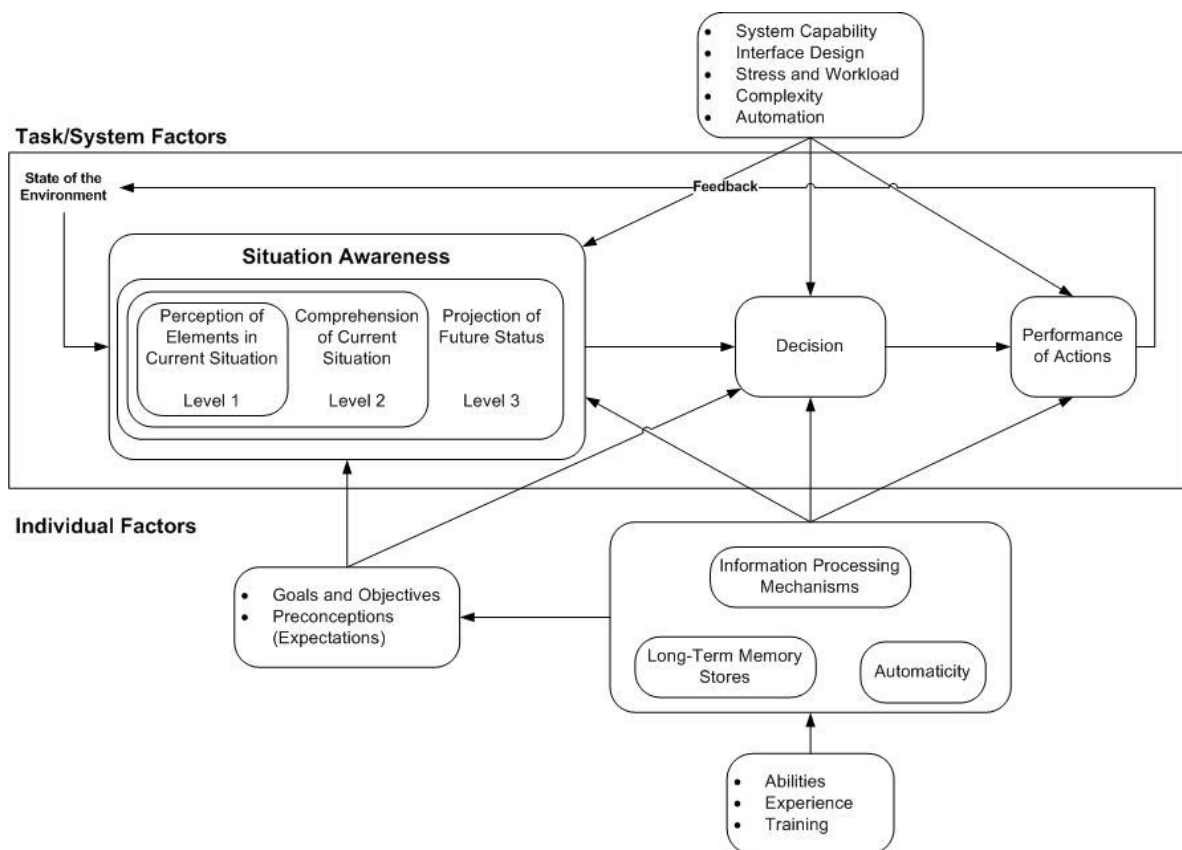


Figura 2-2: Funcionamento de consciência de situação (de Endsley, 2000)

Dourish e Bellotti (1992) definem consciência como um entendimento de atividades de outros, que fornece um contexto para a própria atividade. Esta informação assegura que contribuições individuais tenham ligação com as atividades do grupo, e capacita indivíduos a avaliar as ações dos outros com respeito a metas e ao progresso do trabalho, que por sua vez permite aos indivíduos ajustar seu comportamento ou tomar medidas de acordo com a situação. Elaborando sobre estes conceitos, Gutwin e Greenberg (2002) definem consciência de área de trabalho como o entendimento das interações de outras pessoas com a área de trabalho compartilhada. É conhecimento sobre o grupo trabalhando no ambiente, que cria uma compreensão das pessoas dentro

de uma área de trabalho. Estas duas definições especializam a definição de SA ao definir o ambiente como uma área de trabalho colaborativa, onde informação sobre outras atividades e estado do grupo é um ativo valioso.

O modelo de consciência de Rodden (1996) considera um conjunto de objetos no espaço, que reagem baseados em seus níveis de consciência. A consciência é manipulada via foco e nimbo, subespaços aos quais um objeto dirige sua presença ou atenção. A consciência é a sobreposição entre nimbo e foco, onde:

- Nimbo é a informação distribuída por cada elemento em espaço que pode ser percebido por outros; e
- Foco descreve os elementos em que um operador dirige sua atenção.

Assim, a consciência do indivíduo A em direção ao indivíduo B é a interseção entre a informação sendo distribuído por B e a informação na qual A está interessado, como mostrado em Figura 2-3. Este modelo detalha como informação para manter consciência é obtida e mostra que a atenção é um aspecto chave no processo, e como a consciência é afetada pelo foco de cada indivíduo e pela informação fornecida por cada elemento (nimbo).

Num ambiente co-localizado, indivíduos distribuem uma quantia grande de informação enquanto trabalham, que podem ser captadas por qualquer um prestando atenção a elas. Em cenários computacionais, operadores distribuem informação através das aplicações com as quais eles trabalham e do sistema operacional, que normalmente não é transmitido a outros.

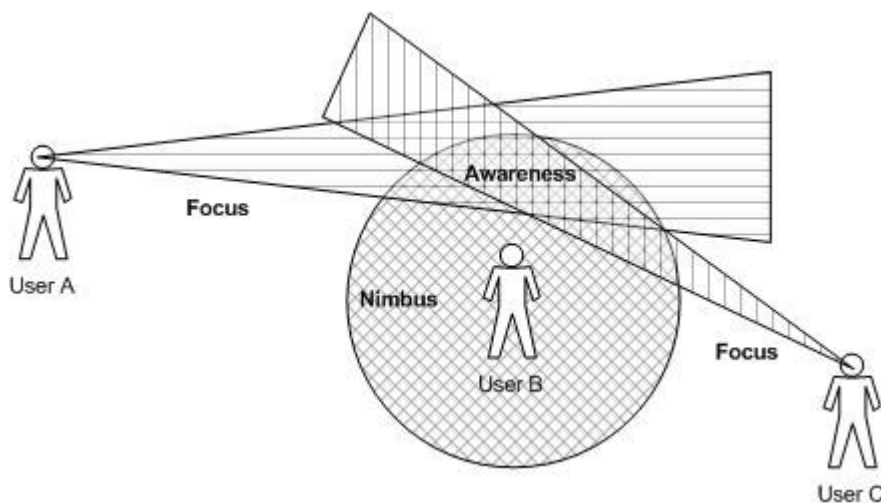


Figura 2-3: A consciência como o sobreposição entre foco e nimbo

Num ambiente colaborativo, consciência envolve conhecimento sobre colaboradores (presença, identidade e autoria), atividades nas quais eles trabalham (ações, intenções e artefatos manipulados) e onde (situação de trabalho, direção de olhar, vista e alcance individual). Informação histórica de consciência também inclui ação, artefato e histórico de acontecimentos, e devem ser fornecidos em situações assíncronas de trabalho (GUTWIN, GREENBERG, 2002). Esta estrutura fornece um ponto de referência para projetistas considerarem elementos de consciência e que informações fornecer em determinadas situações (baseado em que questões devem responder), e é resumido na Tabela 2-2.

Tabela 2-2: Os elementos de consciência de área de trabalho e as perguntas que eles endereçam (adaptou de GUTWIN, GREENBERG, 2002)

		<i>Presente</i>	<i>Passado</i>
Categoria	Elemento	Perguntas específicas	Perguntas específicas
Quem	Presença	Há alguém na área de trabalho?	Quem estava aqui e quando?
	Identidade	Quem participa? Quem é isso?	
	Autoria	Quem faz isso?	
O que	Ação	O que eles fazem?	O que tem uma pessoa fez?
	Intenção	De que meta essa ação é parte?	
	Artefato	Em que objeto trabalham?	
Onde	Situação	Onde trabalham?	Onde uma pessoa foi?
	Olhar	Onde olham?	
	Visão	Onde podem ver?	
	Alcance	Onde podem alcançar?	
Como	História de ação		Como que operação aconteceu?
	História de artefato		Como este artefato veio estar neste estado?
Quando	História de acontecimentos		Quando esse acontecimento aconteceu?

Em trabalho posterior, Tam e Greenberg (2006) trabalharam sobre os conceitos envolvidos em consciência de mudanças assíncronas. Eles desenvolvem uma visão de cada elemento de consciência de acordo com três perspectivas diferentes (artefato, pessoa e área de trabalho), que corresponde aos três modos em que um operador pode necessitar ver a área de trabalho compartilhada em dado momento. Estes são resumidos em Tabela 2-3. Quando alternando entre tarefas, esta informação torna-se mais importante, como um indivíduo deve poder rapidamente alternar entre esferas de

trabalho, e necessita estar atualizado com o estado atual do projeto colaborativo para poder se integrar com outros participantes.

Tabela 2-3: Elementos assíncronos de consciência de mudança de acordo com diferentes perspectivas, e perguntas que devem ser consideradas (adaptado de Tam e Greenberg, 2006)

Categoria	Elementos de informação	Perguntas específicas		
		Visão baseada em artefato	Visão baseada em pessoa	Visão baseada na área de trabalho
Onde	História de situação	Onde este artefato estava? Onde o artefato está agora?	Onde na área de trabalho uma pessoa visitou?	Onde as pessoas estiveram na área de trabalho? Onde artefatos estavam na área de trabalho?
	Fite história		Onde na área de trabalho uma pessoa olhou?	Que partes da área de trabalho as pessoas olharam?
	Edite história	Onde este artefato foi durante o tempo que eu fui longe?	Onde na área de trabalho uma pessoa fez mudanças?	Em que partes das de área de trabalho as pessoas fizeram mudanças?
Quem	História de presença	Quem olhou este artefato?	Com quem esta pessoa interagiu?	Quem esteve na área de trabalho?
	Identidade	Quem mudou este artefato?		
	História de conjunto de leitores		Quem fez mudanças com esta pessoa?	Quem olhou a área de trabalho?
	História de autoria			Quem fez mudanças à área de trabalho?
O que	História de ação	Que mudanças foram feitas no artefato?	Que artefatos uma pessoa olhou? Que artefatos uma pessoa mudou? Em que atividades uma pessoa trabalhou?	Que mudanças ocorreram na área de trabalho? Que artefatos foram vistos? Que artefatos foram mudados?
Como	História de processo História de resultado	Como este artefato mudou?	Como uma pessoa mudou coisas?	Como a área de trabalho mudou?
Quando	História de acontecimento	Quando este artefato foi mudado? Quando uma mudança particular a este artefato foi feita? Em ele que mandam eram mudanças feitas neste artefato?	Quando uma pessoa fez mudanças? Quando uma pessoa fez uma mudança particular? Em ele que manda fez esta pessoa faz mudanças?	Quando as mudanças foram feitas à área de trabalho? Quando um particular mudou na área de trabalho ocorre? Em ele que manda fez mudanças à área de trabalho ocorre?
Por que	História cognitiva História motivacional	Por que este artefato foi mudado?	Por que uma pessoa fez essa mudança?	Por que mudanças foram feitas na área de trabalho?

Pesquisas paralelas introduzem a idéia de contexto em ambientes colaborativos. O contexto é uma descrição complexa de conhecimento em vários níveis (e.g., físico, social, histórico, etc.) dentro do qual ações e acontecimentos ocorrem (BORGES et al., 2004). Conhecimento contextual é usado quando indivíduos executam tarefas, quando refletem sobre o contexto subjacente para decidir sobre cursos de ação. Em trabalho de grupo, contexto pode ser visto como informação em uma de 5 categorias: pessoas e grupos, tarefas programadas, relacionamentos entre as pessoas e tarefas, ambientes onde a interação acontece e tarefas e atividades já concluídas (BORGES et al., 2004).

Kraut e colegas (1990b) definem “interações não planejadas” como interações que ocorrem sem planejamento prévio por qualquer participante. Estas interações ocorrem durante encontros semi-espontâneos (ex., encontros no corredor ou na pausa para o cafezinho), porque há algum contexto compartilhado entre as partes ou uma possibilidade para intercâmbio de informações. A colaboração é baseada em interações entre as pessoas. Kraut (1990b) define quatro categorias de interação, descritas abaixo:

- Interações *planejadas* são programadas, reuniões formais.
- Interações *intencionais* acontecem quando uma pessoa explicitamente procura outras.
- Interações *oportunistas* acontecem quando uma parte pretende conversar com outra, mas se encontram por acaso.
- Interações *espontâneas* não são antecipadas por qualquer das partes e acontecem completamente por acaso.

Interações intencionais, oportunistas e espontâneas são consideradas não planejadas, mas acontecem diariamente quando um grupo compartilha a mesma área de trabalho. Estas interações acontecem mais frequentemente que as planejadas, e contribuem para aumentar a consciência do que acontece na área de trabalho e sobre o que outros estão trabalhando (KRAUT et al, 1990a; KRAUT et al, 1990b). Este conhecimento gera uma maior compreensão do contexto em que as pessoas estão inseridas, aumentando as possibilidades de novos projetos colaborativos entre pares (MCEWAN, GREENBERG, 2005). No entanto, se o grupo não compartilha a mesma área de trabalho, estas interações informais param de acontecer e as possibilidades de colaborações novas, espontâneas acontecerem também caem.

Em ambientes co-localizados, informações para manutenção da consciência são reunidas principalmente através de (GUTWIN, GREENBERG, 2004):

- Comunicação intencional, i.e., comunicação pretendida pelo remetente, tal como conversa e gesticulação;
- Comunicação conseqüente, ou transferência de informação que acontece em conseqüência da atividade do indivíduo dentro do ambiente, obtido pela observação de ações de outros ou posições corporais; e
- Feedthrough, que é o mecanismo de determinar as ações da pessoa através de pistas deixadas nos artefatos com os quais eles interagem, tal como posição, orientação ou movimento.

Em colaboração distribuída, comunicação intencional é freqüentemente o único meio de manter consciência do ambiente, o que cria problemas significativos, como a necessidade de lembrar de informar aos outros sobre acontecimentos importantes e enviar mensagens. Como envolve conhecimento sobre um ambiente dinâmico, a consciência deve ser mantida atualizada com as mudanças no ambiente. Ela é mantida através da percepção da informação reunida do ambiente (através de interações com o ambiente); e é geralmente secundária a outras metas. Enquanto permanecer ciente de outros não é complexo em interações cotidianas, torna-se complicado em sistemas distribuídos, onde comunicação e recursos de interação são pobres (GUTWIN, GREENBERG, 2002).

Estudos etnográficos mostraram que essa consciência permite aos membros de grupo administrar o processo de trabalhar junto, e é necessário para coordenação de atividades de grupo (DOURISH, BELLOTTI, 1992). Estar ciente de atividades dos outros numa área de trabalho permite participantes entender melhor os limites de suas ações, que por sua vez os ajuda a localizar suas próprias ações no fluxo de atividade colaborativa. Isto também permite aos grupos melhor administrar os níveis de acoplamento entre suas atividades, ajudando indivíduos a decidirem com quem eles necessitam trabalhar e quando fazer as transições de trabalho mais solto a mais forte (HEATH, LUFF, 1991), e decidindo quando estabelecer contato (GREENBERG, JOHNSON, 1997). É esta consciência do espaço de trabalho que permite a indivíduos

em situações fracamente estruturadas ajustar seu curso de ação para compensar acontecimentos imprevistos.

Além disto, a consciência simplifica a comunicação por permitir que indivíduos se refiram ao ambiente compartilhado e aos elementos dentro dele: quando discutem artefatos compartilhados, a área de trabalho pode ser usada como uma base para comunicação (BRINCK, GOMEZ, 1992). Isto faz da consciência uma peça importante na construção de cognição de equipe (GUTWIN, GREENBERG, 2004), bem como na criação e manutenção de um contexto compartilhado. Ela também viabiliza o entendimento compartilhado que permite indivíduos compor um melhor sensação do trabalho que está sendo executado por outros (GUTWIN et al, 2005), além de contribuir para a integração do grupo (NARINE et al, 1997).

2.5 Atenção e Foco do Usuário

A definição de consciência de Rodden (1996) se refere a outro conceito importante, o de atenção. Uma das grandes dificuldades que indivíduos encaram hoje em dia é lidar com a crescente quantidade de informação a qual eles são expostos diariamente. Decidir o que deve receber atenção tornou-se uma tarefa freqüente em ambientes de trabalhadores de informação. Como informação consome atenção, o aumento na informação disponível cria uma exigência de atenção que é difícil de suprir (VERTEGAAL, 2003). Em anos recentes, pesquisadores verificaram que a atenção está escassa em organizações e indivíduos.

Cada email enviado é um pedido de atenção, que deve ser considerado pelo destinatário. Neste cenário, cada indivíduo funciona como um fornecedor de informação, tentando de atrair atenção a seu ou seus memorandos, correios eletrônicos, projetos, etc (DAVENPORT, BECK, 2001). Controlar a atenção torna-se então uma parte importante da vida cotidiana, e um determinante de produtividade, uma vez que a informação alcança consciência só quando atenção lhe é dedicada (CSIKSZENTMIHALYI, 1997, apud DAVENPORT, BECK, 2001).

A atenção é importante para a tomada de decisão, visto que indivíduos necessitam refletir sobre problemas para tomar boas decisões. A atenção é foco mental num item particular de informação (DAVENPORT, BECK, 2001). Ela ocorre entre uma fase de estreitamento, onde muito da entrada sensorial é descartada, e uma fase de decisão, em que o indivíduo decide agir na informação. Há um relacionamento causal

entre consciência, atenção e ação, como visto na Figura 2-4. A atenção é um processo seletivo, cognitivo, através da qual os indivíduos absorvem informação selecionada.

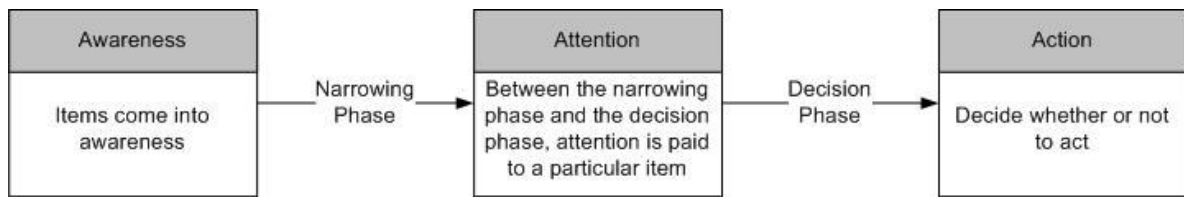


Figura 2-4: Relacionamento entre consciência, atenção e ação (DAVENPORT, BECK, 2001, p. 21)

Atribuição de atenção apropriada é fundamental ao indivíduo, e uma pessoa normalmente escolherá entre as muitas demandas de atenção de acordo com os desfechos esperados (ex, aprender algo novo, consertar algo que esteja errado ou reforçar um relacionamento). Esta atenção, uma vez gasta, não volta, então é razoável supor que indivíduos priorizarão sua atribuição de atenção de acordo com metas e contratos estabelecidos com outros. Em configurações fracamente estruturadas, trabalhadores frequentemente alternam entre várias colaborações, dividindo seu tempo e atenção de acordo com a urgência das tarefas (GONZALÉS, MARK, 2005).

Os dois principais problemas encontrados quando trabalhando em grupos reconfiguráveis são *lembrar-se* das pessoas e atividades e *comunicar-se* com outros (NARDI, 2002). Um problema que se repete é a administração da atenção e se lembrar de que arranjos foram feitos e que tarefas estão vencendo em ordem “não deixar nada cair a peteca” (GONZALÉS, MARK, 2005). Uma *Esfera de Trabalho* agrupa um conjunto de acontecimentos envolvendo um grupo particular de pessoas, um objetivo comum e um conjunto único de recursos (GONZALÉS, MARK, 2004).

Sistemas de apoio a trabalhadores envolvidos em múltiplas equipes virtuais devem ajudar os indivíduos a administrar suas múltiplas esferas de trabalho. Um meio de fazer isto é ajudá-los a entender o contexto social do trabalho que eles executam: quem está envolvido com este projeto e em que capacidade? Que outros recursos são relevantes? A determinação de colaboradores e identificação de temas de colaboração é um dos passos em direção dessa metas. Uma *Esfera de Colaboração* não é composta só de um conjunto de pessoas, recursos e ferramentas, mas também das relações mútuas entre eles, que afeta o trabalho em andamento ao definir como pessoas se relacionam, se organizam e realizam seu trabalho. O contexto do grupo, informalmente construído e dinâmico, contém hierarquias, papéis e designações de tarefa. A descoberta destes

contextos é um primeiro passo em direção à criação de sistemas que apoiem trabalhadores envolvidos neles.

Os trabalhadores de informação tipicamente são envolvidos em múltiplas colaborações ao mesmo tempo, e devem dividir seu tempo (GONZALÉS, MARK, 2005). Deste modo é possível que várias subseções diferentes de uma rede intensional estejam ativas simultaneamente, e que o usuário deva alternar sua atenção entre as diferentes esferas de trabalho de acordo com as necessidades. Uma das dificuldades neste cenário é lembrar-se dos diferentes contextos e administrar as transições entre eles. Trabalhadores se valem de uma variedade de métodos para lembrar-se de coisas, pessoas e do contexto atual, incluindo visitas a colaboradores, verificação constante de email, mantendo listas ou anotações para si próprio (GONZALÉS, MARK, 2005).

A atenção é dedicada às pessoas e projetos nos quais um usuário tem maior interesse (ou investimento mais alto). A análise de interação extrai as prioridades do operador e capacita um sistema a manter um perfil atualizado de colaboração. A atribuição de atenção entre pessoas fornece as prioridades em relação à colaboração continuada. Se cada email recebido é um pedido de atenção, cada email respondido é uma ação clara sobre a informação contida no email anterior. Esta dinâmica de envio e recebimento dura enquanto houver uma meta em conjunto a ser alcançada ou um foco comum de atenção. Interações continuadas mostram compromisso em conjunto com um projeto colaborativo.

Assim, uma análise de interações deve viabilizar a descoberta de como a atenção é atribuída entre os muitos conhecidos do usuário. Pesquisas prévias indicaram que uma análise histórica é capaz de extrair padrões que espelham a mudança de interesses de uma pessoa (PERER et al, 2005). O mesmo estudo indicou uma intensificação em interações durante trabalho colaborativo.

2.6 Estrutura de Locais (Locales Framework)

A Framework Locales (FITZPATRICK, 1998) fornece um conjunto de abstrações para apoiar o projeto e análise de trabalho colaborativo. Ele procura ligar o perspectivas sociais e técnicas, e foi concebido como uma ferramenta para análise e projeto de sistemas de apoio ao trabalho colaborativo. É baseado na Teoria de Ação de Anselm Strauss (STRAUSS, 1993 apud FITZPATRICK et al., 1995), de acordo com a qual, ação individual sempre é embutida em interações, que são executados por um ou

mais interagentes. É baseado principalmente na noção de ação continuamente desenvolvida e de *Mundos Sociais*. De acordo com a Teoria de Ação, Ação acontece no contexto de condições estruturais, que facilitam ou impedem o curso das interações, e que podem mudar com tempo.

Mundos sociais são grupos de pessoas que compartilham um compromisso com ação coletiva, e formam a principal mecanismo estrutural para interação (STRAUSS, 1993 apud FITZPATRICK et al., 1995). Um mundo social é uma unidade interativa que se ativa quando um número de indivíduos decide agir coletivamente (que frequentemente exige coordenação de perspectivas separadas e compartilhamento de recursos). Estes indivíduos executam trabalho coletivo em locais (sites) e empregam meio (means) para fazê-lo. Os locais são os espaços (e.g. sistemas de arquivo compartilhados) e meios são objetos contidos nestes espaços (e.g. os arquivos e documentos armazenados neste sistema de arquivo). O papel do sistema em muitos casos é único fornecer um cenário, mas não estruturar colaboração.

Estas observações estão em consonância com as observações de Suchman (1996) de como trabalhadores de co-localizado usam seu ambiente disponível como um cenário para seu trabalho, e se reorganizam no espaço de acordo com as necessidades. Seu estudo de um grupo de operadores de torre de controle ao lidar com um incidente aéreo mostra como indivíduos passam para colaboração mais próxima quando necessário, fisicamente se movendo para criar grupos diferentes ou virando para mudar o foco. Os participantes se empenharam em resolver um problema em conjunto devido à sua consciência parcial da situação compartilhada, e o grupo organizou-se como necessário. Enquanto o incidente ocorreu num período curto de tempo, o estudo mostra como, através da manipulação do espaço compartilhado e mobilização de recursos, o grupo podia passar de atividades individuais a um curso conjunto de ação. A área de trabalho compartilhada foi construída colaborativamente pelos membros do grupo.

Um *Locale* é um conceito abstrato que surge do uso de espaço e recursos por um grupo. Ele traça o relacionamento entre um Mundo Social (e suas necessidades de interação) e os *locais* e *meios* que seus membros usam para suprir essas necessidades. (FITZPATRICK et al., 1998).

Dado que indivíduos normalmente estão envolvidos em múltiplos mundos sociais num tempo, mundos sociais diferentes estão interligados pelos indivíduos que os

compõem, e ações em um mundo social podem refletir em outro. Além disso, cada indivíduo tipicamente está envolvido em múltiplas tarefas ou atividades, em mais de um mundo social. Isto está de acordo com observações por Gonzalés e Mark (2005), que descrevem os processos envolvidos na alternância entre várias colaborações.

Uma análise automatizada dos Locales usados por um grupo deve resultar em padrões que refletem o trabalho do grupo. Que mundos sociais estão ativos, que priorização está sendo usada entre eles e como se ligam um ao outro.

2.7 Rede Social e Análise de Interação

Associações interpessoais são os elementos através dos quais indivíduos formam grupos e as sociedades (VANDENBERGHE, 2001). George Simmel diz que interações entre indivíduos são o elemento mais básico de grupos, servindo simultaneamente como o meio de associação e de socialização entre membros (ibid, p.50), e portanto devem ser tomados como os objetos de estudo.

Simmel estava mais preocupado com as formas que com o conteúdo da interação, dirigindo seu trabalho para uma ciência das estruturas de relações sociais (MERCKLÉ, 2004). É possível estudar relações sociais porque estas são regulares e estáveis, e figuram em diferentes partes da vida social: relacionamentos típicos, tal como dominação, competição, imitação ou conflito podem ser encontrados através do estudo destas estruturas. O trabalho de Simmel inspirou o desenvolvimento de Análise de Redes Sociais (SCOTT, 1991; VANDENBERGHE, 2001).

J L. Moreno desenvolveu sociometria como um meio de desenhar laços entre indivíduos e estabelecer sua força. É um instrumento de estudo de estruturas sociais à luz de atrações e repulsões que se manifestam dentro de um grupo (MORENO 1954, apud MERCKLÉ, 2004). O sociograma resultante é, ao mesmo tempo, um meio de exposição e exploração do grupo, que permite a determinação da posição que cada membro ocupa ou as relações de aceitação ou rejeição estabelecidas entre indivíduos.

Análise de redes sociais (Social Network Analysis – SNA) avalia os relacionamentos entre indivíduos e suas regularidades ao analisar comportamento individual e de grupo, trazendo para o comportamento individual a complexidade ou os sistemas sociais dentro dos quais eles operam (MERCKLÉ, 2004). Isto reflete o fato de

que a maioria das atividades estão inseridas em contextos sociais, e que as redes nas quais eles estão inseridos é limitada por regras e práticas definidas por seus membros.

Análise de redes sociais é usada freqüentemente em ciências sociais e comportamentais, assim como em economia. Lida com o estudo de entidades sociais e seus relacionamentos: comunicação entre indivíduos, comércio entre negócios ou tratados entre as nações. Os elementos básicos de uma rede são nós e os laços que os unem. Num contexto de rede social, nós são atores individuais, e os laços são relacionamentos entre estes indivíduos, que podem ser hierárquicos, familiares, de amizade, de trabalho, etc. (SCOTT, 1991). SNA considera estruturas tal como o sociograma, um gráfico que representa indivíduos e as relações entre eles (WASSERMAN, FAUST, 1994). Estas relações podem ser de natureza diversa (comunicação, presença de partido, difusão de informação, etc.), e normalmente são expressas como gráficos e matrizes (Figura 2-5); sobre as quais a análise pode ser executada (SCOTT, 1991). SNA usa teoria de gráfico e técnicas de álgebra linear para determinar propriedades estruturais dos indivíduos dentro de cada rede. Assim, atributos de rede tal como centralidade (quantos laços ligam a um certo nó (FREEMAN, 1979)) são ligados a uma interpretação sociológica (no caso, importância no grupo). Analistas de rede social olham o mundo em termos de padrões ou regularidades em relacionamentos entre atores.

-	A	B	C	D	E	F	G	H	I
A	-	1	0	1	0	0	0	0	0
B	1	-	0	1	0	0	0	0	0
C	0	0	-	1	0	0	0	0	0
D	0	1	0	-	0	0	1	1	1
E	0	0	0	0	-	0	0	0	1
F	0	0	1	1	0	-	1	0	0
G	0	0	0	0	0	1	-	0	0
H	0	0	0	1	0	0	0	-	0
I	0	0	0	0	0	0	0	0	-

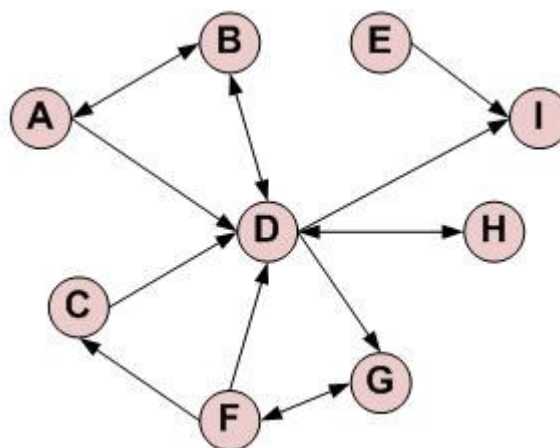


Figura 2-5: Exemplo de sociograma e tabela correspondente de ligações entre elementos.

Como o ator participa em grupos diferentes (GONZALÉS, MARK, 2005; SIMMEL, 1955), ele ou constrói uma rede pessoal ao redor dele, que nunca é exatamente igual de qualquer outro (WELLMAN, GULIA, 1999). Os laços que unem pares de atores são construídos individualmente, à medida que relacionamentos progredem (GABARRO, 1990). Para estabelecer novos laços com outros, atores

comumente se valem de parte da rede que os cerca (FERRAND, 1997, apud MERCKLÉ, 2004).

A rede do ator está sujeita à sua interpretação e entendimento, e afetará a maneira na qual ele usa ou navega pela rede. A extensão de uma rede é variável, com alguns pesquisadores dizendo que toda sociedade está emaranhada num conjunto de redes de indivíduos e grupos (LÉVI-STRAUSS, 1952, apud MERCKLE, 2004). Pesquisas também mostraram a importância de redes pessoais, e como laços fortes e fracos têm papel importante numa vida do indivíduo, como são usados para achar trabalho e formar novas associações (GRANOVETTER, 1973). Mais tarde estudos de documentaram a existência e importância destas redes em situações grandes de escala, tal como no Vale de Silício (SAXENIAN, 1994; CASTILLA et al., 2000), onde membros se encontram informalmente, trocam informação, discutem idéias e mudam de companhia frequentemente, criando um ambiente inovador.

Dois tipos de análises são possíveis: enquanto a análise de rede *sociocêntrica* olha estruturas de relacionamento de uma perspectiva global (e.g., um gráfico de comunicação entre todos membros de um departamento ou grupo), a análise de rede *egocêntrica*, por outro lado, focaliza no indivíduo (*ego*), e analisa suas relações com um conjunto de outros (*alters*) (GARTON et al., 1999). Este tipo de rede foi usado para estudar o ambiente social que cerca indivíduos como as famílias e estruturas sociais de apoio (WASSERMAN, FAUST, 1994).

Em grupos distribuídos, indivíduos normalmente interagem eletronicamente, e interações eletrônicas deixam vestígios, tal como email, fora ou mensageiro registra. Estas interações exibem certos ritmos que correspondem a padrões de trabalho do indivíduo (PERER et al., 2005), e podem ser usados para, por exemplo, estudar a evolução dos interesses do indivíduo (e de como ele se relaciona a outros em momentos diferentes) ou esforços colaborativos: por exemplo, intensa troca de mensagens normalmente acompanha trabalho cooperativo. Padrões individuais de troca de email também podem indicar hierarquia e posicionando num grupo (FISHER, 2004).

Na maioria dos casos, redes sociais e análise de interação são usadas apenas como ferramentas de análise. Nesta tese, procuramos aplicar rede social e análise de interação como elemento ativo, para viabilizar o raciocínio por uma entidade inteligente e adaptação de sistema. Construimos uma rede egocêntrica baseada em registros de

comunicação eletrônica. Esta rede é navegada para descobrir colaboração em andamento e informações relacionadas a projetos em conjunto.

2.8 Busca e Recuperação de Informação

Em geral, técnicas de recuperação de informações representam documentos como vetores de palavras-chave, calculados com base em seu conteúdo (BAEZA-YATES, RIBEIRO-NETO, 1999). Estas palavras-chave são extraídas de corpos de documentos e “pesadas” de acordo com sua frequência em cada documento. Este processo de extração pode ser refinado através da remoção de stopwords (palavras que tem frequência alta mas adicionam pouco aos documentos, tal como “o”, “um”, “os”, etc.) e *stemming* (um processo através do qual sufixos são retirados, ficando apenas os radicais para serem processados como palavras-chave). A relevância de cada palavra (para definir o conteúdo de cada documento) é calculada de acordo com sua frequência no documento e sua frequência em todos os documentos, usando a fórmula TFiDF (Term Frequency, Inverse Document Frequency) (SALTON, MCGILL, 1997): a relevância de uma palavra em um documento particular é igual à sua frequência no documento dividido por sua frequência no conjunto de documentos. Assim, palavras que aparecem muito frequentemente no conjunto de documentos perdem seu poder de distinção. Documentos são associados a vetores contendo as palavras-chave que melhor os representam.

A comparação entre documentos é realizada através de uma comparação entre os vetores de termos que os representam e cálculo de sua similaridade. A medida de cosseno é frequentemente utilizada, onde o cosseno do ângulo entre dois vetores é calculado e, quanto mais próximos os vetores, mais semelhantes os documentos (BAEZA-YATES, RIBEIRO-NETO, 1999).

2.9 Esferas de Colaboração como Pontes entre Membros de Grupos

Viabilizar a auto-coordenação e o trabalho de grupo em cenários fracamente estruturados, envolve a disponibilização de informação aos sobre o trabalho em andamento. Esta informação diz respeito a acontecimentos relacionados às pessoas e recursos envolvido em cada cenário de trabalho.

Visto que as redes intensionais e esferas de trabalho representam pontos de vista individuais sobre as colaborações, nós deixamos de lado as abordagens sociocentricas e

colocamos o foco no indivíduo e suas relações com outros. Adotamos uma abordagem egocêntrica para auxiliar a gerência de múltiplas tarefas e manutenção de consciência. Enquanto o indivíduo trabalha só, em uma atividade pertencente a uma esfera de trabalho, outras esferas de trabalho podem estar incorrendo ação de outros membros. Esta informação deve ser detectada e disseminada para capacitar uma transição mais fácil entre esferas de trabalho. Baseado neste ponto de vista sociológico, nós adotamos uma arquitetura baseada em pares para prover flexibilidade máxima (indivíduos podem se ligar a outros como desejado) e usamos tecnologia de agentes para criar um sistema pró-ativo que acompanha atividades e colaborações.

Em cenários fracamente estruturados tal como os descritos, indivíduos freqüentemente trabalham sem apoio centralizado, usando email e outros meios de comunicação eletrônicos de comunicação para se coordenar e trocar informação e tarefas (NARDI et al., 2002). Grupos diferentes podem selecionar ferramentas diferentes ou ter recursos diferentes disponível a eles (GONZALÉS, MARK, 2005). Determinar colaboradores, recursos e informação trocada, numa situação distribuída onde uma variedade de ferramentas são usadas, é um problema complexo mas necessário para facilitar a colaboração.

Visto que subredes se tornam ativo por comunicação (NARDI et al, 2002) e trabalho freqüentemente é empreendido por trocas lingüísticas (ZACKLAD, 2003), propomos um método a para a determinação de colaboradores (esferas de trabalho ativas) e de atividades que pertence a diferentes esferas de trabalho baseadas em análise de registros de interação e conteúdo. Com isto, esperamos determinar grupos ativos de colaboradores em uma rede com o que sua atividade lida. Estes então podem ser combinados à atividade individual para descobrir que atividades pertencem a que esferas de trabalho, e acompanhar a evolução da área de trabalho conjunto.

Criar uma vista unificada do grupo, preenchendo a lacuna entre trabalho individual e em conjunto, e múltiplos laços unindo esferas de trabalho, introduzimos o conceito de esferas de colaboração. Estas agrupam pessoas e recursos, junto com elementos contextuais relacionado ao trabalho em conjunto. Os acontecimentos que ocorrem dentro de uma esfera de colaboração são pertinentes a membros desta esfera e podem ser transmitido a outros membros. Nos próximos capítulos, nós elaboramos sobre este conceito e apresentamos um método para constituir esferas de colaboração.

Awareness est une compréhension des activités des autres, qui fournit un contexte pour l'activité de soi même (DOURISH, BELLOTTI, 1992). Elle concerne information sur les autres, leurs tâches et activités et l'environnement de travail en général. Awareness est obtenu atravers la perception d'informations récoltes de l'environnement et doit être maintenu lors des changes sure l'environnement (GUTWIN, GREENBERG, 2002). La manutention du awareness est un problème dans environnements computationnelles, parce-que les ressources de communication et interaction sont plus pauvres que ceux des interactions face a face. En plus, la distance introduit plus de difficultés pour la manutention des informations sur l'environnement partagé. Dans cette thèse nous proposons une méthode pour résoudre ce problème avec la création des sphères de coopération.

3 Trabalhos Correlatos

Ce chapitre présente une révision de systèmes pour l'organisation et distribution d'informations. Y compris systèmes de recommandations, systèmes de conscience et recherche sur personnalisation.

Esta seção apresenta uma revisão de sistemas que lidam com organização e disseminação de informações. Isto inclui sistemas de recomendação, sistemas de apoio à consciência e sistemas de redes sociais, bem como trabalhos em personalização.

3.1 Organização do Trabalho

Muitos sistemas baseados em redes sociais têm como meta a localização de especialistas. ReferralWeb (KAUTZ et al, 1997) é um protótipo para localizar especialistas dentro de uma companhia, tomando como base referências fornecidas por usuários sobre quem eles chamariam para falar sobre determinados temas. Agentes seguem estes fluxos de recomendação para encontrar usuários que possam ajudar com problemas. O sistema também processa arquivos dos usuários, extraindo palavras-chave das mensagens trocadas entre eles, para construir a visão do usuário de cada um dos seus contatos. McArthur e Bruza (2003) apresentam um método semelhante para construir redes de pessoas e palavras-chave a partir de dados de email. Groth (2003) utiliza redes sociais para localizar indivíduos com determinadas habilidades e disponibilidade através de uma análise de suas atividades e tarefas. Nas abordagens supracitadas, a ênfase está na busca por especialistas, e em navegar a rede social para descobrir quem sabe o que.

ContactMap (NARDI et al, 2002) é um sistema que permite aos usuários visualizar e organizar suas redes de contatos. É um passo no sentido da construção de sistemas de apoio ao trabalho através de redes intensionais, e foi construído para facilitar gerência de redes pessoais individuais. ContactMap tem como objetivo ajudar as pessoas a lembrar-se de outros em suas redes, de tarefas em que eles estão envolvidos, das conexões entre diferentes conhecidos, de trocas de documentos e do

estado e prazos do trabalho.. O sistema utiliza análise de emails para construir uma rede social pessoal, onde o usuário pode escolher que contatos exibir. O usuário então pode expor contatos, agrupando-os como apropriado. É uma ferramenta de organização que ajuda os usuários a se lembrar de certos contatos e os agrupa de acordo com suas relações mútuas. ContactMap é um sistema de gerencia pessoal de contatos, que viabiliza a manutenção de visão atualizada de suas redes sociais pessoais, baseado em registros de email.

Vizster (HEER, BOYD, 2005) permite aos usuários visualizar e acessar informação contida em suas redes sociais pessoais. A visualização é rica em informação sobre interesses do usuário, extraídas do Friendster (www.friendster.com), um sistema de redes sociais. Este sistema é especialmente útil para a exploração de uma rede social, pois permite agrupar membros e descobrir interesses e atividades, viabilizando a navegação na rede.

Trabalhar através de ligações envolve mais que saber quem está na rede e quem pode fazer o que. CommunityBar (MCEWAN, GREENBERG, 2005) foi projetado seguindo o Locales Framework (FITZPATRICK, 1998). É um sistema que permite aos usuários criarem ligações com outros através da seleção de diferentes “mídias” para comunicação (notas, janelas de chat, webcam, etc.). O sistema fornece informação de identidade, presença e alguma informação de atividade. Os usuários selecionam os canais que querem acompanhar, e o sistema fornece informação passada pelos canais selecionados. Os usuários recebem quaisquer informações que passem pelo canal. Os contatos e itens de mídia podem ser agrupados representar trabalho em diferentes grupos. Foco e nimbus são estabelecidos manualmente (o usuário pode determinar quanto ele quer receber por cada canal particular, e quanto ele permite que os outros vejam por cada canal). O ambiente fornece uma maneira fácil de estabelecer interações, uma vez que conversas podem ser iniciadas com qualquer participante.

ActivityExplorer (MULLER et al, 2004; GEYER et al, 2006) foi concebido para preencher a lacuna entre aplicações formais e estruturadas e ferramentas colaborativas informais, de fácil inicialização de colaboração. O sistema apóia atividades colaborativas informalmente estruturadas ou oportunistas através do controle de objetos compartilhados. O sistema viabiliza o compartilhamento de recursos com granularidade fina. Cada objeto compartilhado contém informação persistente e define uma lista das pessoas que têm acesso ao seu conteúdo (e.g., arquivos, mensagens, logs de conversas,

etc.). Cada objeto compartilhado gera notificações a seus usuários de qualquer atividade realizada nele. Isto fornece consciência de alta granularidade de quem está trabalhando nele, o que pode servir como um gatilho para colaboração oportunista. Para organizar a colaboração, usuários combinam e agregam manualmente objetos compartilhados heterogêneos em coleções estruturadas. O Activity Explorer fornece uma visão integrada de atividades em andamento e o que está relacionado a elas. As entrevistas com usuários mostram que o sistema é útil como uma ferramenta de coordenação e para melhor fazer sentido do que o usuário e outros faziam. Usuários indicaram que substituíram email e mensageiro pelo Activity Explorer, que integrou características adicionais mas ainda viabilizando o envio de email e início de conversas.

3.2 Sistemas de Recomendação

Os sistemas de recomendação ajudam e apóiam processos sociais através de algoritmos de filtragem de informação. Um sistema típico de recomendação recebe e agrega dados e dirige os resultados às pessoas corretas (Resnick, Varian, 1997). Estes resultados ajudam as pessoas a tomar decisões, baseado na informação fornecido.

Maes (1994) propõe o uso de agentes para apoiar cooperação. Ela descreve um sistema que permite aos usuários encontrar assuntos que talvez sejam de interesse a outros usuários através da análise de perfis de usuário, os agentes procuram ainda temas que podem ser de interesse a ambas as partes. Foner (1997) apresenta um sistema multiagentes para unir pessoas que compartilham interesses. Para alcançar este fim, o sistema faz uma análise dos documentos do usuário (emails, arquivos de texto, artigos, etc., construindo perfis de usuário baseados em palavras-chave e combinando estes com outros perfis do usuário.

Duas abordagens diferentes podem ser usadas para ferramentas de multiagentes e matchmaking. O primeiro é ter um servidor central encarregado do matchmaking, tal como usado em Webhound e Webdoggie (SHARDANAND, MAES, 1995) e HOMR, Ringo ou Firefly (LASHKARI et al, 1994). Estes sistemas constroem perfis através da observação das ações dos usuários, e salvam conteúdo de acordo com seus temas (ex, salvando email sobre algoritmos genéticos numa pasta de Inteligência Artificial).

A outra abordagem é distribuir o sistema: Kautz, Milewski e Selman (1995) descrevem um protótipo para localizar especialistas em uma companhia com base em referências fornecidas por usuários sobre outros que eles talvez chamassem para discutir

certos temas. Os agentes seguem esta corrente de recomendações para achar usuários que podem ajudar com um problema. Neste sistema, usuários devem fornecer explicitamente ao sistema os dados que este necessita para trabalhar. Para cada usuário, o sistema mantém uma lista de contatos e uma lista de áreas de especialidade para estes contatos. Cada modelo de usuário é construído usando técnicas padrão de vetores de palavras-chave nos arquivos do usuário, enquanto que o perfil de cada contato é construído usando a mesma técnica em mensagens enviadas entre o usuário e seus contatos.

Payton e colegas (1999) descrevem um sistema onde os interesses dos usuários são inferidos a partir do seu acesso à informação. Estes interesses se acumulam com acessos repetidos e decaem em proporções variáveis. Os usuários podem usar uma interface especial, que exhibe os indivíduos como nós numa rede social para alocalizar colaboradores em potencial. As recomendações são feitas com base no histórico de acesso à informação.

3.3 Sistemas de Consciência

A manutenção de consciência é facilitada pela proximidade física, uma vez que é possível perceber uma quantidade grande de informação simplesmente andando pelo escritório, ouvindo conversas de outros por acaso ou entrando em breves conversas de corredor (Kraut et al., 2002).

Fornecimento de informações relacionadas à consciência em ambientes computacionais de colaboração não é tarefa simples, especialmente quando os participantes estão distantes de si. Dispositivos de entrada e saída geram menos informação que as situações face a face e as interações dos usuários com áreas de trabalho computacionais também geram informação menos rica que a gerada em um ambiente físico (Gutwin e Greenberg, 2004).

Ferramentas de apoio ao trabalho em grupo devem lidar com a diminuição na consciência causada pela distância entre participantes e pela mudança para meios de comunicação mediados por computador. Uma maneira de suprir esta necessidade é através do fornecimento de informação sobre o trabalho de outras pessoas e a identificação de possíveis oportunidades de colaboração. Diversas aplicações foram projetadas para fornecer informação de consciência em aplicações de colaboração eletrônica.

Os primeiros trabalhos em consciência foram voltados para o uso de vídeo para apoiar consciência pessoal e interações informais: CRUISER (Root, 1988) era um ambiente virtual usando áudio e canais de vídeo para apoiar interação social informal e visão social. A possibilidade de facilmente visualizar pessoas empenhando esforços semelhantes gerou interações espontâneas que não teriam acontecido sem o sistema. Pelo sistema, usuários eram capazes de explorar o ambiente de trabalho virtual, fortalecendo seus relacionamentos com colegas.

Polyscope e Portholes (DOURISH, BELLOTTI, 1992) são sistemas que usam uma visão compartilhada da área de trabalho, para tornar usuários mais cientes do ambiente através de meios de comunicação compartilhados. Polyscope captura imagens da área de trabalho e as apresenta numa tela no computador do usuário. Estas imagens mostram atividade em áreas públicas e escritórios. Portholes foi continuação de Polyscope, estendendo os meios de comunicação compartilhados para cobrir uma maior distância e ligar a Xerox PARC nos EUA à EuroPARC.

Portholes (DOURISH, BELLOTTI, 1992) foi um sistema projetado para prover consciência entre locais remotos. Portholes capacitava visualização informal de outros, tal como visões noturnas, pedaços de canções sendo tocadas em locais remotos ou indivíduos trabalhando até tarde. A disseminação desta informação criou um sentido de consciência do grupo de trabalho como um total, fortalecendo a sensação de comunidade entre indivíduos, que se identificavam com seus colegas remotos. Os principais problemas foram falhas de imagens (devido a problemas tecnológicos) e o fato de que acontecimentos levavam um tempo para serem mostrados: a transmissão lenta de informação levou à perda de motivação para usar o sistema. Estes sistemas também ajudaram as pessoas a identificar o momento correto para iniciar uma conversa por permitir verificação de disponibilidade das pessoas, um fator importante em ambientes não co-localizados.

A maioria das ferramentas de áudio e de vídeo tem uma limitação séria: embora muitas destas ferramentas forneçam aos usuários informação suficiente para levar à interação, a maioria deles não oferecem meios eficientes de passar da visão à interação, o que compromete a capacidade dos usuários de tirar proveito de oportunidades que possam surgir. Adicionalmente, estas ferramentas não levam em conta o conteúdo da informação, distribuindo indiscriminadamente o que passa pelo canal de comunicação.

Pesquisas posteriores se focaram no fornecimento de informação sobre outros usuários executando tarefas semelhantes no mesmo momento: Piazza (ISAACS et al, 1996), por exemplo, fornecia informações de consciência sobre outros trabalhando em tarefas semelhantes em seus computadores, expondo oportunidades para interação ou cooperação. Piazza permitia a um usuário, enquanto trabalhava nas suas tarefas, ver o que outros usuários estavam fazendo. Sua medida de semelhança era baseada em três dimensões: dados sendo manipulados, quando eram manipulados e por que aplicação. Pessoas usando os mesmos dados em aplicações diferentes eram consideradas próximas e eram exibidas aos outros.

Outras abordagens tomam como base um único objeto do trabalho colaborativo, tal como um documento que esteja sendo redigido por várias pessoas simultaneamente. GROVE, PREP e QUILT (DOURISH, BELLOTTI, 1992) seguem esta abordagem: estes sistemas funcionam como editores de texto, mas tem funcionalidades de comunicação e sincronização adicionais embutidas. Estas permitem aos usuários se manterem cientes, de que o documento está sendo alterado.

PIÑAS (MORAN et al, 2002) é um arcabouço para apoiar colaboração potencial e real através do agrupamento de pessoas e recursos. Para atingir este fim, PIÑAS define espaços colaborativos (denominados PICoS) para integrar e estruturar serviços correlatos. Os serviços permitem aos usuários se agrupar em um PICoS compartilhado com outros usuários e recursos, que recebem informações periféricas sobre outros usuários e recursos pertencentes ao mesmo PICoS, através do envio de mensagens. Os usuários devem se logar em um ambiente compartilhado para ter acesso aos serviços, embora a arquitetura seja descentralizada, ligando usuários distribuídos através de um middleware. Os usuários precisam entrar no PICoS relevante e fazer uma “assinatura” dos ambientes sobre os quais eles queiram saber. O sistema acompanha as atividades ocorrendo nestes espaços compartilhados e informa ao usuário sobre estas através de uma interface de messenger.

CommunityBar (MCEWAN, GREENBERG, 2005), descrito anteriormente, é uma ferramenta que exhibe imagens (fotos ou vídeos) e nomes das pessoas com quem o usuário está trabalhando e fornece perfis manualmente construídos com informação de contato. Um usuário pode permanecer consciente do que seus colegas estão fazendo e iniciar uma interação se necessário.

O sistema CSCW3 (GROSS, 1999) adiciona uma série de funcionalidades colaborativas a um navegador web, possibilitando que usuários vejam quem está visitando a mesma página ou o fez previamente, passando à interação (via interface de chat) quando desejado e trocando informações em diferentes formas (bookmarks ou históricos de visita). Usuários informaram que gostaram de ver que mais estava na mesma página e que raramente fechavam a visualização quando navegando.

O ambiente Nessie fornece informação baseada em perfis de interesse. Ele distribui informação sobre outros usuários que estão logados e sobre eventos que aconteceram desde seu último login que vão de encontro com seus perfis de interesse. Estes perfis contêm uma lista de eventos relevantes para cada usuário, e devem ser manualmente especificados pelo usuário.

Elvin (FITZPATRICK et al, 2002) é um sistema de notificação baseado em eventos desenvolvido para apoiar diferentes aplicações de distribuição de informação, que não foi originalmente não projetada para apoiar colaboração. No entanto, foi aplicada ao apoio à consciência e interação. Elvin é um distribuidor de eventos concernentes a objetos observados. Os produtores de eventos enviam informações a Elvin, que as compara com tabelas de assinatura contendo regras criadas por consumidores. Estas regras determinam para quem cada evento será propagado. Um usuário pode “subscrever” a eventos acontecendo a outros usuários ou artefatos de sua colaboração compartilhada. No entanto, caso estas preferências mudem, nova configuração será necessária.

Entre as aplicações cliente desenvolvidas, tickertape e CoffeeBiff permitem que usuários deixem outros saber sobre suas intenções de fazer uma “pausa para o cafezinho”. O número das pessoas em pausa para café é mostrado a outros, o que os deixa saber quando um grupo está se juntando (o que, de acordo com os autores, indica um acontecimento social em andamento). Tickertape exibe informação de eventos em uma pequena barra de rolagem (semelhante a um visualizador de ações) e permite aos usuários saberem quando determinados eventos (que eles selecionaram) acontecem. A informação é distribuída de acordo com configurações explícitas dos usuários de “assinatura” de eventos, que indica em que “produtores de eventos” eles querem se concentrar em e que regras usar para filtrar mensagens pelo conteúdo. Uma terceira aplicação foi desenvolvida ligar o CVS (sistema de controle de versionamento de arquivos) a Elvin, possibilitando notificações de mudanças em arquivos. Através de

estudos com usuários, os autores determinaram que estas aplicações forneceram apoio a interações informais, tornando ações digitais visíveis e capacitando filtragem social e baseada em conteúdo de mensagens.

3.4 Sistemas baseados em agentes

Sistemas multi-agentes (MAS) são um bom meio de modelar sistemas de apoio a trabalho distribuído e em ambientes flexíveis, tal como o sendo estudado. Um MAS é uma rede fracamente acoplada de resolvidores de problema (agentes) trabalhando em conjunto para resolver um dado problema (WANG et al. 1999). Agentes inteligentes são entidades que percebem seus ambientes através de sensores e agem sobre este baseados em um raciocínio interno (RUSSEL, NORVIG, 1995). São sistemas computacionais encapsulados situados num ambiente e capazes de ação autônoma e flexível nesse ambiente para alcançar seus objetivos (Woolridge, 1997). As técnicas orientadas a agentes já foram aplicadas a um número de aplicações comerciais, tais como: telecomunicação e aplicações industriais, e estas vêm crescendo à medida que empresários e desenhistas compreendem seu potencial (JENNINGS,2001).

Aplicações baseadas em agentes são apropriadas para o desenvolvimento de sistemas complexos e distribuídos, compostos de subsistemas relacionados com organização variável (JENNINGS, 2001). A adoção de uma abordagem baseada em agentes viabiliza uma representação mais precisa da natureza descentralizada de um problema, de seus múltiplos pontos de controle e de perspectivas diferentes ou interesses envolvidos. Os agentes necessitam interagir uns com os outros para alcançar objetivos individuais e administrar interdependências, o que significa que devem ter capacidades sociais, e leva a seu uso em sistemas de groupware (BERGENTI et al., 2002).

Um arrazoado das pesquisas sobre a aplicação de agentes em groupware e CSCW pode ser encontrado em (ELLIS, WAINER, 1999) e (WAINER, ELLIS, 1998). NEEM (ELLIS, et al., 2002), Personal Assistant (ENEMBRECK, BARTHÈS, 2002) e COLABORADOR (BERGENTI et al., 2002) são exemplos de abordagens de agentes usadas para desenvolver ferramentas colaborativas. Em um cenário fracamente estruturado, agentes podem ser imbuídos da autonomia necessária para controlar a intensidade de comunicação e troca de informações quando indivíduos tornam-se mais próximo, ou ajustar seu comportamento de acordo com as necessidades dos grupos.

3.5 Arquiteturas para Disseminação de Informação

Diferentes arquiteturas foram propostas para viabilizar a distribuição de informações. Big Watch (BW) (Kirsch-Pinheiro et al., 2003), por exemplo, é uma estrutura para apoiar a construção de sistemas de awareness de acontecimentos passados. Este framework manipula informação de eventos definidos e distribuídos de acordo com determinações dos projetistas. A distribuição acontece em três fases:

- na primeira fase (registro), o projetista registra eventos de groupware com o framework,
- na segunda fase (controle), atividades que acontecem dentro do groupware são passadas ao framework;
- Na terceira fase, (notificação), membros do grupo são notificados da ocorrência de eventos aos quais tenham “assinado”.

O projetista necessita identificar de antemão que atividades são importantes para efeitos de awareness e registrá-las com o framework. O sistema possibilita a definição de filtros baseada em papéis, membros ou uma combinação dos dois, de modo que certos eventos podem ser declarados como sendo de interesse a certos papéis ou membros, para reduzir a quantidade de informação fornecida.

A estrutura de NAJA conta com restrições para equilibrar a privacidade e a interrupção dos usuários. Em sua implementação, um *mediador* é um controlador central, que controla a comunicação entre dois elementos (chamado *peers*). Os *peers* coletam e/ou distribuem informação, e o mediador filtra esta informação de acordo com as restrições estabelecidas. É possível ainda definir filtros que restringem a liberdade individual, para representar circunstâncias e condições particulares; prevenindo incômodos e protegendo privacidade (Gross, 1999). Vários tipos de restrições podem ser definidos, baseados em estruturas sociais, físicas ou legais.

Aldunate et al. (2002) apresentam uma arquitetura baseada em agentes que utiliza dois agentes: um agente de perfil e um agente técnico. Os primeiros coletam e recuperam informação sobre o usuário, construindo um modelo do usuário, enquanto o segundo cuida da infra-estrutura de rede e comunicação. Estes agentes fornecem uma estrutura baseada na descoberta de serviços, e os agentes de perfil tentam combinar perfis de usuários que sejam similares.

Oh (2004) propõe outro sistema de consciência baseado em agentes, onde cada usuário tem um agente que mantém informação de presença sobre seu usuário e a transmite sobre demanda. Esta informação é colhida via agentes perceptivos que coletam informação via câmeras ou Bluetooth. Ela é passada a cada usuário via seus agentes de consciência que atualizam as informações sobre os outros usuários.

Estas arquiteturas implementam ao menos dois níveis, um de infra-estrutura, onde comunicação com o sistema operacional ou aplicações é realizada e um para raciocinar sobre preferências ou eventos. Eles também adotam distribuição baseada em messaging, definindo acontecimentos que são de interesse ao usuário e informando-o quando estes ocorrem.

3.6 Considerações

A consciência é conhecimento sobre o estado de um ambiente dinâmico, que deve ser mantido com o passar do tempo. A maioria das tecnologias de consciência é implementadas como servidores de notificação, onde usuários “assinam” certos eventos e são informados quando esses acontecem (RAMDUNY et al, 1998). Isto normalmente significa que algum esforço é necessário para configurar o sistema, especificando que acontecimentos observar e que notificações enviar. No entanto, necessidades de consciência mudam de acordo com a situação, o que quer dizer que configurações de notificação necessitam ser revisadas constantemente para serem mantidas atualizadas com as necessidades dos usuários. Apesar da grande quantidade de esforços para lidar com a consciência, a maioria dos projetos ainda lidam quase que exclusivamente com a pergunta de como fornecer mecanismos para integrar e disseminar informação, com pouca ou nenhuma consideração sobre o conteúdo fornecido ou a seleção entre elementos diferentes. A consciência é mantida pela coleta e disseminação de informação.

Poucos destes sistemas possuem a capacidades de perfis de usuário. Adicionalmente, a maioria deles só apóia grupos preexistentes; não permitem a descoberta de novos parceiros nem novas oportunidades de colaboração. Ao contrário, a maioria dos sistemas de recomendação procura fazer apresentações e ajuda ao usuário expandir sua rede de contatos, mas não fornece nenhuma informação de consciência nem de acompanhamento de atividades.

La plupart des systèmes présentés doivent être configurés pour l'utilisateur, ce qui pose une barrière pour son utilisation. En plus, la majorité des systèmes n'a pas de capacité de construire profils d'utilisateur ni de découvrir nouveaux groupes dès qu'ils se forment, et permettent seulement groupes prédéterminés. En revanche, systèmes de recommandation ont comme but de présenter nouvelles informations à l'utilisateur et permettre l'expansion des réseaux de contacts mais ne donnent pas d'information de conscience ou activités.

4 Estudos de Caso

Pour bien comprendre le travail faiblement structuré, deux études de cas ont été faites, avec deux différents groupes de collaborateurs. Le premier était un projet de longue durée avec un groupe de 16 collaborateurs de différents pays de l'Union Européenne. Le deuxième, un projet court avec un groupe de 8 personnes qui travaillaient souvent ensemble et se connaissaient bien

Para melhor entender trabalho colaborativo fracamente estruturado, dois estudos de caso foram conduzidos, com dois grupos diferentes de colaboradores. O primeiro era um projeto de longo prazo envolvendo um grupo de 16 colaboradores de países diferentes na União Europeia, e o segundo um projeto de curto prazo envolvendo um grupo pequeno de colaboradores próximos.

4.1 Caso 1: Caso Europeu

O objetivo deste estudo é obter melhor entendimento de colaboração remota fracamente estruturada. O grupo estudado é composto de participantes de um projeto de pesquisa e desenvolvimento financiado pela Comunidade Europeia envolvendo 16 parceiros de países diferentes. Um subconjunto destes parceiros (conhecido como “parceiros técnicos”) está diretamente envolvido no projeto e desenvolvimento e prototipação de sistemas. O projeto tem 4 anos de duração, e ainda está em andamento.

O projeto tem por objetivo o desenvolvimento de ferramentas flexíveis e interoperáveis para apoiar serviços de eGovernment. O desenvolvimento segue o modelo de código aberto, e necessidades do usuário estão sendo extraídas através do desenvolvimento de sistemas-piloto em países diferentes. Mais tarde no projeto, pesquisas socioeconômicas avaliarão o impacto destas tecnologias nas administrações municipais.

Este é um projeto grande envolvendo vários colaboradores que trabalham em locais diferentes e trocam informações conforme necessário. Dada a natureza complexa do projeto, muitos problemas estão abertos para pesquisa e experimentos com novas

tecnologias. Um dos resultados esperados é um conjunto de diretrizes para pesquisas futuras e para a implementação de outros sistemas em toda a Europa. Devido à superficialidade da definição e natureza flexível do projeto, parceiros tiveram que definir metas concretas e entregas, definindo expectativas para os próximos anos de trabalho.

O estudo foi conduzido através de um misto de observação participativa, onde o pesquisador participou em um encontro de dois dias entre os parceiros técnicos e análise de interação, onde aproximadamente 18 meses de mensagens no fórum e 2 sessões de conversa via IRC foram disponibilizadas para análise. Entrevistas informais foram conduzidas para esclarecer alguns pontos. Com este estudo, nós tentamos compreender melhor os tipos de interação existentes no grupo, quanto tempo é gasto em cada uma, como a consciência é mantida e como a coordenação e auto-adaptação são alcançadas.

4.1.1 Organização da Equipe e Sistemas de Apoio

O projeto começou em resposta a uma chamada de propostas pela União Européia (UE). Participantes iniciais contataram uns aos outros (usando suas redes intensionais) até que um grupo foi formado que satisfizesse os requisitos constantes da chamada da UE. Desta maneira, não havia nenhuma figura central controlando a composição do grupo. No início do projeto, a definição passou por um processo de refinamento, pois tinha sido descrita em termos vagos e sem uma especificação formal. Os participantes refinaram a visão do projeto, construindo o planejamento de entregas durante os anos vindouros. Uma vez feito o planejamento, a modularização e divisão do trabalho foi decidida, com os parceiros espontaneamente se oferecendo para trabalhar em partes diferentes do projeto, de acordo com seus interesses ou habilidades. O trabalho de coordenação dos diferentes módulos é dividido entre os membros.

Várias equipes compõem o grupo de projeto, e cada equipe tem funções e responsabilidades diferentes, definidas no início do projeto. Neste estudo, nos concentramos nas equipes técnicas, que realizam pesquisa e desenvolvimento, uma vez que estas estavam mais acessíveis. Estas equipes têm à sua disposição um website, que funciona como o principal site para o projeto, onde documentos relacionados ao projeto são armazenados e conversas podem ser empreendidas através de uma ferramenta de fórum. Entregas parciais são colocadas no site com pedidos de feedback, assim como especificações finais do projeto e planejamento de próximas etapas. Interações ocorrem

através de fóruns, num espaço privado aberto só aos participantes. Interações nos fóruns tornam-se públicas a todos os membros do grupo. Assim, o site provê um repositório central e público de dados e informação do projeto, e fornece um meio para interação entre participantes.

Os membros tentaram se organizar de modo a reduzir a interdependência entre eles, dividindo tarefas de modo a só precisar colaboração fraca e reduzir a interdependência entre eles. A estrutura de controle é variável: apesar de haver um membro no papel de coordenador, os diferentes parceiros assumem responsabilidade por diferentes etapas do projeto, distribuindo as tarefas de coordenação entre todos. No entanto, parceiros se tratam como iguais, sem haver tratamento especial para o coordenador designado ou o coordenador geral. O parceiro no papel de coordenador tem por incumbência a escrita de relatórios consolidados e lembrar outros participantes de prazos vindouros, mas de cada participante é esperada contribuição com os próprios relatórios e o fornecimento de feedback e sugestões nos documentos antes destes serem enviados à Comissão Européia (CE). Portanto, a responsabilidade é compartilhada entre os membros, que são todos responsáveis pelas suas partes e pelo projeto como um todo.

Os membros do grupo têm relativa liberdade ao decidir como trabalhar no projeto e como organizar e dividir seu trabalho. Isto gera um número de problemas, visto que os parceiros se organizam baseados apenas num entendimento tácito de que o projeto deve ser completado até certa data. Eles não estão organizados hierarquicamente: em vez disso, subgrupos se organizam e coordenam os próprios esforços para alcançar sua meta final. Adicionalmente, sendo de organizações diferentes, membros do grupo estão freqüentemente empenhados em múltiplos projetos ao mesmo tempo (incluindo projetos fora do escopo), portanto devem gerenciar restrições e interdependências advindas da existência de focos diferentes. Esta configuração corresponde a nosso foco de pesquisa, visto que é um grupo fracamente estruturado e auto-organizado.

Apesar de reuniões ocasionais, os parceiros têm dificuldade de coordenar suas tarefas, trocar idéias e manter-se a par do trabalho de outros. Isto acontece principalmente porque há certo nível de improviso e adaptação que dentro de cada equipe, e membros acabam por invadir os limites de trabalho uns dos outros, às vezes duplicando trabalho em conseqüência de necessidades individuais imediatas. Alguns parceiros colaboram mais proximamente, tendo interações individualizadas via email ou

messenger e compartilhando trabalho e idéias ou tomando decisões sobre diferentes aspectos do projeto. Outros membros ficam de fora destas conversas, e, portanto, têm pequena oportunidade de reagir ou fornecer sugestões ou idéias numa etapa onde estas poderiam ser mais facilmente incorporadas.

Para entender o trabalho do grupo técnico e como este organiza suas tarefas e interações, nós:

- Conduzimos uma análise do histórico de suas mensagens online no fórum (que é acessível por todos os membros do projeto);
- Conduzimos uma análise de duas sessões de conversa online conduzidas no início do projeto;
- Participamos numa reunião de dois dias, durante a qual nós observamos a dinâmica de interação do grupo e anotamos problemas que surgiram.

4.1.2 Interações Assíncronas: Fórum

O fórum é a principal forma de interação de grupo, uma vez que está disponível para todos os membros. No momento da análise, o fórum principal continha 76 conversas (threads - 708 mensagens), o primeiro datado de 22/06/2004. Analisamos os logs até 18/11/2005, compondo um ano e 5 meses de mensagens. O comprimento médio dos threads era de 9,32 mensagens, e índice médio de participação de 4,23 participantes distintos por thread. As mensagens do fórum foram analisadas individualmente e codificadas, sendo algumas referências particulares destacadas (tal como menção a trabalho individual ou comunicação via meios não públicos). Os threads representam conversas entre membros, portanto era interessante verificar quais os principais focos de interação. As mensagens foram codificadas de acordo com seu conteúdo geral, e threads foram classificados em uma das seguintes categorias, dadas as mensagens que as compunham:

- Programação – estes threads normalmente continham duas vertentes distintas de conversa: a primeira, a respeito de programação de reuniões entre os parceiros, revolvendo ao redor de concordância sobre uma localização (que poderia ser online via IRC) e busca das melhores datas para interação síncrona para todos os membros. Estes threads incluíam também instruções sobre como chegar ao local da reunião, onde se

hospedar, etc. A segunda linha de conversação era sobre agendas para as reuniões. Uma agenda inicial era sugerida pelo membro encarregado da coordenação (que variava dependendo da etapa do projeto). Eram então trocadas mensagens sobre os itens que deveriam estar na agenda, e participantes contribuíam novos itens quando viam necessidade. A agenda final era definida em função desta negociação. Além disso, participantes às vezes colocavam documentos com suas idéias ou apresentações nestes threads, de modo que outros pudessem dar uma olhada e discuti-los na reunião.

- Anúncios – estes threads consistiram principalmente de mensagens anunciando um acontecimento iminente (e.g., uma conferência ou a próxima data de revisão) ou uma referência que o participante pensou ser de interesse aos outros. Estes eram normalmente threads de uma só mensagem, contendo mensagens que não necessitavam de resposta.
- Pedidos de Contribuição – visto que muito trabalho era conduzido em paralelo e sem esquema central de coordenação, era esperado que cada participante fornecesse suas contribuições aos relatórios consolidados e documentos ser apresentado à CE. Nestes threads o coordenador do módulo do projeto solicita a cada participante que envie suas contribuições, para então combiná-las em um único documento ser enviado à CE. O coordenador normalmente coloca este documento no thread para revisão pelos outros antes de enviá-lo à CE. Estes threads às vezes incluíam mensagens de clarificação, onde parceiros tentavam esclarecer o que exatamente estava sendo pedido e o que era esperado deles.
- Conversas técnicas – estes threads normalmente começam com um dos participantes fazendo uma pergunta ou solicitando opiniões sobre um tema, e outros então oferecendo seus pontos de vista no assunto. Uma conversa se segue, com participantes trocando idéias e informação técnica. A interação normalmente trata de uma pergunta técnica em particular ou de algum assunto que os parceiros acreditam que possa surgir durante uma revisão da CE.

- Informação de Status – estes threads eram semelhantes a threads de anúncio, exceto por conterem informes para um (ou mais) membros. Os participantes colocavam informações sobre uma reunião não conduzida no fórum ou documentação gerada, por exemplo, de modo que outros pudessem verificá-la, e tornar-se ciente do que eles estavam fazendo e fornecer feedback. Alguns relatórios de status geravam discussões. A diferença principal entre estes e anúncios, é o fato que anúncios informam sobre acontecimentos externos ao grupo, ao passo que relatórios de status informam sobre atividades ou artefatos gerados por um dos membros do grupo.
- Planejamento – threads de planejamento contêm conversas sobre subdivisão de trabalho, designação de tarefas, interdependências, prazos e adaptações planejadas. Estas às vezes eram formalizadas em documentos a ser enviados à CE, e estavam sujeitos à revisão freqüente. Threads de planejamento também incluem aquelas em que parceiros trocam informações para se organizar, coordenar uma apresentação de grupo ou outra atividade coletiva. Nestes, normalmente era decidido quem apresentaria o que, os participantes compartilhavam suas apresentações e outros forneciam opiniões e idéias.
- Lembretes – este tipo de thread aparece menos freqüentemente. Consiste principalmente de parceiros lembrando outros da necessidade de atualizar as informações no site, ou que devem publicar seus artigos no site. Normalmente estas mensagens não geravam respostas.

A maioria dos threads do fórum recaiu em um das categorias supracitadas. Em algumas ocasiões, o tema mudou de uma categoria a outro no meio da conversa. Nestes casos, nós utilizamos a tendência predominante do thread para classificá-lo. A distribuição de mensagens no fórum é mostrada na Tabela 4-1.

Nós também tivemos acesso a duas interações síncronas que aconteceram no grupo: reuniões face a face, na qual muitos membros estiveram presentes e o projeto foi discutido durante dois dias, e duas conversas online via IRC, onde participantes se logaram num momento combinado previamente para discutir o projeto. Analisamos dois

logs de conversas via IRC e assistimos a uma reunião face a face de dois dias onde observamos as dinâmicas do grupo.

Tabela 4-1: Distribuição de mensagens no fórum

Tipo de atividade	Número de mensagens	Percentual
Programação	278	39.27%
Solicitação de Contribuições	198	27.97%
Planejamento	108	15.25%
Conversas técnicas	76	10.73%
Informação do Status	31	4.38%
Anúncios	13	1.84%
Lembretes	4	0.56%
<i>Total</i>	<i>708</i>	<i>100</i>

4.1.3 Reunião Face a Face

Os seis parceiros técnicos estavam presentes na reunião face a face à qual nós participamos. Anotamos os temas discutidos e a dinâmica do grupo. Nós também fizemos perguntas informais para esclarecer arranjos cooperativos entre os participantes. A reunião tinha sido previamente acertada no fórum, e uma agenda tinha sido postada de acordo com a percepção do grupo dos pontos que necessitavam ser discutidos. A reunião durou dois dias, e caracterizou-se por conversas técnicas sobre aspectos diversos do projeto no primeiro dia, mas esteve dividida entre conversas técnicas e planejamento/organização no segundo dia.

Durante o primeiro dia, oito temas foram tratados, dos quais só dois de planejamento (uma clarificação sobre algo que havia sido previamente decidido e sobre o que seria apresentado na próxima revisão da CE). As seis restantes eram conversas técnicas sobre aspectos diversos do projeto. Estes seguiram o mesmo formato geral: um dos participantes apresentava um trabalho em andamento e solicitava comentários dos outros participantes, que então forneceria seus pensamentos. Durante uma destas discussões, sobreposições foram achadas entre trabalho em andamento de membros diferentes, e a discussão então reverteu para a gerência destas interdependências para evitar duplicação de trabalho. Neste caso, ficou claro que os participantes ignoravam o trabalho uns dos outros, e algumas duplicações já tinham acontecido devido a isso.

Durante o segundo dia, nove temas diferentes foram discutidos. Durante a primeira discussão técnica, na qual quase todos os participantes se envolveram, algum planejamento também aconteceu, na medida em que os participantes decidiram sobre

calendários de entregas, marcos, e quem seria responsável por que tarefas. Depois que divisão de trabalho foi acertada, a conversa técnica continuou. A discussão seguinte também foi técnica e a terceira foi um planejamento e estabelecimento de interdependências entre participantes: um dos participantes solicitou dados de outro, desde que ela não podia executar testes de seu módulo sem estes. A quarta conversa novamente recaiu sobre planejamento, com um dos participantes pedindo aos outros a formalização do que estaria no relatório. Depois disso, durante uma conversa informal antes do almoço, todos os presentes verificaram as anotações de um dos participantes e decidiram que estas seriam usadas como ata da reunião. Os participantes vinham tomando notas individuais (não havia ninguém tomando notas consolidadas) e foi decidido que estas estavam mais bem organizadas, então seriam mantidas como minutas oficiais. Depois do almoço do segundo dia, duas das discussões foram técnicas e as duas restantes de planejamento, sobre (1) onde publicar documentos e recursos compartilhados e (2) o que apresentar na próxima revisão da CE.

No segundo dia em particular, um grande número de interações ocorreu "em paralelo": participantes conduziram conversas informais em grupos de dois ou três e concordaram em colaborar em certos aspectos do projeto, planejando empreender certas tarefas em conjunto. Estes arranjos informais eram possíveis devido à proximidade e à repentina facilidade em determinar interesses mútuos ou trabalho semelhante. Por vezes estes esforços conjuntos eram anunciados ao grupo, mas não sempre. Estas colaborações não foram executadas "em aberto", no fórum para todos verem, mas por email, messenger ou em reuniões face a face somente entre os participantes envolvidos. Um dos participantes permaneceu um dia extra para trabalhar com o anfitrião em uma de suas tarefas. Assim, outros se tornam alheios à certas atividades até que estas sejam colocadas no fórum.

4.1.4 Reuniões via IRC

A primeira reunião de conversa de IRC começou com meia hora de acertar na tecnologia e organização, onde participantes mais acostumados às ferramentas sendo usadas ajudavam os demais a se logar, antes da reunião realmente começar.

Quatro membros do grupo técnico estavam presentes na primeira reunião de IRC. Dos 10 temas discutidos nesta reunião, oito foram assuntos técnicos, com abundância de discussão entre os participantes. Uma conversa foi sobre planejamento e

a outra sobre a definição de programa para a próxima reunião. Em geral, as dinâmicas eram semelhantes ao face a face encontrando, e conversas paralelas levam a arranjos particulares e mais reuniões entre os subconjuntos dos participantes apresenta também acontecido aqui.

Todos os seis participantes técnicos estavam presentes na segunda reunião online. Esta começou com um relatório de estado, onde participantes informaram aos outros sobre as tarefas que tinham realizado (ou que era necessário realizar). A interação seguinte serviu para estabelecer interdependências entre participantes e o que eles necessitariam uns dos outros, cobrindo também o segundo item da agenda. Isto se deu juntamente com o planejamento de uma reunião vindoura, onde os participantes decidiram o que seria apresentado. Uma longa discussão técnica e algum planejamento sobre ontologias seguiram, com participantes tentando decidir sobre abordagens e designar responsabilidade pelas partes do trabalho. A parte final da reunião ficou para a programação de um encontro mais tarde, a ocorrer antes da revisão de projeto.

Nas reuniões via IRC, assim como nas reuniões face-a-face, ocorreram algumas conversas paralelas. Estas envolviam um subconjunto do grupo planejando seu trabalho e estabelecendo acordos de cooperação individualizados. No IRC, isto era mais difícil de administrar, visto que não havia nenhum meio de estabelecer a quem uma sentença era endereçada. No entanto, ter uma conversa paralela à vista de todos faz com que seja mais fácil para os participantes entenderem que acordos estão sendo feitos entre os outros participantes e melhor entender o grupo como um todo.

4.1.5 Considerações

Há uma distinção clara entre os tipos de interação que se dão no fórum (assíncronas) e face a face ou via IRC (síncronas). Mais da metade das mensagens no fórum eram sobre programação e planejamento (386 mensagens) enquanto que, nas reuniões face a face e via IRC, as discussões técnicas prevaleceram (19 discussões técnicas no total, contra 13 relacionadas a planejamento e programação).

O número grande de pedidos de contribuições no fórum mostra a forma como o trabalho é organizado no grupo: tarefas são designadas a participantes e destes é esperado que forneçam relatórios para a geração de documentos consolidados a ser entregues à CE. Isto mostra que participantes tentaram se organizar de maneira a reduzir a interdependência, como também observado por Olson e Teasley (1996). Os relatórios

periódicos, no entanto, indicam que os participantes estão cientes da necessidade manter uns aos outros a par do trabalho em andamento, e que dão valor à revisão e sugestões que outros venham a dar sobre seu trabalho. Estas interações sobre entregas ou trabalho em andamento permitem que participantes alinhem seu trabalho de modo que este possa ser mais facilmente integrado numa etapa posterior. Elas também asseguram que participantes mantenham um entendimento compartilhado do projeto e de que necessidades existem.

Os participantes claramente preferiam meio síncronos de discussão de conteúdo técnico, que viabilizavam uma decisão mais rápida. Enquanto as mensagens no fórum normalmente eram respondidas em um dia, às vezes levava até uma semana para alguém responder. Frequentemente, nenhuma resposta era dada a não ser que algum participante fosse explicitamente nomeado e solicitado a fornecer alguma opinião. Isto traz à tona outro aspecto interessante do trabalho do grupo: a regulamentação é exercitada socialmente. Enquanto há um participante encarregado de coordenar tarefas e produzir os relatórios finais, cada participante está ciente da sua própria responsabilidade com o grupo e não quer ser aquele que "deixa peteca cair". A pressão social leva os participantes a tentar obedecer a prazos acertados e manter certo nível de compromisso. Quando não era possível fornecer o que foi solicitado no prazo, os participantes forneciam uma desculpa apropriada e uma data esperada para a informação necessária.

O trabalho e interações paralelas não estão evidentes nos números acima. Muitos threads continham referências a trabalho acertado fora dos fóruns (em conversas privadas via email ou face a face) ou requisições para comunicação via outros meios (fora do fórum, via email, telefone ou face a face). Havia ao menos uma destas mensagens em cada thread com mais de 10 mensagens. Uma vez que um thread se tornasse longo demais ou a discussão muito acirrada, participantes preferiam trocar para meios de comunicação síncronos, onde a interação podia ser continuada mais facilmente e sem o atraso de tempo existente no fórum.

Mesmo não sendo o melhor meio para interação, o site provia um repositório compartilhado valioso para os participantes. Em caso de dúvidas concernentes à designação de tarefas, horários ou decisões prévias, os participantes se referiam a documentos arquivados no site (especialmente aqueles entregues e aprovados pela CE). Os participantes frequentemente pediam que documentos fossem colocados no fórum

além de serem enviados via mensagem, de modo que fossem permanentemente arquivados e para que os outros pudessem ter acesso a estes documentos. O site web é uma rica fonte de informação, visto que toda documentação de projeto é arquivada nele, bem como produtos entregues e relatórios enviados à CE.

Em um caso interessante, uma interação entre dois participantes foi realizada no fórum, o que permitiu que todos os demais participantes do grupo a acompanhassem. Após alguns dias, um terceiro participante decidiu participar, fornecendo suas idéias sobre o assunto. Logo depois, um quarto participante se juntou a eles, e o thread transformou-se em uma discussão técnica animada, na qual todos os participantes técnicos participaram. Interações realizadas em público permitem que os participantes fiquem cientes de assuntos e participem quando necessário ou quando interessados.

Esta situação é semelhante à encontrada em áreas de trabalho compartilhadas: dois participantes discutindo nos corredores ou durante uma pausa para cafezinho (ou mesmo em um de seus escritórios) poderiam ter sido ouvidos por outros e estes poderiam se juntar à discussão facilmente. Isto é possível porque a interação ocorre em um espaço público. Estivessem todos os participantes em uma situação de escritório, é provável que estes tivessem colaborado mais freqüentemente, visto que a maioria dos “acordos paralelos” teria sido realizada abertamente, e outros teriam tido a oportunidade de contribuir quando apropriado.

Deve ser notado ainda o grande número de negociações entre usuários. A maioria das interações de planejamento e programação envolvia negociação de datas e tarefas, mas algumas soluções técnicas também envolveram decisões sobre a melhor solução e como esta deveria ser implementada. Quando uma solução era apresentada ao grupo ou quando sugestões eram solicitadas, o grupo discutia e determinava se outras abordagens preferíveis. Mesmo simples pedidos de contribuições às vezes envolviam negociação: por exemplo, quando um participante solicitou outras contribuições e alguém não conseguiria fazê-lo dentro do prazo proposto, um processo de negociação aconteceu, com participantes propondo alternativas até que um consenso fosse alcançado e um novo acordo fosse feito.

As principais dificuldades mencionadas por usuários e observadas durante as reuniões diziam respeito ao acompanhamento das atividades dos outros (quem estava fazendo o que), modos para evitar duplicação de trabalho ou discutir a abordagens

decididas em etapas anteriores. Um dos operadores expressou desejo de saber quando participantes mudavam suas configurações, trazendo alguém novo para o grupo. Isto acontecia com frequência, uma vez que as organizações participantes também tinham outros projetos e membros tinham que lidar com estes. Adicionalmente, conversas paralelas e sub-grupos se formavam para lidar com atividades em particular, e estes também devem ser apoiados.

4.2 Caso 2: Caso Artigo

Neste caso, analisamos as atividades e comunicação entre membros de um grupo pequeno e coeso de indivíduos escrevendo um artigo científico para revista. O objetivo deste estudo era melhor entender os colapsos de comunicação dentro do grupo, e o foco foi em um incidente que causou considerável desconforto dentro do grupo: a duplicação de uma atividade (a tradução do artigo para inglês). O grupo era composto de 4 membros de um grupo de pesquisa que, em um momento inicial, discutiu pessoalmente e concordou em escrever o artigo, motivado por uma figura hierarquicamente superior (seu supervisor de tese, comum a todos).

Na reunião inicial, os participantes concordaram no tema do artigo, selecionaram uma revista alvo e determinaram o prazo de entrega. Todos tinham extensa experiência com email, tinham escrito artigos e tinham conhecimento individual para trazer ao artigo. Deste modo, todos conheciam as tarefas a serem executadas e os meios de comunicação disponíveis, tinham experiência de trabalho via email e dados para trazer ao artigo. O grupo trabalhou junto neste artigo durante um período de 3 semanas. Apesar de 6 reuniões face a face, a maior parte do trabalho foi realizada individualmente, usando email como principal meio de comunicação, com indivíduos coordenando tarefas e negociando aspectos diferentes do texto eletronicamente.

A estratégia selecionada para análise foi o mapeamento via redes sociais da comunicação entre os membros de grupo, traçando colaborações, comunicação e troca de artefatos. Com a ajuda dos participantes, extraímos o fluxo de comunicação e atividades do grupo e construímos grafos de redes sociais (sociogramas) da seguinte maneira:

- Cada nó da rede representa um ator;
- Os arcos representam comunicação entre atores;

- Comunicação pode incluir transferência de arquivo (representado por um ícone de arquivo);
 - Arcos tracejados representam comunicação em *background* (não realizada explicitamente via email, para todos os participantes).
- Cada grafo é representa um passo no processo do grupo.

Assim, uma seqüência de grafos representa a dinâmica de interação durante o projeto e traça a comunicação e intercâmbio de arquivos entre os membros de grupo (isto poderia ser transformado em uma animação). O grafo completo é desenhado com a passagem do tempo: a cada mensagem trocada, um arco é fortalecido ou novos membros são introduzidos. Isto nos permite ver os padrões emergindo a partir dos dados e analisar a estrutura do gráfico resultante.

O projeto começou em 31 de março (houve uma reunião face a face antes do início do fluxo de mensagens), e estendeu-se até 11 de abril. Um histórico de mensagens e acontecimentos foi construído, juntamente com um mapa de atividade. Os principais membros do grupo são J, V, W e Z, que trabalharam mais ativamente no projeto. Membros periféricos são T, M, U, X e B, que foram consultados por razões diversas durante a escrita do artigo (em particular, T é o tradutor, que verificou o artigo antes que este fosse enviado). Com exceção do tradutor, nenhum destes participantes periféricos enviou mensagens para o resto do grupo: todos foram contatados em reuniões face-a-face e não foram envolvidos no trabalho em andamento. Os sociogramas foram construídos a partir dos registros de intercâmbios de email entre J, V, W e Z (um histórico parcial é mostrado na Tabela 4-2). Usamos o conjunto completo de emails trocado durante o projeto para construir os sociogramas. O Pré-processamento destes dados envolveu a retirada de redundâncias, uma vez que emails foram fornecidos tanto por remetentes quanto por receptores, causando duplicação no conjunto.

Além das mensagens de email, houveram reuniões offline freqüentemente ao longo do projeto. Nos últimos dias, o grupo descobriu que tinham incorrido em retrabalho. Houve um encontro inicial entre J e W no dia 28, outro entre V e W em 2 de abril. Houve reuniões em 3, 5 e 6 de abril, devido à proximidade do prazo final, e uma última no dia 10. Nestas, membros coordenaram suas tarefas e discutiram os temas e conteúdo do artigo, extraindo comentários dos participantes periféricos. Na reunião do

dia 10, participantes descobriram que tinham realizado trabalho desnecessário, pois o envio do artigo ao tradutor não tinha sido anunciado e um dos participantes tomou para si a tarefa de traduzir o artigo (desnecessariamente). Alguns atores sequer estavam cientes que outros tinham entrado no grupo até o último dia, na reunião final.

Tabela 4-2: Log parcial de mensagem para o projeto

Data/ Hora	De	A	Conteúdo e Anexos
31 Março			
12:16	V	W, J	Estrutura proposta; WebServices2006.doc
2 Abril			
23:15	V	J, W	Atualização de estado, próximos passos; diagramas.doc WebServices2006-V.zip
5 Abril			
00:08	V	J, W	Atualização de conteúdo; WebServices2006Artigo1.doc
1:59	J	V, W	Atualização de conteúdo, comentários; WebServices2006Artigo1.2.doc
21:23	W	J	Respostas a comentários, discussão
21:55	J	W	Comentários sobre conteúdo
6 Abril			
01:42	V	J, W	Atualização de conteúdo; Webservices2006Artigo1.3.doc,- base de dados.rmx,
12:09	W	J, V	Correção de informação de filiação Webservices2006Artigo1.4.doc
19:47	W	J	Informação adicional sobre autores
19:30	W	Z	Nova versão de artigo Webservices2006Artigo1.5.doc
8 Abril			
16:23	J	W, V	Versão final, enviar a tradutor Webservices2006Artigo1.5.doc
21:22	W	T, J, V	Solicitação de tradução Webservices2006Artigo1.6.doc
10 Abril			
17:42	J	O, M, U	Artigo final, coordenação, solicitação de comentários Webservices2006Artigo1.7.doc
19:11	W	J	Relatório de estado, coordenação
22:00	J	O, Z, U, W	Resposta de M
11 Abril			
00:18	Z	O J, U, V, W	Confirmação de submissão DAPD-S-06-0039.pdf

O log mostra uma grande quantidade de trabalho sendo realizada, inicialmente entre um grupo menor de colaboradores. Há conversas e versões diferentes do artigo e arquivos com pedaços do mesmo que são trocados. Quando uma versão razoavelmente estável é completada, um novo membro se junta ao grupo, e questiona alguns conceitos. Os membros anteriores continuam a trabalhar, e o artigo é enviado a um tradutor para revisão. Nas reuniões finais, os atores compreendem que tinham duplicado trabalho e que a coordenação não tinha sido adequada. As revisões finais são resolvidas face a face, e um dos atores fica com a responsabilidade de enviar a versão final do artigo depois que o último membro de grupo envie seus comentários.

A atividade do grupo é representada nos sociogramas da Figura 4-1: em um momento inicial, os nós J, W e V trabalham em paralelo e interagem frequentemente, discutindo, trocando informações e versões do artigo. Num segundo momento, Z se junta ao grupo, o arquivo é enviado a ele, que começa a trabalhar no projeto. J e V ignoram a existência do novo membro, e continuam seu padrão normal de colaboração. O terceiro gráfico mostra J e W coordenando o envio do artigo para T, o tradutor. Z não é notificado, e acaba por duplicar trabalho do tradutor. Em seguida, J solicita comentários de M, um membro periférico do grupo. As linhas pontilhadas representam conhecimento do que foi feito. J, V, W e Z tinham concordado que J enviaria o artigo a M. Não houve nenhuma comunicação explícita, mas era de conhecimento comum o que seria feito.

Ao final, o artigo foi escrito e foi submetido como planejado, então talvez pareça que não exista nenhum problema real nem razão analisar este caso. No entanto, em entrevistas com três de nossos quatro participantes, tornou-se claro que a experiência colaborativa foi pouco satisfatória. Todos eles se queixaram do trabalho, da confusão em comunicação e do trabalho adicional que foi realizado sem necessidade. Faltou sinergia ao grupo, que estava claramente fragmentado desde o começo. Comentários tais como "eu não sabia que eles ia enviar o artigo a um tradutor", "desperdicei meu tempo traduzindo isto" e "eu não tinha idéia que ele podia traduzir" foram ouvidos com frequência.

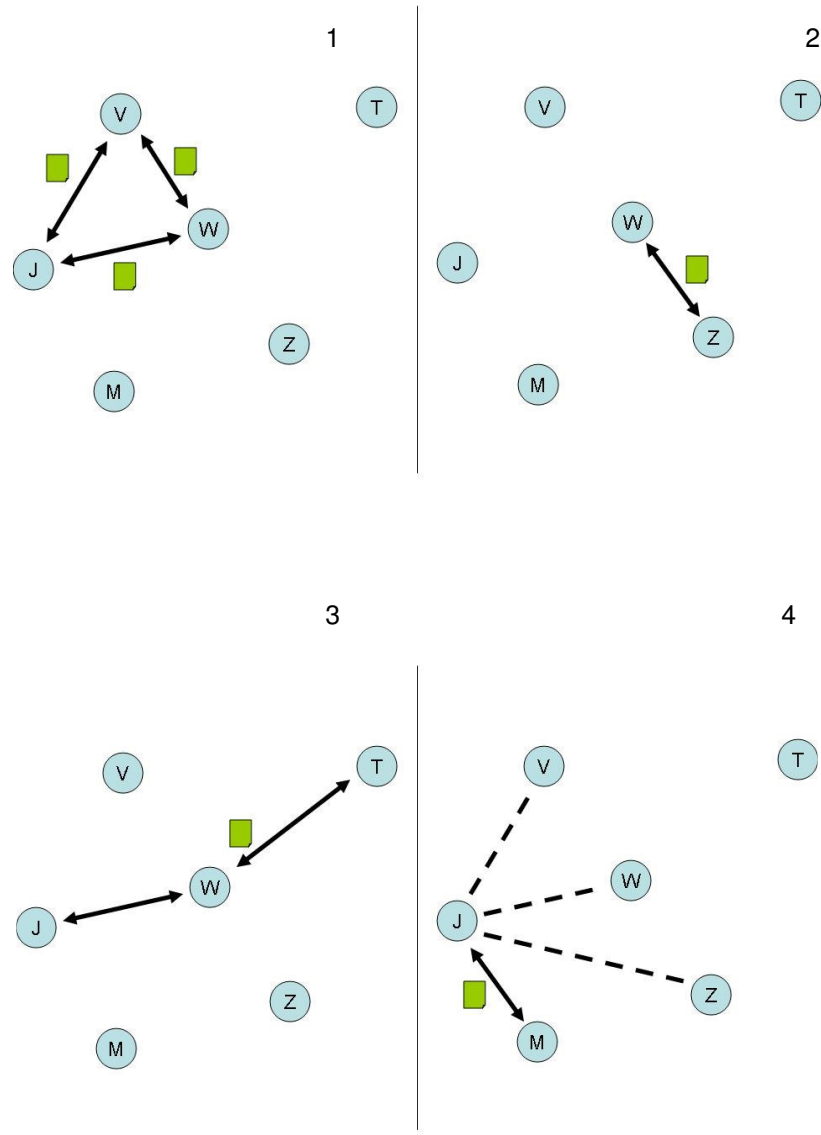


Figura 4-1: Uma série de sociogramas correspondente a momentos diferentes nas atividades do grupo

Por exemplo, um dos participantes se queixou de não ter sido informado que o artigo foi enviado ao tradutor: ele passou a noite revisando o artigo desnecessariamente, e seu trabalho foi desperdiçado. Outro queixou-se que alguém que não tinha visto o artigo antes nem acompanhado o processo solicitou várias mudanças alguns dias antes do prazo final. Pior, um dos colaboradores nem sequer sabia que outro estava envolvido até que todos se encontraram no dia antes do prazo final para fazer as últimas adaptações. Ao final do processo, todos tinham a sensação de ter trabalhado demais, e estavam descontentes e irritados com o trabalho. Em geral, o processo envolveu muita

discussão e esforço extra que poderiam ter sido evitados caso os colaboradores estivesse mais cientes do contexto social no qual eles operavam.

Ao olhar os gráficos, fica claro que os limites do grupo estavam mudando (o que é esperado), e que os indivíduos não estavam cientes dessas mudanças. Em dado momento, nó W tornou-se mais central, como indicado por seu grau de centralidade. Com a centralidade aumentada, vem responsabilidade adicional: para evitar duplicação de trabalho e equívocos, participantes precisam ser notificados de ações que impactem no trabalho do grupo (tal como enviar o arquivo a um tradutor ou a adição de uma pessoa ao grupo). Isso foi o que preveniu problemas adicionais quando ator M foi notificado e seus comentários foram solicitados. Visto que ele já tinha um laço mais forte a J, ficou para J lidar com M, enviando pedido de comentários e transmitindo a resposta aos outros, mas era de conhecimento comum que isto seria feito. Como M não era um participante local, ele não esteve pessoalmente envolvido e não podia participar de reuniões face a face.

Em termos mais gerais, numa situação de trabalho entre iguais, onde indivíduos têm tarefas semelhantes no projeto, a configuração ideal seria ter todos os membros recebendo informação sobre todos os acontecimentos, que criaria um grafo conexo. Quando um ator passa a uma posição diferente (como é o caso com ator W, que tornou-se mais central) ou ocorrem mudanças na estrutura do gráfico, há necessidade de que algo seja feito. Sendo esta uma mudança legítima, o ator agora deve acompanhar as interações e fluxo de mensagens, sendo a mudança involuntária (devido a um erro ou esquecimento), a estrutura mais "democrática" é desejável. Neste caso, um sistema subjacente poderia detectar estas mudanças estruturais e consertar este problema transmitindo informações para que outros saibam de novas adições ou acontecimentos no grupo.

As principais causas dos colapsos foram:

- Membros atuais não conheciam as capacidades e habilidades do novo membro
- Novo membro não conhecia combinações prévias
- Os membros não comunicaram uns aos outros a execução de atividades (duplicação)

- Os membros ignoravam a entrada do novo membro no grupo

Esta informação poderia ter sido disseminada para obter melhores resultados. Nas seções seguintes, nós apresentamos uma solução ao problema de disseminação de informação baseada nestas observações.

Les études de cas montrent que informations sur manipulation d'objets, composition du groupe et activités relationnés avec le travail du groupe étaient nécessaires pour le bon fonctionnement du groupe. Basée sur ça, nous avons développé une architecture, un modèle et une méthode pour la distribution d'informations. Dans les sections suivantes, nous présentons une solution pour le problème de dissémination d'informations basée sur ces observations.

5 Uma Arquitetura Baseada em Agentes para Captura e Organização de Informações

Étant donné que les gens fréquemment travaillent avec différents systèmes et plateformes, une architecture basée en agents pour la collection, analyse et distributions d'informations est proposée. Agents assistants intelligents pourraient aider l'utilisateur dans la gestion de ses liaisons avec ses collègues. Nous proposons une architecture basée sur agents pour la collection, organisation et distribution automatique d'informations.

Dado que em ambientes fracamente estruturados as pessoas frequentemente selecionam e trabalham com sistemas e plataformas diferentes, propomos a utilização de uma camada baseada em agentes como uma solução integradora. Agentes assistentes inteligentes poderiam ajudar usuários a administrar seus múltiplos laços, ligando-os às diferentes aplicações e fontes de dados. Propomos uma arquitetura baseada em agentes para apoiar a coleção, organização e disseminação automática de informação. Dada a relativa independência de cada indivíduo, adotamos uma arquitetura baseada em peers.

5.1 Uma Arquitetura Multiagente para Gestão de Conhecimento

OMAS (Open Multi Agent System) é base de agentes utilizada em uma série de projetos correlatos, que fornece uma infraestrutura para o desenvolvimento de sistemas baseados em agentes. Em OMAS, “coterias” são grupos de agentes cognitivos que trabalham juntos. Não há nenhuma estrutura interna de controle em uma coterie, e coterias podem participar em interações de mais alto nível através de agentes de comunicação. Os agentes trocam mensagens assincronamente, e este intercâmbio de informações permite que eles mantenham seus modelos internos atualizados e persigam suas metas. A arquitetura é mostrada em Figura 5-1. Há dois tipos de agentes na plataforma OMAS:

- Agentes de serviço (Service Agents – SAs): fornecem serviços correspondentes a habilidades específicas (normalmente relacionados a aplicações individuais).
- Secretários particulares (Personal Assistants – PAs): interface com o usuário. Suas habilidades são dedicadas à compreensão e apresentação de informação de maneira inteligente e oportuna.

Os agentes interagem com sistemas de groupware, extraíndo e administrando informações. Por exemplo, um *agente de projeto* é um agente de serviço especificamente projetado para fazer a interface com software de gerência de projeto. Outro agente de serviço, o *agente carteiro* controla mensagens enviadas e recebidas.

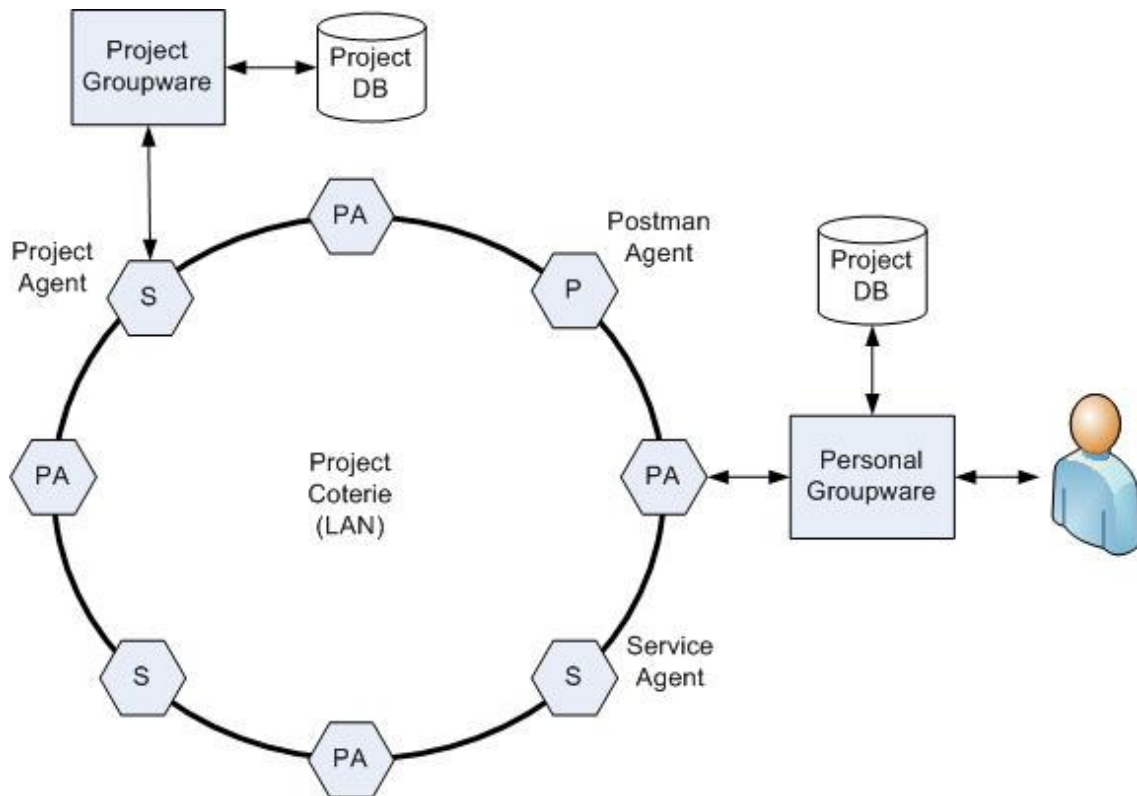


Figura 5-1: Arquitetura multiagente OMAS (adaptado de TACLA, BARTHES, 2003)

Em OMAS, cada agente é um subsistema cognitivo independente, com a estrutura interna mostrada na Figura 5-2. O conhecimento é modelado em várias ontologias diferentes, usadas para raciocínio do agente. Os agentes compartilham uma estrutura básica comum, mas agentes assistentes têm uma interface com o usuário e um modelo de usuário (mostrado em cinza), que os agentes de serviço não têm. Os seguintes elementos compõem cada agente:

- Interface de rede (network interface): implementa protocolos de comunicação.
- Controle (control): é responsável pelo comportamento do agente.
- Habilidades (skills): o conjunto de serviços que um agente pode fornecer. Uma habilidade pode ser expressa como procedimentos ou regras.
- Mundo (world): representação de outros agentes (e suas habilidades) e do ambiente.
- Tarefas (tasks): representação da tarefa atual do agente.
- Ontologia: um módulo para manipular o agente e as ontologias de ambiente. Ontologias são usadas para representar habilidades, tarefas e outras particularidades do agente.
- Próprio (self): representação das habilidades do agente, dados armazenados em sua memória e metas.
- Modelo de usuário (user model): preferências do usuário e um modelo do diálogo com o usuário.
- Interface com o usuário (user interface): capacita comunicação com o usuário.

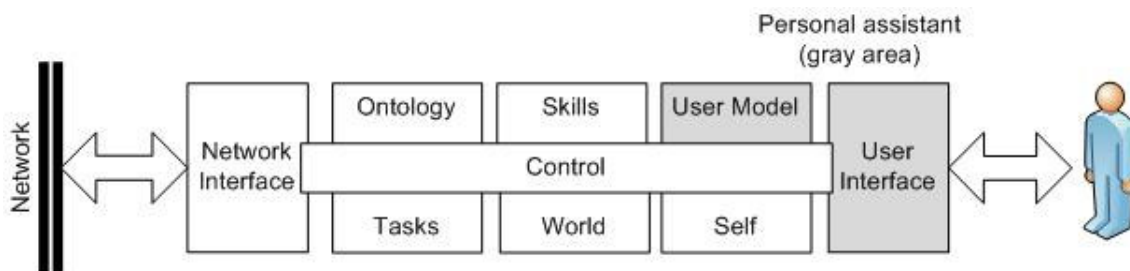


Figura 5-2: Arquitetura interna de agente OMAS (de TACLA, BARTHÈS, 2002)

CUMBIA, nossa estrutura multiagente, ecoa a arquitetura proposta por OMAS (Tacla, Barthès, 2003): Cada usuário é assistido por um grupo de agentes reativos que executam tarefas individualmente para ajudar o usuário. Os agentes colecionam e processam dados, fornecendo informação quando necessário. Além do mais, adotamos uma arquitetura baseada em peers, onde coteries diferentes representando cada usuário trocam informações.

5.2 CUMBIA: uma Estrutura para Apoiar Colaboração Oportunista

CUMBIA é uma arquitetura peer-to-peer baseada em agentes, para apoiar a gestão de múltiplas esferas de trabalho, em particular a formação de grupos improvisados e o estabelecimento de colaboração. CUMBIA baseia-se no conceito de “interações não planejadas”, definidas anteriormente.

5.2.1 Uma Estrutura para Apoio a Colaboração

Como o foco da arquitetura de CUMBIA é colaboração sem planejamento (espontânea), os dois requisitos principais são:

- Facilitar o encontro entre usuários que estão realizando trabalhos em contextos semelhantes;
- Viabilizar a interação entre estes usuários.

Os agentes são projetados para coletar informação sobre as atividades dos usuários e buscar potenciais colaboradores. Assim como em OMAS, CUMBIA fornece a cada usuário um conjunto de agentes de serviço (uma “coterie”) que controlam aplicações e ajuda com a gestão de conhecimento e tarefas de colaboração. Os agentes ficam encarregados de identificar situações cooperativas em potencial e fornecer informações relevantes de maneira oportuna. CUMBIA pressupõe que o grupo de agentes tem um objetivo compartilhado (apoiar a colaboração do usuário), logo nenhuma negociação entre agentes é necessária.

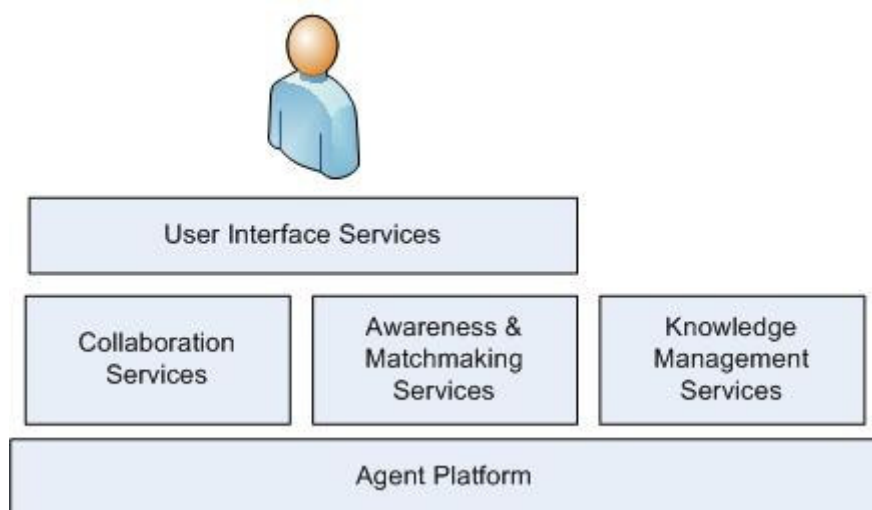


Figura 5-3: Arquitetura de Serviços CUMBIA

No framework, há quatro tipos de agentes de serviços que interagem para cumprir os requisitos predefinidos, fornecendo os serviços mostrados em Figura 5-3.

Quatro tipos de serviços necessitam ser fornecidos para alcançar o objetivo supracitado. Os agentes fornecem os seguintes serviços para apoiar encontros entre usuários distribuídos:

- Serviços de Interface com o usuário: realiza a apresentação de informações e permite ao usuário interagir com o sistema, especificando parâmetros e passando informações aos agentes. Isto é executado pelo PA, ao interagir com o usuário.
- Serviços de colaboração: permitem o rápido estabelecimento de contato quando a possibilidade para colaboração surge, através do fornecimento de ferramentas para cooperação (fóruns, messaging, etc.). Estes são executados por SAs ligados a ferramentas de groupware.
- Awareness e Serviços de Matchmaking: procuram por outros usuários com quem talvez seja interessante estabelecer contato; contacta outros agentes trocando perfis de usuários e contextos de trabalho; compara perfis de usuário ao contexto e ambiente atuais de trabalho. Realizados por SAs intermediários, que não estão necessariamente ligados a ferramentas de groupware mas recebem e processam informações sobre o usuário.
- Serviços de Gestão de Conhecimento: administra dados pessoais do usuário, constrói perfis de usuário baseados nestes dados e acompanha atividades em andamento, incluindo uso de documentos, realização de buscas, colaborações e pesquisas. Realizados por SAs, provêm também uma interface com ferramentas individuais de produtividade (eg., editores de texto) e ao sistema operacional.

A meta da coterie é determinar as múltiplas esferas de trabalho do usuário. Uma esfera de trabalho agrupa documentos, interações, contatos ou laços relacionados a uma colaboração em andamento. Em qualquer momento, um usuário estará trabalhando em uma de suas esferas de trabalho e podem alternar diferentes esferas. Em termos práticos, uma esfera de trabalho determina que recursos devem ser agrupados: documentos e laços podem ser salvos em qualquer lugar no disco rígido do usuário e são associados com um ou mais esferas de trabalho como visto na Figura 5-4. Quando o usuário passa a

uma esfera de trabalho diferente, estes documentos e contatos tornam-se “ativos”, já que são parte do contexto para esta atividade particular. Uma única esfera de trabalho estará em evidência de cada vez, embora múltiplas esferas possam estar ativas simultaneamente, indicando com isso diferentes trabalhos em andamento.

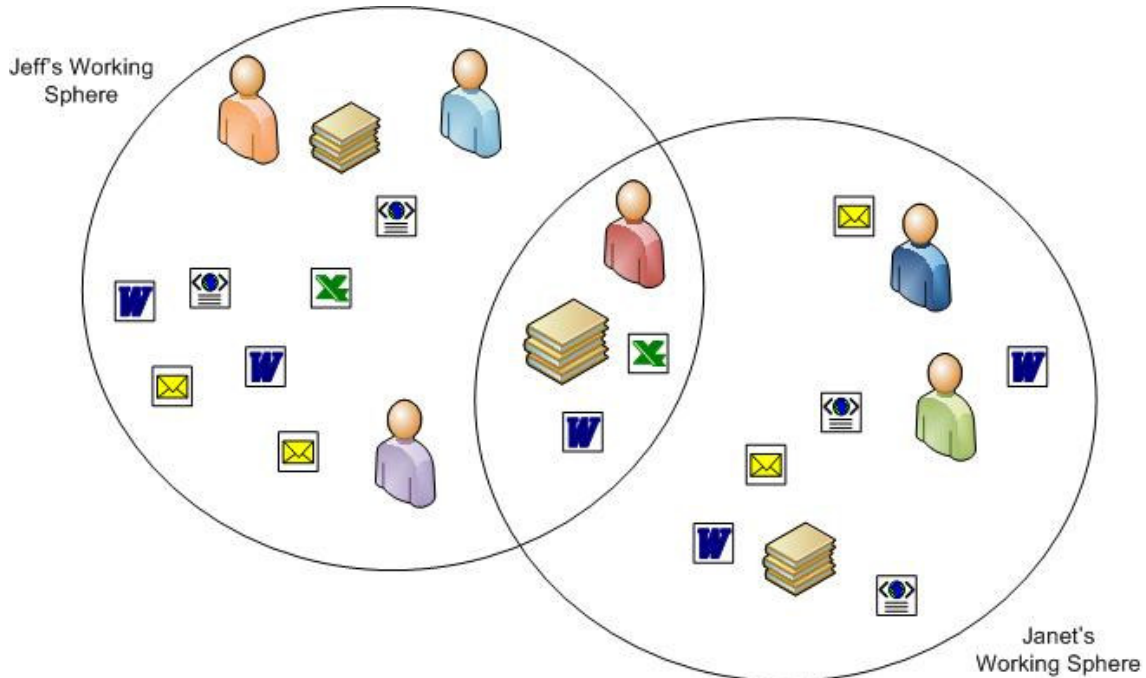


Figura 5-4: Esferas de trabalho

Cada agente de CUMBIA tem um objetivo individual, e se comunica com os outros de modo que a coterie possa alcançar seu objetivo global. Deste modo, agentes podem se auto-ajustar de acordo com o ambiente. Cada um dos serviços fornecido é descrito abaixo, assim como os agentes envolvidos, suas metas e raciocínio.

5.2.2 Serviços de Gestão de Conhecimento

Os serviços de gestão de conhecimento envolvem as atividades de coleta e organização de informações e funcionalidade de construção de perfis ao sistema. Os agentes são encarregados de tarefas relacionadas à coleta e organização de dados de usuário e recursos, e pela construção e manutenção de perfis de usuário que podem ser trocados com outros agentes. Eles também fornecem uma interface para outros groupwares e sistemas individuais de produtividade e o sistema operacional. Os agentes automaticamente inferem perfis de atividade de usuário, mas usuários sempre têm a última palavra sobre seus perfis, podendo corrigir informações e determinar que informação pode ser divulgada e o que deve permanecer privado. Em nosso ambiente, consideramos que uma pessoa sempre trabalha dentro de uma esfera de trabalho.

Esferas de trabalho contêm projetos, documentos, contatos, etc. relacionados ao projeto em questão.

Esferas de trabalho estão relacionadas a documentos, pessoas, colaborações e pesquisas, mas são inerentes a cada usuário, uma vez que papéis e posições hierárquicas diferentes podem gerar visões diferentes sobre o mesmo contexto de grupo. Isto por que indivíduos têm impressões pessoais da realidade e organizam seu trabalho de acordo com estas. As seguintes informações são mantidas por agentes de Gerência de Conhecimento:

- **Pessoas:** uma lista de contato do usuário, classificado em diferentes categorias, tais como contatos pessoais ou de trabalho, colaboradores prévios, correntes ou potenciais, pesquisadores, etc. Parte desta informação pode ser inferida a partir de emails ou mensagens enviadas e recebidas pelo usuário.
- **Contatos:** informação necessária para outra pessoa contatar o usuário: nome, email, telefone, etc. Possíveis fontes de dados para esta informação são email, messengers e software de gerência de contato, que armazenam informação sobre contatos de usuários.
- **Áreas de Interesse:** áreas gerais em que o usuário tem interesse. Estas podem ser automaticamente descobertas ou manualmente informadas, e são classificadas por nível de interesse e atividade. Esta informação pode ser reunida a partir de arquivos manipulados ou sites web visitados, usando técnicas de extração de informação.
- **Histórico:** agentes podem seguir os acessos a páginas quando o usuário navega na Internet. Atividades são registradas e seguidas criando históricos de colaboração que podem dar apoio à interações futuras.

É necessário avaliar que mecanismos podem ser usados para que os usuários possam tomar conhecimento de seus pares ou recursos (locais ou documentos) de acordo com a relevância ao trabalho ou projeto em andamento ou a uma área de interesse, recebendo informação do sistema.

5.2.3 Awareness e Serviços de Matchmaking

Os serviços de matchmaking envolvem atribuição de usuários e recursos as suas esferas de trabalho respectivas, de modo que trabalho e recursos correlatos sejam agrupados em um mesmo contexto colaborativos. As oportunidades para colaboração são identificadas através da combinação de uma esfera de trabalho ativa do usuário com atividades ou interesses de outros usuários. Quando duas esferas de trabalho são semelhantes ou relacionadas (ou quando uma esfera de trabalho e um recurso de outro usuário estão relacionados), é possível que exista uma oportunidade para colaboração. Usando a informação manipulada por cada usuário, nós procuramos documentos que se relacionam a esta, em ambientes de outros usuários.

O matchmaking pode ser realizado em tempo real ou “em background”: matchmaking em tempo real ocorre enquanto o usuário trabalha, inserido num contexto e o sistema busca oportunidades de colaboração baseadas nesse trabalho. Em background procura usuários cujas esferas de trabalho ou recursos podem ser relacionados a esferas de trabalho ou áreas de interesse do usuário. Para acelerar as buscas, agentes podem buscar previamente o espaço e armazenar modelos de usuário simplificados para colaboradores em potencial, que são então usados como ponto inicial para buscas.

5.2.4 Serviços de Colaboração

Quando surge uma oportunidade para colaboração, um usuário deve ser imediatamente notificado. As oportunidades são sensíveis ao tempo, e o usuário deve receber informação sobre o potencial de reciprocidade do outro usuário ou suas motivações para entrar em colaboração. Esta informação pode incluir sociedades passadas e comportamento cooperativo ou laços comuns entre os dois indivíduos.

Uma vez identificadas as oportunidades de colaboração, um indivíduo pode tornar-se um colaborador ativo ou incidental de outro usuário. Colaboradores incidentais fornecem sugestões ocasionais ou participam de reuniões ocasionalmente. Colaboradores ativos estão inseridos no projeto e têm que lidar com horários e prazos finais. Neste caso, é importante saber o estado de cada participante e que tarefas são dependentes dele(a).

A iniciação de colaboração deve ter baixo custo para o usuário o, que significa que a interface deve viabilizar o fácil estabelecimento de um canal de comunicação para

estabelecer contato inicial. Os usuários podem decidir se engajar em interações mais longas, uma conversa ou troca de mensagens. Outra possibilidade é o agente notificar o usuário sempre que localizar alguma informação ou documento que sejam úteis a um colega e enviar essa informação ao colega automaticamente (depois de pedir permissão do usuário). Desta forma, um usuário não tem que preocupar-se em localizar documentos adequados, históricos nem informação apropriada a ser enviada aos outros.

Ferramentas normais de apoio à colaboração (listas de mensagens, sistemas de mensagens instantâneas, whiteboards compartilhados, mecanismos de arquivamento e compartilhamento) podem ser fornecido por agentes, que se tornam ativos de acordo com as necessidades. Isto ajudaria com o problema de lembrança dos canais preferidos de comunicação para cada um dos laços.

5.2.5 Serviços de Interface com o Usuário

Os agentes de interface de usuário (UI) fornecem informação aos usuários e solicitam informação deles. Os agentes de UI também mediam trocas entre outros agentes e usuários. A interface básica deve exibir informação de maneira não intrusiva, para não interromper o usuário. No entanto, informações adicionais devem ser facilmente acessíveis com um click ou rollover do mouse. Um agente de interface pode selecionar diferentes meios de exibir informação dependendo da situação. Ele pode, por exemplo, reorganizar uma lista de contatos quando prioridades mudam ou acender ícones quando um contato se torna ativo.

5.3 Estudo Piloto: Navigator – um Sistema para Promover Encontros entre Colaboradores Potenciais

Analisar o potencial da estrutura de CUMBIA, uma sistema foi implementado como prova de conceito. Navigator é uma aplicação multiagente para apoiar a colaboração oportunista que adere à estrutura CUMBIA. Neste sistema, agentes reúnem informação de atividade dos usuários e constroem um perfil que é trocado com outros agentes de usuário. Navigator retrata atividade com base no tempo e utiliza métodos de comparação de perfis para recomendação de parceiros em potencial. Os usuários são comparados com base em suas atividades atuais, interesses e habilidades.

Em comunidades grandes, encontrar especialistas ou parceiros às vezes pode ser difícil. Para buscar novos parceiros, uma pessoa frequentemente pede recomendações de

conhecidos. Isto é uma forma de navegação social, pela qual uma procura na rede é realizada, por outros que talvez se encaixem nos perfis necessários. O tempo é um fator importante neste tipo de situação, visto que necessidades, interesses e níveis de habilidade mudam rapidamente. Navigator é um sistema para ajudar a navegação social e a busca na rede, baseado na arquitetura CUMBIA. Colaboração oportunista acontece quando participantes usam seu conhecimento sobre outros para identificar oportunidades imprevistas de colaborar. Este tipo de colaboração pode ser visto de dois pontos de vistas diferentes: as interações que levam a trabalho em conjunto e interações com o objetivo de fornecer auxílio. Comunicação informal é vital para colaboração, pois permite que indivíduos se mantenham a par do que está acontecendo em numa organização e das habilidades e disponibilidade dos outros. Navigator é um sistema de distribuição de informação sobre atividades dos outros, para promover colaboração oportunista. Os seguintes agentes foram criados:

- Gestão de conhecimento: agentes de interface com MS Word e PDF (agentes *Miner*), acesso ao sistema operacional e listas de tarefa (agente *Tasker*) e construção de perfis de usuário (agente *Profiler*).
- Colaboração: na implementação, funcionalidade de chat foi construída em um agente de interface. Uma melhor solução seria ligar os agentes a sistemas externos de mensagens.
- Awareness e Matchmaking: os agentes se comunicam com outras coteries e trocam informações sobre seus usuários, comparando perfis e buscando semelhanças (agente *Matcher*).
- Interface de usuário: um agente de interface exibe informação sobre os outros usuários, de acordo com o potencial para colaboração. Uma interface alternativa de radar também foi construída, para permitir a visualização de proximidade entre iguais.

5.3.1 Perfis de Usuário

Navigator implementa 4 agentes na camada de GC: *Miners* de PDF e MS Word, um *Tasker* e um *Profiler*. Quando o ambiente é inicializado pela primeira vez, agentes constroem perfis iniciais buscando e analisando toda a informação fornecida pelo

usuário. O primeiro passo é processar documentos de texto e redes de contato. Os seguintes processos são realizados:

- Processamento de documentos textuais: os agentes *Miner* processam documentos usando técnicas de recuperação de informações. Palavras-chave são extraídas de cada documento e relações entre documentos são estabelecidas. Navigator implementa TFIDF, um algoritmo de extração de palavras-chave, descrito na seção anterior.
- Processamento da lista de tarefas: o agente *Tasker* coleta informações de lista de tarefa do sistema operacional para determinar:
 - Quanto tempo o usuário gastou em um recurso ou tarefa;
 - Há quanto tempo este recurso ou tarefa foi acessado pela última vez.
- Construção de perfis: o agente *Profiler* recebe as informações extraídas do ambiente e aplicações externas e as consolida em perfis de usuário.

Quando o usuário está trabalhando, seus comportamentos são observados para extrair informações adicionais. Enquanto trabalha, agentes observam seu comportamento e documentos manipulados e quando são acessados. As seguintes informações adicionais são extraídas do comportamento dos usuários:

- Tempo gasto em documentos: o tempo total gasto em cada documento é medido;
- Frequência de acesso: o sistema acompanha quantas vezes o documento foi acessado, e quando foi o último acesso.

Os interesses dos usuários mudam com o tempo, logo o que é útil ou necessário em um momento pode não ser no próximo. Para lidar com este aspecto, cada perfil de usuário tem duas dimensões: uma dinâmica, de curto prazo, e uma estável de longo prazo. Estas são construídas com base em uma análise de documentos e atividades dos usuários, como mostrado na Figura 5-5. Tanto a dimensão de longo prazo quanto a dimensão de curto prazo são calculadas como uma função do tempo gasto com os recursos.

Um perfil contém um ID do usuário no sistema, um nome ou alias e uma lista das suas áreas de habilidade. Estas áreas de habilidade são calculadas a partir do conteúdo dos documentos. Palavras-chave são extraídas dos recursos de trabalho do usuário (documentos, arquivos pdf, páginas da web, emails, etc.), e suas contagens de frequência são verificadas, usando a técnica de recuperação de informações descrita no capítulo 2.

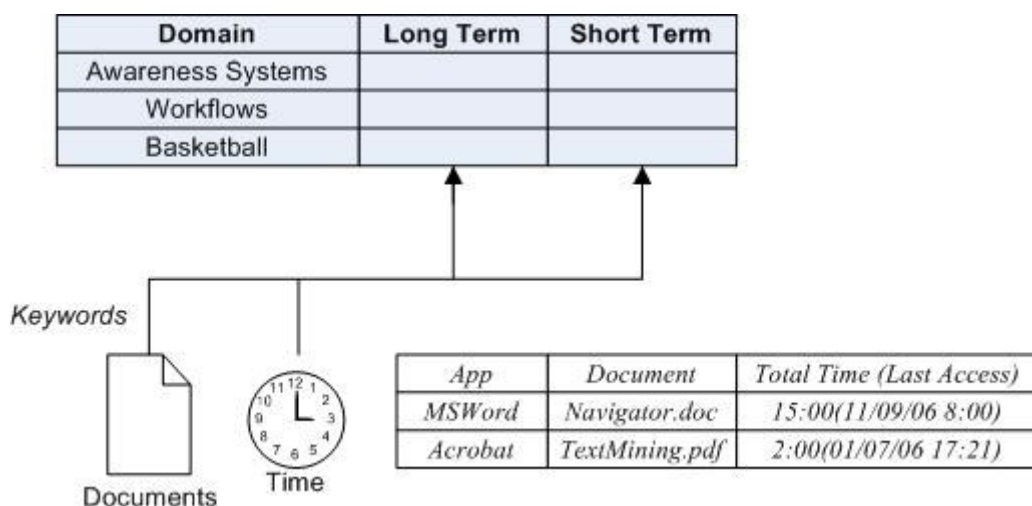


Figura 5-5: Perfil de Usuário usado em Navigator

Documentos são agrupados de acordo com a co-ocorrência de palavras-chave, para determinar “áreas de atividade” gerais. Na maioria dos sistemas, apenas estes dados de palavras-chave são usados para construir perfis de usuário. Em nosso sistema, nós também levamos em conta o tempo gasto em cada documento, que é usado para calcular as habilidades do usuário e o nível de interesse em cada assunto (respectivamente, o as dimensões de longo e curto prazo do perfil). Estes são obtidos por um agente especializado que anota a atribuição de tempo do usuário enquanto no este trabalha no computador. A lista de atividades é uma lista de aplicações e documentos usados pelo usuário, o tempo acumulado de uso em cada um e o último momento em que eles foram acessados pelo usuário.

A dimensão de longo prazo do perfil reflete a habilidade do usuário em determinado domínio. O tempo acumulado gasto com estudo ou trabalho em documentos deve levar a uma estimativa de quão bem informado ele está nas áreas às quais o documento pertence. Assim, a dimensão de longo prazo é cumulativa, e representa a experiência acumulada do usuário em certo assunto. Os perfis são enviados a outros usuários através da rede.

Supomos que os padrões de acesso do usuário a recursos refletem seus interesses correntes ou trabalho em andamento. No entanto, interesses e prioridades mudam frequentemente (por exemplo, quando projetos começam ou acabam, ou indivíduos se mudam). A informação que é valiosa ou interessante em um momento pode não o ser no próximo. Assim, a dimensão de curto prazo é passageira, e os valores designados a cada um dos territórios decaem com o tempo, representando a mudança de interesses e foco de atenção do usuário. Quanto mais um documento é o objeto da atenção do usuário, o mais “quente” ele se torna. Acesso repetido ao mesmo recurso indica interesse renovado ou continuado no documento ou assunto.

Todos dados relacionados aos domínios de conhecimento são rotulados de acordo com quão recentemente foram usados e quanto ao tempo gasto neles. Cada vez que um usuário acessa um domínio diferente do seu perfil, este é reforçado em suas dimensões de longo e curto prazo. Desta forma, tanto as habilidades quanto os interesses são calculados como uma função do tempo. Na implementação atual, cada recurso tem um peso de acordo com seu tempo total de uso. Deste modo, palavras de um documento que foi usado mais frequentemente são considerados melhores identificadores de habilidade que os de outros documentos.

Para calcular o perfil de interesse, documentos são ponderados pelo tempo de uso, mas os pesos se deterioram de acordo com o tempo gasto desde o último acesso. Assim, estes valores diminuirão com a passagem do tempo, representando interesse decadente. No entanto, se o usuário volta aos recursos depois de um tempo, o interesse é outra vez renovado. Acreditamos que uma análise temporal de atividade deve ser mais útil do que uma baseada em palavras-chave simples, visto que ela permite extrair mudanças em interesses ou necessidades de trabalho. Em nossa abordagem inicial, supusemos que 1 hora de trabalho equivale a 24 horas de interesse, isso é, para cada hora gasta em um documento, o interesse associado durará 24 horas. Assim, se um indivíduo gastasse 3 horas num documento, seu interesse seria reduzido a zero após 3 dias.

5.3.2 Awareness e Matchmaking

A identificação de uma oportunidade para colaboração é um processo realizado pelo agente *Matcher*. Este processo começa por achar indivíduos em contextos semelhantes (é melhor não muda os contextos de trabalho do usuário) e então achando

recursos que estes usuários talvez possam compartilhar. Uma vez um elo tenha sido estabelecido, usuários podem ser direcionados a colaborar sincronamente e a trocar idéias sobre o trabalho em si.

Navigator tenta identificar possíveis colaboradores. Para fazê-lo, o sistema processa perfis recebidos de diferentes pares e calcula as semelhanças entre eles e o perfil do seu usuário. O sistema então mostra os outros usuários em uma lista com as respectivas relações de semelhança calculadas. A semelhança é calculada utilizando técnicas de recuperação de informações já descritas.

Este cálculo capacita a extração dos termos que melhor definem uma interação (composto por um conjunto de mensagens textuais) e documentos (recursos) sendo manipulados pelo usuário. A combinação dos dois pode ser executada via uma comparação de vetores de palavras-chave representando os diferentes documentos para calcular sua semelhança. A medida de co-seno é freqüentemente utilizada para este fim: o co-seno do ângulo entre os dois vetores é calculado e quanto mais próximos os vetores mais semelhantes os documentos (Baeza-Yates e Ribeiro-Neto, 1999). Desta forma, torna-se possível interligar recursos e indivíduos por sua manipulação de artefatos e inserí-los em suas respectivas esferas de trabalho.

Quando dois usuários trabalham (lendo, escrevendo, buscando) em assuntos semelhantes, o sistema deve indicar proximidade, visto que seus focos atuais (atividades) são semelhantes. Perfis de longo prazo indicam quem são os especialistas em certo assunto. Pela comparação de perfis de longo prazo, o sistema pode determinar se há outros usuários que sabem sobre o assunto em que um usuário está trabalhando. Embora um especialista talvez não esteja atualmente trabalhando no assunto, o sistema ainda pode usar informação histórica para compará-lo a outros.

Ter um perfil bidimensional permite ao sistema executar diferentes tipos de comparações, que correspondem a diferentes oportunidades para colaboração. Quando se considera dois indivíduos, um usuário é o ponto de base de comparação e cada par é um colaborador potencial. Vislumbramos as seguintes oportunidades para colaboração (resumidas na Tabela 5-1):

- O perfil de curto prazo do usuário, quando combinado com um perfil de curto prazo de outro usuário: representa uma possibilidade de colaboração e trabalho compartilhado, visto que ambos os indivíduos

têm necessidades imediatas e trabalham nos mesmos assuntos. Isto pode levar à cooperação ou divisão de trabalho;

- Um o perfil de curto prazo do usuário, quando combinado com um perfil de longo prazo de outro: o usuário está trabalhando ativamente no assunto, e o outro trabalhou no assunto antes e tem experiência que pode ser valiosa ao trabalho em andamento. O usuário pode pedir auxílio com o trabalho ao seu par;
- Um perfil de longo prazo do usuário, quando combinado com um perfil de curto prazo de outro: o usuário está em posição de ajudar alguém que está trabalhando em um assunto em que ele trabalhou antes. Isto é uma oportunidade para o usuário oferecer seus serviços como um consultor;
- Um perfil de longo prazo do usuário, quando combinado com um perfil de longo prazo de outro: ambos trabalharam com o assunto antes (e possivelmente não juntos). Podem compartilhar experiências e lições aprendidas, discutir alternativas e escolhas feitas, refletindo sobre elas e talvez chegando a novas conclusões se o problema surgir novamente. Estas conversas podem ser recursos valiosos para trabalhadores que estão atualmente atacando estes problemas.

Tabela 5-1: Tipos de oportunidades gerados pela combinação de dimensões distintas de perfis

	Igual	
Usuário	<i>Curto prazo</i>	<i>Longo prazo</i>
<i>Curto prazo</i>	Cooperação	Solicitar ajuda
<i>Longo prazo</i>	Fornecer ajuda	Conversa

A semelhança entre perfis dos usuários representa a semelhança entre seus interesses e habilidade. A semelhança é calculada usando a medida de co-seno de distância entre um perfil de curto prazo do usuário e o longo e perfis de curto prazo de cada um dos seus pares. Este cálculo fornece um valor entre 0 e 1, onde quanto mais próximo a 1, mais semelhante são os perfis dos usuários (quanto mais próximos a 0, mais diferentes os perfis dos usuários). A maior vantagem deste modelo é a utilização do tempo gasto em cada documento, que permite privilegiar termos que são mais freqüentemente ou mais recentemente acessados (dependendo do que procuramos).

5.3.3 Colaboração e Interface

Navigator implementa um agente para exibir a informação reunida e cálculos do sistema, assim como controlar a comunicação direta com outros usuários por uma interface de messenger.

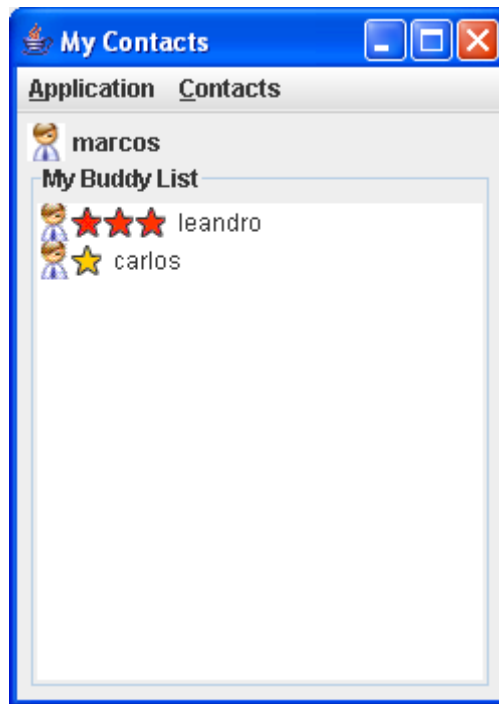





Figura 5-6: Lista de contatos

A interface principal é semelhante a uma lista de contatos de aplicações e mensagens instantâneas (Figura 5-6) e mostra usuários e níveis de semelhança calculados. Semelhança recente (trabalho em andamento) é representada pelo número de estrelas, onde quanto maior o número de estrelas, mais alta a semelhança (ver Tabela 5-2). A habilidade é representada pela cor da estrela: amarelo, laranja ou vermelho, indicando níveis de habilidade crescente, como visto na Tabela 5-3. Os usuários podem contatar uns aos outros usando uma funcionalidade de chat do sistema.

Tabela 5-2: Semelhança de trabalho recente

O número de Estrelas	Semelhança
★	0% - 25%
★★	25% - 50%
★★★	50% - 75%
★★★★	75% - 100%

Tabela 5-3: Semelhança de habilidade

Cor da Estrela	Nível de habilidade
	0% - 33%
	33% - 66%
	66% - 100%

O sistema também tem uma visão de radar, mostrada na Figura 5-7, que exibe o usuário no centro e os outros ao redor dele, com proximidade indicando proximidade de trabalho e cor da camisa indicando habilidade (como mostrado acima). Isto não só permite a visualização de proximidade com o usuário, mas também entre outros usuários e permite aos usuários diferenciar entre dois contatos que pode estar à mesma distância.

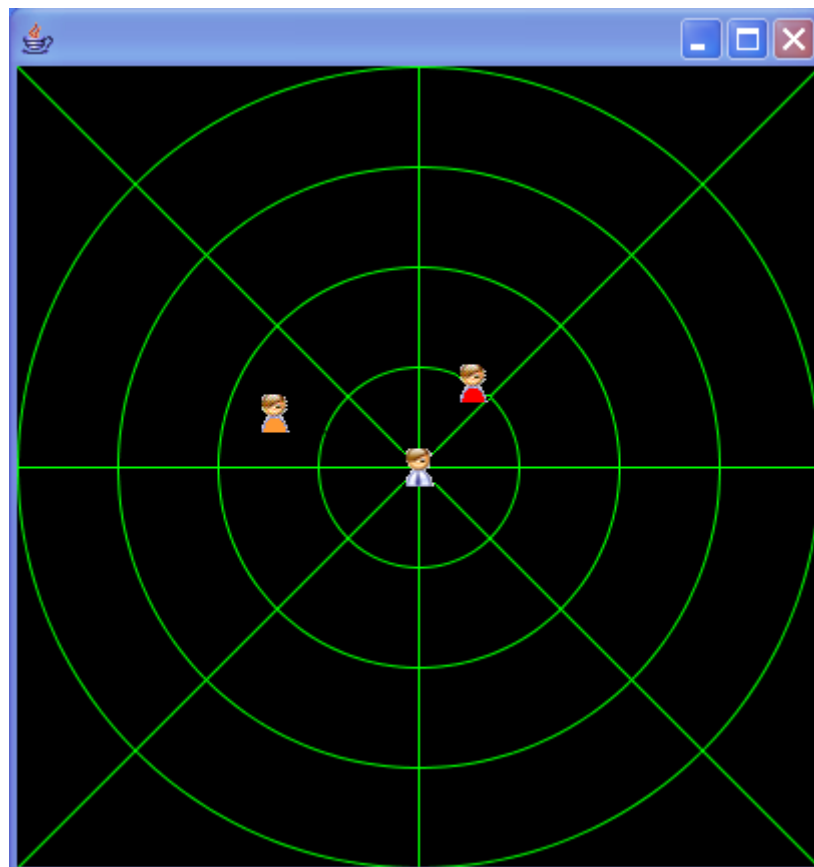


Figura 5-7: Visão de radar

A Figura 5-8 mostra uma visão geral da interface do Navigator, mostrando a lista de contatos, onde, para cada contato, uma lista de documentos também é mostrada. Estes são os documentos sendo manipulados pelos outros usuários nesse momento. Isto contribui para aumentar consciência de atividades dos outros usuários, mostrando no que eles trabalham que é relacionado ao usuário. A interface de chat também é

mostrada, que é o canal de comunicação fornecido para que dois usuários possam entrar em colaboração.

5.3.4 Detalhes técnicos

Navigator foi implementado usando a estrutura COPPEER (Miranda et al, 2006). COPPEER fornece uma infra-estrutura para facilitar o desenvolvimento de aplicações baseadas em agentes peer-to-peer. Os agentes COPPEER podem agir em paralelo e reagem às ações de outros agentes. O comportamento coletivo é o resultado das ações individuais de cada agente. A aplicação foi escrita em Java, e usa Lucene (LUCENE, 2007) para indexação e busca de texto, PDFBox (PDFBOX, 2007) para acessar arquivos PDF e Jacob (JACOB, 2007) para arquivos do MS Office.

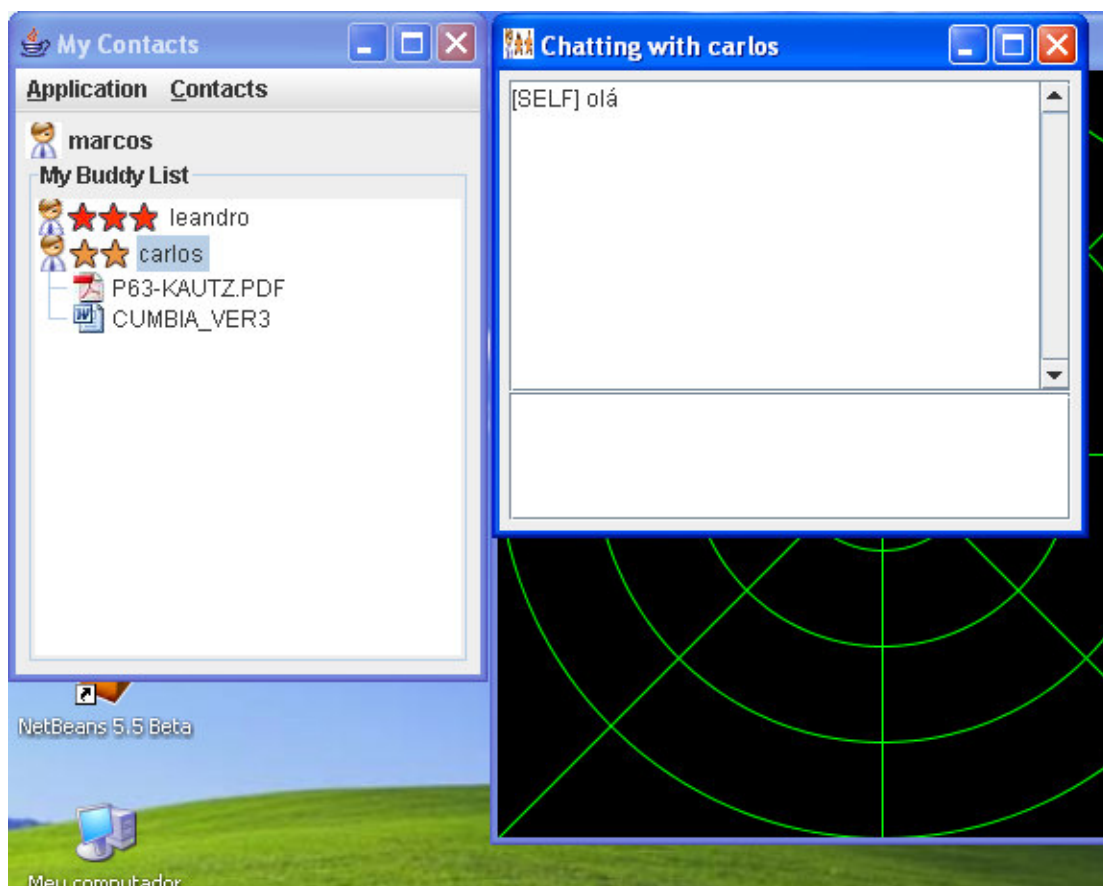


Figura 5-8: A visão geral da interface Navigator: janela de lista de contatos, janela de chat e radar (parcialmente mostrado)

Navigator permite uma flutuação constante de usuários. Cada usuário administra a própria conexão com outros iguais, sem controle central. Navigator implementa cinco tipos de agentes, que já foram mencionados em seções prévias e são listados aqui: agentes mineiradores de texto (*Miner*), coletores de tarefa (*Tasker*), construtores de perfil (*Profiler*), matchmakers (*Matcher*) e interface (*Interfacer*).

- *Miners* analisam documentos dos usuários e extraem os dados para construir perfis. Estes agentes são especializados de acordo com tipos de documentos: atualmente existem agentes para ler MS Word (*DocMiner*), PDF (*PdfMiner*), e arquivos em texto simples (*TxtMiner*). Os agentes criam diferentes índices e os combinam em um índice completo. Este índice consolidado será usado pelo agente construtor de perfil (*Profiler*) para criar os perfis dos usuários.
- O agente *Tasker* controla os processos correntes buscando informações diretamente do sistema operacional, capturando informação sobre a execução de aplicações e uso de recursos. Esta é a informação de atividade usada para calcular perfis dos usuários.
- O agente *Profiler* agrega as atividades e frequência de palavras-chave geradas por *Miners* e *Tasker*, em um perfil, calculando as dimensões de longo (habilidade) e curto (interesse) prazo. Estas são montadas usando uma função de agregação, como descrito anteriormente.
- *Matcher* é responsável por comparar o perfil do seu usuário com outros perfis de usuários. Ele gera compara perfis de usuários gerados pelo *Profiler* e perfis que vêm da rede. Duas comparações diferentes são calculadas para cada par: um perfil de curto prazo do usuário contra os perfis de curto e longo prazo de outro usuário. Em cada caso, isto fornece duas medidas de proximidade, e cada par será categorizado deste modo.
- O agente *Interfacer* coordena a interface gráfica e a comunicação com o usuário. Ele usa os cálculos feitos pelo agente *Matcher* para fornecer uma visualização de oportunidades para colaboração. A Figura 1-8 mostra a interface de usuário: há uma lista de contatos, onde cada contato é associado um conjunto de estrelas coloridas. O número de estrelas indica a proximidade do perfil de curto prazo (possibilidade de divisão de trabalho), e as cores indicam a avaliação de habilidade (possibilidade de consulta com um especialista). Um duplo clique em um contato exhibe sua lista atual de atividade. Uma visualização alternativa é fornecida: uma visão de radar onde o

usuário é o centro, cercado por seus contatos. A proximidade indica proximidade em atividades recentes, cores indicam habilidade.

Devido à necessidade manter o perfil do usuário atualizado, o *Miner* e *Tasker* executam continuamente, registrando informação em intervalos regulares. O intervalo é customizável pelo usuário (o default é 30 segundos). Os dados são enviados de um agente para outro, e o perfil é atualizado deste modo. A Figura 5-9 mostra a troca de mensagens entre agentes Navigator. O usuário interage somente pela interface, onde os nomes dos usuários e graus de semelhança são mostrados. Pela interface, usuários coletam conhecimento sobre outros em seu domínio e a possibilidade deles ajudarem com estudos, problemas ou tarefas. Os agentes são pró-ativos em suas recomendações e na construção de perfil. Quando há uma possibilidade de colaboração, isso é mostrado sem intervenção de usuário.

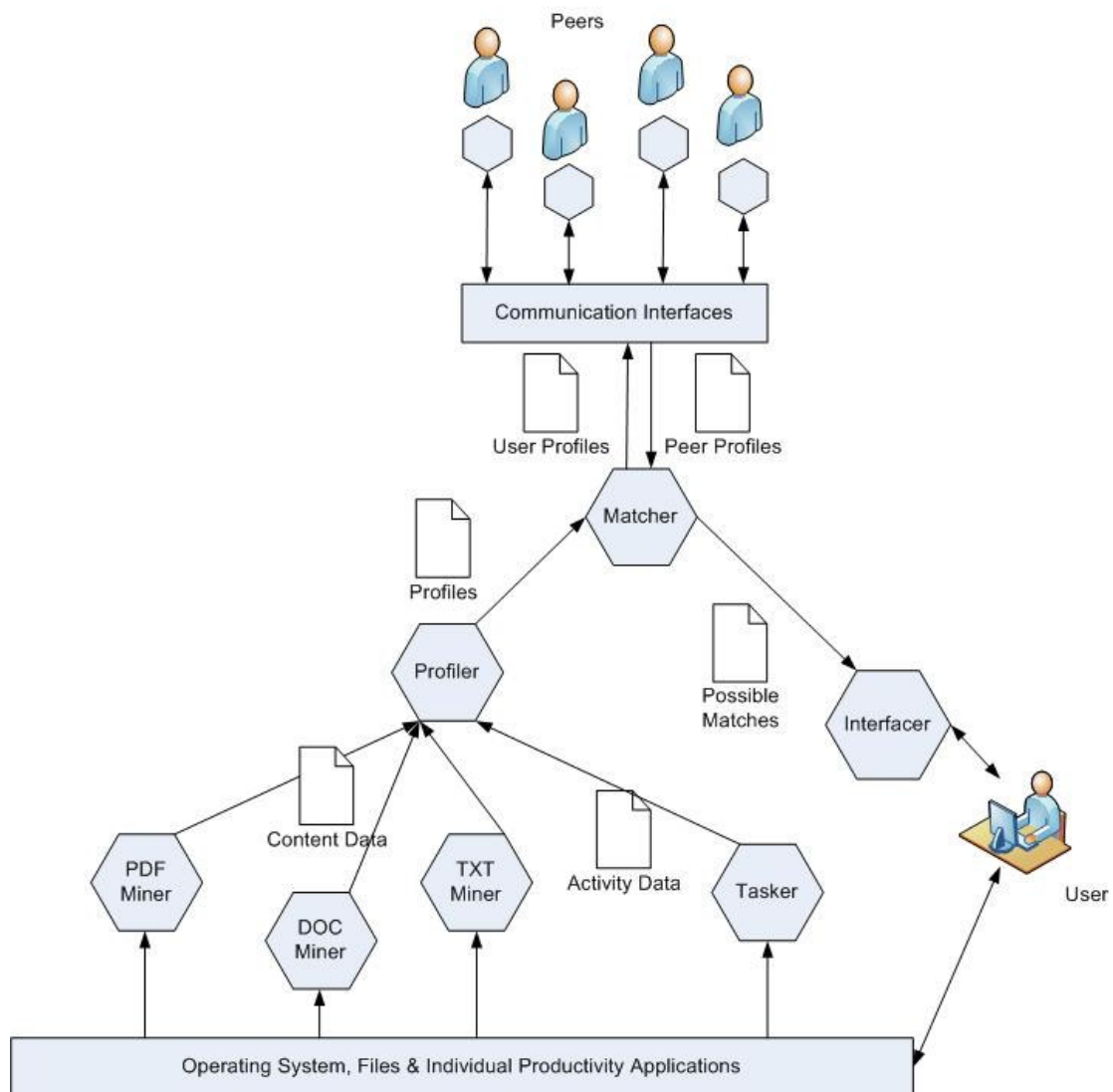


Figura 5-9: Arquitetura de Agentes de Navigator e sua Comunicação

5.4 Lições Aprendidas com CUMBIA e Navigator

Um das vantagens de Navigator é a forma na qual a informação é processada para criação de perfis de usuário. Como em muitos sistemas semelhantes (FONER, 1997) (RESNICK, VARIAN, 1997), agentes Navigator procuram nos documentos do usuário termos chave e suas frequências. No entanto, Navigator estende essa aproximação para levar em conta o tempo gasto nestes recursos e quão recentemente foram usados. Isto altera o cálculo da relevância de cada documento, para refletir o foco da atenção do usuário.

A meta de Navigator era apoiar colaboração oportunista fracamente estruturada, em cenários remotos, para facilitar interações não planejadas. Os agentes localizam colaboradores potenciais baseados numa análise de atividades de seus usuários. Mesmo sem nenhuma experiência formal ter sido conduzida, alguns problemas e observações foram recolhidos que ajudam diretamente nossa pesquisa. Um dos problemas encontrado ao rodar o sistema foi o número grande de mensagens enviadas e recebidas entre agentes. Mesmo com apenas 5 usuários, o sistema rapidamente causou overflow de memória e parou. O sistema foi rapidamente inundado com mensagens, dado o grande número de dados sendo transmitidos. Isto é uma dificuldade técnica que pode ser remediada através de configuração do comportamento de envio de mensagens (quão frequentemente mensagens são enviadas, etc.).

Uma reflexão sobre a experiência com CUMBIA e Navigator nos leva a crer que estavam abertos demais. O fato de que Navigator busca qualquer um numa rede potencialmente grande levará a um grande número de comparações e resultados, gerando um overflow de informação que um usuário talvez tenha dificuldade em manipular. Uma grande quantidade de informações não seria muito útil ao usuário, pois gera necessidade de análise manual por parte do usuário.

Navigator parte de artefatos para chegar a pessoas, que é uma estratégia frequentemente usada. Acreditamos que esta estratégia não funcionaria tão bem para colaboração em andamento, pois não analisa o envolvimento de contatos com os recursos sendo analisados, focalizando só nas atividades e interesses. Enquanto pode ser boa para achar novos contatos (buscar as pessoas sem levar em consideração seus relacionamentos prévios com outros usuários), esta estratégia não é necessariamente a melhor para colaboração em andamento. Nesta situação, parece mais apropriado

procurar inicialmente por colaboradores e então pelos os artefatos manipulados usado em sua colaboração. Com isto em mente, nós passamos para projeto de um método que colocasse maior ênfase na colaboração em andamento, selecionando contatos e informação que seriam de interesse ao usuário dadas as atividades em grupo.

Expériences avec systèmes pilotes (Navigator et CUMBIA) ont montré que l'architecture basée sur agents est une option viable. Par contre, nos expériences ont montré que la quantité d'informations est trop grande pour l'utilisateur et qu'une méthode de réduire cette quantité est nécessaire. Dans le prochain chapitre, nous présentons un modèle et une méthode de raisonnement d'un agent assistant.

6 Uma Abordagem Computacional para a Descoberta de Esferas de Colaboração

Pour être capable d'aider son utilisateur dans un environnement dynamique en réseau, un agent assistant doit avoir un modèle actualisé de le réseau de contacts de son maître. Le système doit aussi être capable de déterminer quels parties du réseau de son maître son actives (quelle est la situation de chaque liaison) et de quoi s'agit chaque liaison. Le modèle doit indiquer qui sont les collaborateurs courants d'un utilisateur et quel est son contexte coopératif partagé.

Para auxiliar um usuário em um ambiente dinâmico, ligado em rede, um sistema inteligente necessita ter um modelo atualizado da rede de contatos do usuário. O sistema deve também ser capaz de determinar as partes desta rede que estejam ativas (qual o estado de cada ligação), e os interesses de cada laço. Este modelo deve poder indicar quem são os colaboradores atuais do usuário e qual o seu contexto colaborativo compartilhado.

Inspirados pelo conceito de *esfera de trabalho* definido por Gonzalés e Mark (2004), nós definimos uma *esfera de colaboração* para representar uma estrutura que une porções ativas da rede intensional do usuário e os temas e recursos relacionados a elas. É semelhante à noção de *esfera de trabalho*, que une eventos envolvendo um grupo de pessoas com um motivo comum e recursos, sob o ponto de vista de um único usuário. Uma esfera de trabalho representa uma visão individual de um projeto colaborativo, preferências, hierarquia, etc. Uma esfera de colaboração faz a ponte entre várias esferas de trabalho, fornecendo uma visão unificada de um projeto colaborativo. Esferas de trabalho pertencem a indivíduos diferentes, mas se relacionam ao mesmo esforço conjunto, tal que quando uma atividade acontece em uma esfera de trabalho, esta pode ser correlacionada a outra pessoa trabalhando em sua esfera de trabalho através da esfera de colaboração que as une. Deste modo, eventos envolvendo um só membro de um grupo podem ser amarrados a outros membros do grupo e suas esferas

de trabalho. Uma visão esquemática de esferas de colaboração é mostrada na Figura 6-1. Uma esfera de colaboração liga elementos de diferentes esferas de trabalho, e é possível que estes sejam elementos que pertençam a uma esfera de trabalho, mas que devem permanecer privativos a um indivíduo. Da mesma maneira, indivíduos podem formar subgrupos dentro de uma esfera de colaboração e constituir sub-esferas de colaboração através destas interações.

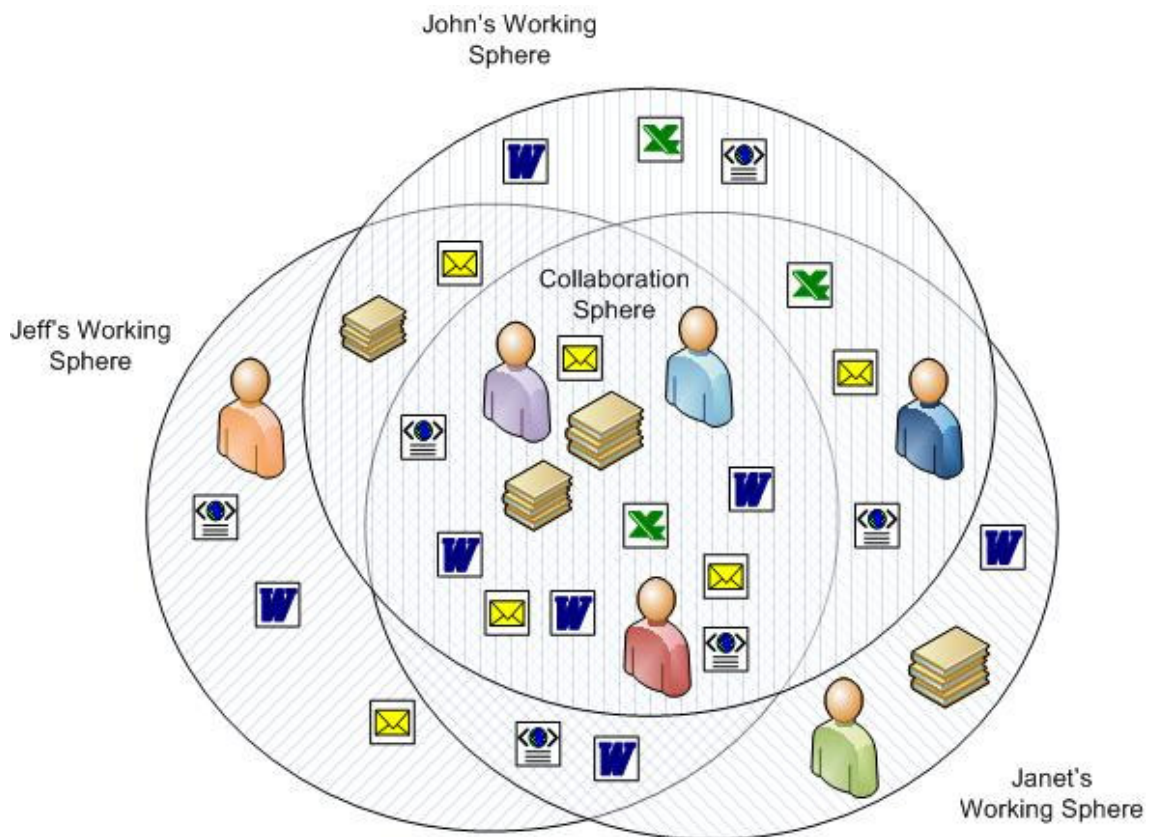


Figura 6-1: Esferas de Trabalho e Esfera de Colaboração

Uma esfera de colaboração une partes da rede intensional do usuário e os temas e recursos relacionados a estas com partes de redes de outros usuários. Ela descreve um relacionamento entre duas ou mais partes e contém informação sobre as pessoas, a organização e os aspectos de coordenação de um projeto sendo realizado através destes relacionamentos. Adicionalmente, ela contém informação sobre atividades anteriores, que estabelecem conhecimento comum e geram expectativas de comportamento individual. Isto ajuda os usuários a decidir com quem eles trabalharão e a ajustar seu comportamento de modo adequado.

Uma esfera de colaboração contém um registro de interações passadas e em andamento. Ela deve refletir o estado de cada laço, incluindo projetos em andamento,

com ênfase especial em variáveis que influenciam ações do usuário. As características de uma esfera de colaboração são:

- Estado: se a esfera está ativa ou dormente;
- Natureza: se é pessoal ou negócio;
- Formalidade: se é formal ou informal;
- Participantes: pares envolvidos no projeto;
- Estrutura de projeto: descreve as estruturas de organização adotadas para o projeto. Estas podem ter sido definidas formalmente (designadas pela organização) ou distribuídas informalmente (voluntários escolhem as tarefas que vão executar).
 - Subdivisão de tarefas: como as tarefas estão divididas entre participantes;
 - Interdependências: onde estão as interdependências entre as tarefas;
 - Estrutura hierárquica: a quem um ator deve informar quando seu trabalho esteja terminado? A quem devem os resultados ser enviados?
 - Designação de papéis: quais são os papéis neste projeto e quem está em que posição?;
 - Trabalho em andamento: como o trabalho está progredindo dentro do projeto de grupo?
- Restrições: que restrições existem para este contexto (ex. sigilo, NDAs, prazos finais, ferramentas disponíveis, etc.)
- Temas: com o que são os assuntos este negócios de contexto?
- Metas: o que são as metas desta colaboração?;
- Locais: descrevem os locais e meios através dos quais a colaboração é empreendida.
 - Sítios (espaços ou ferramentas compartilhados e meios de comunicação adotados);

- Meios (artefatos e recursos compartilhados).

Estas variáveis foram extraídas de elementos contextuais descrito por Borges e colegas (Borges et al., 2004), elementos descritos como influências a tomada individual de decisão em ambientes colaborativos descrito por Bernoux (1995), e de nossas observações nos estudos de casos. Elas são mostradas na Figura 6-2.

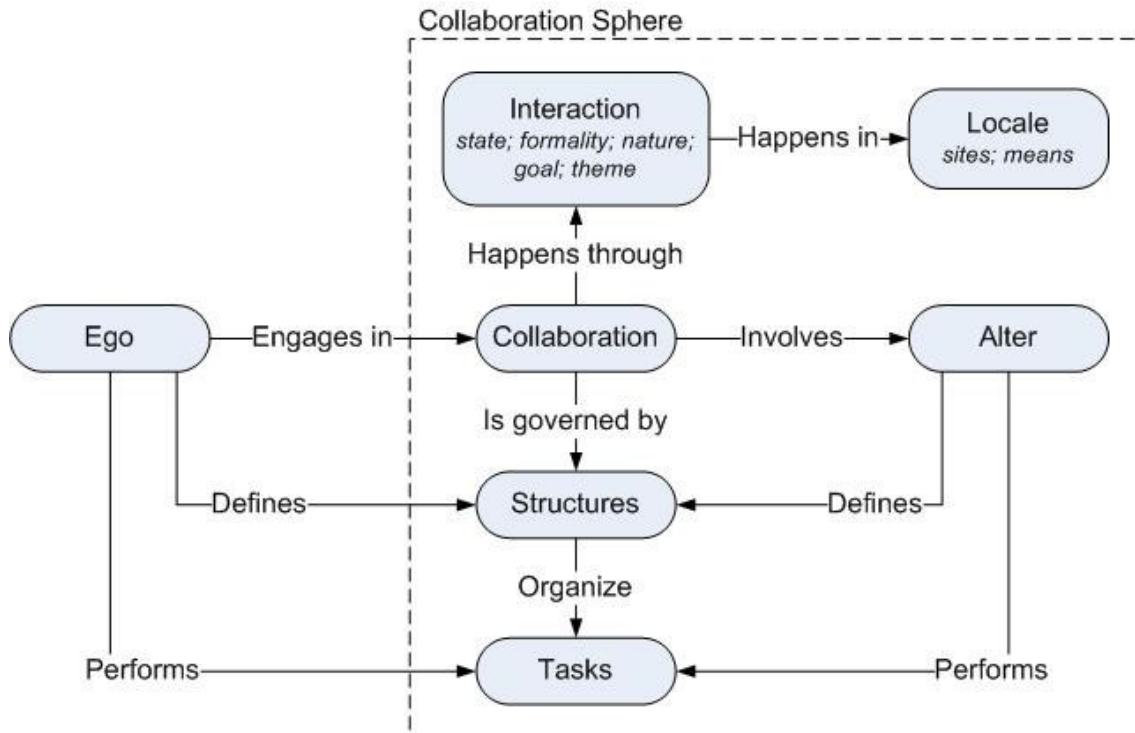


Figura 6-2: Ontologia de conceitos da esfera de colaboração

Um laço é um conceito rico e individualizado, que reflete o relacionamento entre dois indivíduos. Ele existe por si próprio, independente de trabalho ou esferas de colaboração e é construído ao longo do tempo, o que significa que ele inclui históricos de interação e preferências pessoais ou afinidades. Um laço pode ter várias esferas de trabalho associadas a ele, visto que um usuário pode ter colaborado com uma pessoa mais que uma vez, e cada uma destas descreveria um projeto em conjunto. Cada pessoa também tem certas características individuais que refletem seus níveis de habilidade e capacidades, segundo a visão do usuário. Deste modo cada usuário deve manter um modelo de cada um dos seus pares para ser capaz de melhor navegar sua rede intensional.

Características tais como natureza, formalidade e meios de comunicação também podem ser vistas como características do laço ao invés de um contexto particular. Muitos relacionamentos têm tanto uma faceta pessoal quanto uma de

negócio, e maiores pesquisas devem ser realizadas para determinar como as pessoas administram estes aspectos e até que ponto uma esfera de colaboração é influenciada por eles.

Em ambientes ligados em rede, indivíduos necessitam gerenciar as diferentes colaborações em que estão envolvidos, seus papéis em cada e restrições de organização às quais estão sujeitos. Administrar múltiplas colaborações envolve um entendimento dos ambientes sociais, capacidade de alocar pessoas e recursos a cada um destes, e de identificar oportunidades para interação. Dois desafios que se apresentam aos indivíduos trabalhando nestes cenários são lembrança e comunicação. Por exemplo, lembrar-se que uma atividade ou um documento tem ligação com um grupo particular e comunicar-se com os outros sobre atividades fortalece laços individuais e mantém o fluxo do trabalho de grupo. Para facilitar a alternância entre várias colaborações, contextos sociais e atividades necessitam ser interligados, pois isto capacitaria que a informação fluísse, tornando o contexto mais fácil de compreender pelos participantes. Isto é realizado através das esferas de colaboração.

Neste cenário, certas situações podem constituir uma oportunidade para fortalecer um laço, ou é possível que haja informação pertinente a colaborações em andamento que poderiam ser transmitidas a outros colaboradores. Detectar e apoiar relacionamentos envolve determinar o estado de um laço (se ativo ou dormente) e que artefatos e informações estão envolvidas nesta colaboração.

Apoio ao trabalho fracamente estruturado, em condições dinâmicas, envolve manter o usuário a par do trabalho em andamento, e lhe fornecer informações que viabilizem sua tomada de decisão e adaptação a situações inesperadas, ajudar o usuário a se lembrar das suas múltiplas esfera de trabalho e colaborações e se comunicar com os outros conforme necessário. Nesta seção, nós apresentamos nossa abordagem à identificação de esferas de colaboração através de análise de interação. Uma esfera de colaboração é um conceito rico que envolve vários elementos que descrevem arranjos colaborativos. Está além do escopo desta tese tentar inferir todos eles. Tentamos determinar os dois elementos fundamentais, as pessoas (quem) e artefatos (o que) envolvidos em um projeto colaborativo. A informação que apóia adaptação e auto-organização diz respeito a eventos relacionados aos elementos contidos em cada esfera de colaboração, especialmente as pessoas e artefatos que estão envolvidos nele.

6.1 *Constituindo Esferas de Colaboração*

Nosso objetivo nesta fase é desenvolver um método para construção de esferas de colaboração. Concebemos um método em dois níveis para alcançar essa meta, descobrindo colaboradores e unindo estes a recursos ou atividades (o que envolve manipulação de informação). Tendo estes elementos à mão, torna-se possível analisar atividades dos usuários e atribuí-las a esferas de colaboração como apropriado, distribuindo informação de eventos conforme necessário. O método apresentado é baseado em nossas experiências com os projetos CUMBIA e Navigator, descritos na seção anterior. Testes iniciais com CUMBIA e Navigator rapidamente levaram à compreensão de que mecanismos de filtragem eram necessários para reduzir a quantidade de informação coletada e distribuída. Com este método, damos um passo na direção da meta de ajudar a administração de laços para apoio de colaboração fracamente estruturada, através da identificação de esferas de trabalho e unindo-as a esferas de colaboração, focando nos colaboradores correntes e fornecendo informação relacionada a tarefas em conjunto.

No restante deste capítulo, nós supomos sistema distribuído, baseado em pares (como descrito no capítulo prévio), ligando os usuários a seus pares, para provisão de awareness relacionada a esferas de colaboração ativas. Neste cenário, cada usuário é ajudado por uma coterie que analisa interações e atividades dos usuários e troca informações com outros peers, criando um fluxo de informação para manter membros de grupo informados de atividades em andamento. Assim, a meta de cada coterie é *manter consciência entre pares através da distribuição de informação sobre as atividades de seus parceiros*. Para alcançar esta meta, os agentes:

1. Coletam informação gerada pelo usuário enquanto trabalha no seu computador;
2. Trocam informação com outros usuários; e
3. Fornecem informação ao usuário sobre atividades dos seus alters.

Dada a grande quantidade de informações disponíveis, mecanismos de filtragem devem ser utilizados para selecionar a informação a ser distribuída e exibida. Os agentes se focam em esferas de colaboração ativas.

Visto que a comunicação está na base de qualquer colaboração e usuários em situações colaborativas fracamente estruturadas se coordenam principalmente através de

intercâmbios lingüísticos (Zacklad, 2003), elegemos a análise de interações como a pedra fundamental de nossa abordagem. Esperamos que isto funcione especialmente bem em casos onde usuários se valem pesadamente de email, messenger e outros meios de comunicação para estabelecer contratos e realizar trabalho, como é o caso em grupos distribuídos.

As interações são analisadas estruturalmente e pelo conteúdo. Atenção é um recurso valioso e usuários a gerenciam ativamente. Dar atenção à informação ou mensagens implica que estas são importantes para eles, de modo que equiparemos atenção dispensada com importância e atividade. Os focos de atenção são determinados através de uma análise das interações correntes do usuário, e consideram que colegas que recebem mais atenção são aqueles com quem o usuário está colaborando ativamente. Apoiar colaborações fracamente estruturadas envolve a descoberta de laços ativos na rede e o trabalho ao qual eles se referem.

A comunicação ativa partes diferentes de uma rede (Nardi et al, 2002): trabalhos prévios mostraram que padrões emergem sobre tempo, e que algum destes padrões refletem colaboração mais próxima (Perer et al, 2005). Estas observações nos levam a uma abordagem baseada na análise de registros eletrônicos de interação (ex. registros de email, forum ou messenger). Uma limitação inerente desta escolha é que esta abordagem só funcionará nos casos em que a maioria das interações são realizadas utilizando meios de comunicação eletrônicos, deixando um vestígio que pode ser analisado computacionalmente.

Uma ontologia dos conceitos envolvidos é mostrada na Figura 6-3. Esta informação está organizada para representar relacionamentos em andamento e focos de interesse. O *usuário* é um indivíduo que executa seu próprio trabalho num ambiente computacional e interage com outros via email ou outros meios de comunicação por computador, e é o centro de análise. Um *laço* é um relacionamento entre dois usuários, construído no tempo, através de uma série de *interações*, que podem ser de trabalho ou pessoais, criando um relacionamento multifacetado. Uma interação envolve uma série de *troca de mensagens* (mensagens e suas *respostas*) entre dois usuários. Estes podem acontecer via email, messenger, fóruns, em pessoa ou no telefone. Mensagens podem conter *anexos*, que são também *recursos* manipulados por usuários e incluídos nas esferas de trabalho. Mensagens e recursos são descritos em termos de *palavras-chave*

que fornecem um tema geral e ajudam a estabelecer proximidade entre diferentes esferas de trabalho.

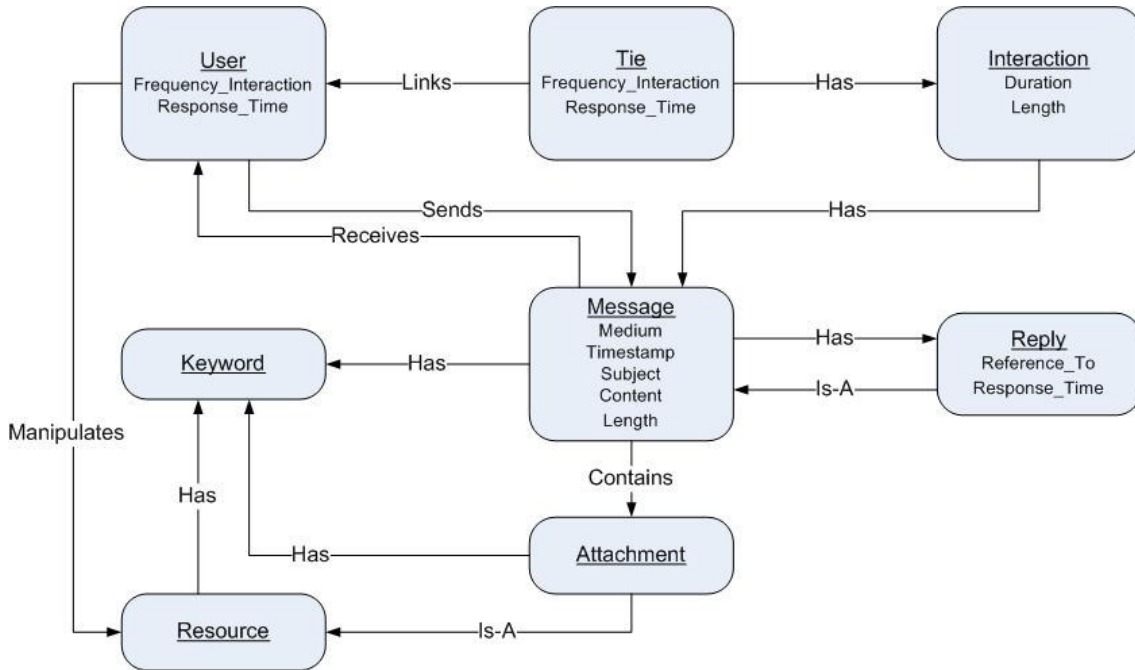


Figura 6-3: Ontologia de conceitos envolvidos na descoberta de esferas de colaboração

Uma rede de egocêntrica é construída ligando o usuário aos seus contatos. Esta rede pode ser vista como uma árvore com ego (o usuário) na raiz e seu ou alters (as pessoas que enviaram ou receberam mensagens de ego) no primeiro nível. Os recursos e informação relacionados às interações (temas, artefatos, etc.) são então adicionados como um segundo nível. A determinação de colaborações em andamento remete à determinação de que laços estão ativos. Para reduzir o espaço de busca, ramos no primeiro nível são pesados para representar intensidade de atividade. Para fornecer informações de atividade (para manter consciência), o sistema deve atribuir atividades e recursos às esferas de colaboração apropriadas, para que seja mais fácil acompanhar eventos que podem se relacionar ao trabalho em andamento.

A esta alocação de atenção do usuário chamamos de *foco de interesse*, e sua determinação é um processo de dois passos, mostrado em Figura 6-4:

1. Descobrir o que o usuário deve estar interessado em acompanhar (selecionando nós no primeiro nível); e
2. Compreender que atividades se relacionam a esta esfera de colaboração.

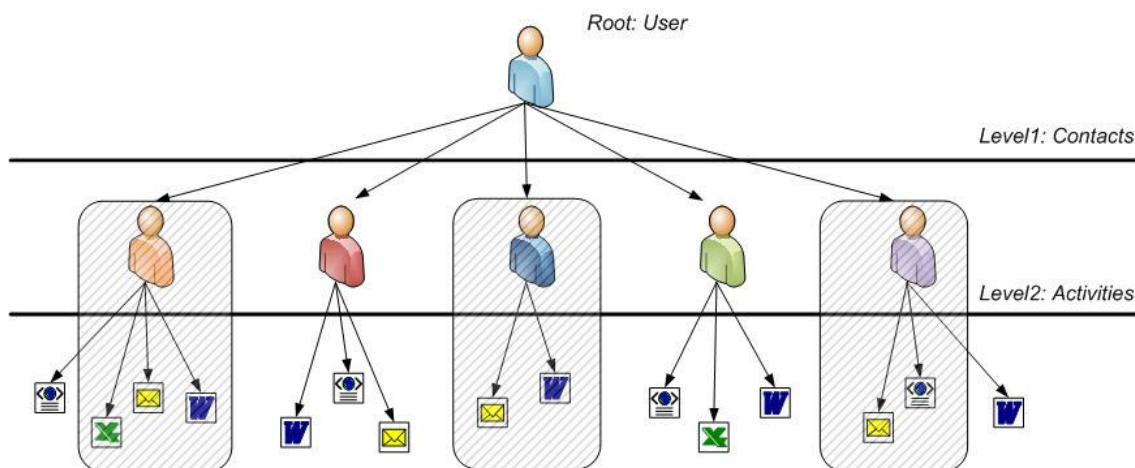


Figura 6-4: Busca em dois níveis para distribuição de informação

6.2 Interações Baseadas em E-Mail

Tendo em vista que email tornou-se um dos meios de comunicação mais freqüentemente usados, uma instanciação destes conceitos foi criada para interações baseadas em email. Neste cenário, informação sobre interações em andamento é reunida a partir de históricos de email e analisada para determinar as necessidades do usuário. Supomos que cada mensagem contém um remetente e um destinatário, data, assunto, conteúdo e anexos. Quando uma mensagem é uma resposta a outra mensagem, o tempo de resposta também é salvo, para verificar urgência. Na rede do usuário:

- *Ego* é o usuário (normalmente determinado extraindo-se os valores do campo *From* de emails enviados, ou informado manualmente) e
- *Alters* são os outros remetentes e destinatários de email (retirados dos campos *From*, *To*, *CC* e *BCC*), que formam a lista dos conhecidos do usuário.
- Cada *Mensagem* pode conter um ou mais *Anexos*
- *Interações* são compostas de mensagens de email e suas respostas (determinados via *Assunto*, *Message-ID* e *In-Reference-To* dos headers de mensagens) agrupadas, com cada resposta qualificada pelo tempo que o usuário levou para responder (extraído do campo *Data*). Uma interação é qualificada por seu comprimento (número de emails) e duração (tempo desde a primeira mensagem).
- Um *laço* é caracterizado pela freqüência de interação entre alters, i.e. quão freqüentemente trocam mensagens. Isto ajuda a determinar

comportamento “normal” em relação a cada alter, que pode ser usado para detectar aumento de comunicação indicativo de colaboração.

A Figura 6-5 mostra um modelo UML dos dados como extraídos de fontes para processar. Como esferas de trabalho e redes intensionais são controladas individualmente, adotamos um ponto de vista egocêntrico, analisando registros de interação de cada pessoa e partindo daí para os grupos com os quais que eles se relacionam.

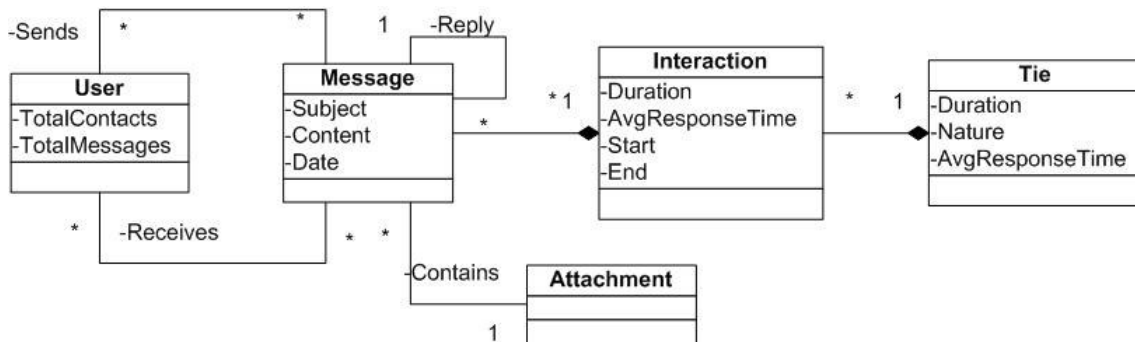


Figura 6-5: Modelo UML de interações de usuário, aplicado ao envio de mensagens de email

Frequência média de interação (quão frequentemente ele envia/recebe email) e tempo médio de resposta (quão rapidamente ego responde a qualquer email) também são calculados como um total e por alter (quão frequentemente ego envia email ao alter A e quão rapidamente responde a mensagens de alter B). Os grupos de co-trabalhadores são determinados pela co-ocorrência em headers de mensagem, e fornecem uma formação inicial de uma esfera de colaboração.

6.2.1 Descoberta de Colaboradores Ativos

O conjunto completo de conhecidos do usuário é extraído dos logs de interação (todos os contatos que enviaram ou receberam mensagens do usuário). O primeiro passo para descobrir os colaboradores é obter uma lista de todos colaboradores potenciais (*alters*), isto é, todos indivíduos com quem o usuário (*ego*) se comunicou em algum momento. Este conjunto forma a rede social pessoal completa do usuário, dos quais os colaboradores serão retirados. Estes podem ser obtidos através da extração de remetentes e destinatários (*to*, *from*, *cc*, *bcc*) de mensagens enviadas e/ou recebidas por ego. Cada alter é relacionado ao usuário através de um laço quantificado pelo número de mensagens enviadas de ego para alter e recebido de alter por ego.

Assim, definimos os seguintes conjuntos (em termos de interações de email) :

Senders = contatos encontrados no header FROM

Recipients = contatos encontrados nos headers TO, CC, BCC

E as seguintes variáveis:

Contato = qualquer indivíduo que envia ou recebe email

Sender (Msg_i) = contato encontrado no header FROM de uma mensagem específica

Recipient(Msg_i) = contato encontrado no header TO, CC ou BCC de uma mensagem específica

Este conjunto completo é frequentemente bastante grande e contém vários alters que podem ser descartados. Interpretamos cada mensagem recebida como uma tentativa de capturar a atenção do usuário, e cada mensagem enviada como um gasto ciente de atenção. Tyler e Tang (2003) sugerem que o conteúdo do Outbox é mais importante que o conteúdo do Inbox neste tipo de análise, uma vez que este reflete interações nas quais o usuário realmente decidiu participar. Assim, do conjunto inicial, retiramos os alters a quem ego não enviou mensagens (isso é, indivíduos que tentaram conseguir atenção de ego, mas não a receberam, então estão provavelmente envolvidos numa esfera de colaboração de prioridade mais baixa). A operação de *Descarte* é definida:

Discard(contact) :

if (*contact* ∈ *Senders*) ∧ (*contact* ∉ *Recipients*)

Discard = *True*

else

Discard = *False*

Depois de retirar alters que não receberam nenhuma atenção da rede, o comportamento “normal” de ego é inferido a partir de interações prévias. O ritmo médio de interação é calculado em termos de: número de mensagens enviadas por dia/semana/mês e tempo entre o recibo de uma mensagem e sua resposta. Estes indicadores são calculados para cada alter, e esta informação é adicionada a cada laço entre ego e alter. Esta informação também é calculada globalmente, para aquelas situações em que um novo alter aparece na rede (seja como remetente ou destinatário de mensagens) e não existem dados prévios com os quais avaliar o comportamento do ego em relação a eles:

MessageSent (ego) = Mensagem onde Ego é o remetente

MessageReceived (ego) = Mensagem onde Ego é o destinatário

Mensagem (ego, contact) = mensagem de ego para contato

TotalTime = tempo total desde que a primeira mensagem no log foi enviada

TotalTime (ego, contact) = tempo acumulado de interação entre ego e contato (tempo entre a primeira mensagem de ego para um contato contido no log)

A co-ocorrência em mensagens (múltiplos destinatários) indica os grupos sociais dos quais um usuário é parte. As esferas de colaboração frequentemente envolvem mais que pares ego-alter, elas frequentemente envolvem múltiplas pessoas (ego e vários alters). A descoberta de múltiplos participantes simultaneamente empenhados em um esforço colaborativo se dá através da análise de co-ocorrência em sequências de mensagens: grupos de alters participando na mesma conversa tem este thread como um ponto de interesse comum (assim, destinatários das mesmas mensagens seriam alocados a um mesmo mundo social). O fato de cada destinatário ser conscientemente selecionado pelo remetente (seja ego ou um alter) como um destinatário numa mensagem é levado em conta neste passo.

Group(contact_i, contact_j) :

if (contact_i ∈ Recipients(Msg_x)) ∧ (contact_j ∈ Recipients(Msg_x))

Group = True

else

Group = False

Para determinar que grupos detêm a atenção do usuário, o sistema procura discrepâncias entre o comportamento atual e o comportamento normal do usuário. As variáveis que caracterizam trocas de email são número de mensagens trocadas e tempo de resposta. Para cada alter, comparamos o comportamento atual ao comportamento normal (a média previamente calculada).

AvgActivity(ego) = Count(MessageSent(ego)) / TotalTime

AvgActivity(ego, contact) = Count(Message(ego, contact)) / Time(ego / contact)

Tendo estabelecido “comportamento normal”, um sistema pode verificar quando o comportamento corrente cai fora destes parâmetros. Quando os níveis de comunicação (em termos de número de mensagens trocadas em determinado espaço de tempo) sobem

acima da média, o sistema considera este como uma subrede ativa, onde ego está engajado em colaboração. A função de atividade é definida:

$$TimeBetweenMsgs(ego, contact) = Time(Msg_i(ego, contact) - Msg_{i-1}(ego, contact))$$

$$Active(contact) = \text{if}(TimeBetweenMsgs(ego, contact) < AvgActivity(ego, contact))$$

$$Active(contact) = True$$

else

$$Active(contact) = False$$

Quando não há dados prévios de atividade para um contato, o método substitui $AvgActivity(ego, contact)$ para $AvgActivity(ego)$.

Um alter ou um grupo recebendo mais atenção que o normal são aqueles nos quais o usuário está ativamente interessado. Uma série de respostas num período de tempo mais curto que a média, ou uma intensificação de mensagens (i.e., mais mensagens estão sendo trocadas que normal) indica colaboração em andamento. Os grupos nos quais o usuário está muito ativo são de mais interesse, com atividade fornecendo uma indicação do foco de atenção.

O corte de primeiro nível tenta responder à seguinte pergunta: dado o universo de conhecidos do usuário, com quem o usuário está colaborando? A ênfase está em encontrar esferas ativas de colaboração, para procurar informação em atividades relacionadas a trabalho em conjunto.

Há certos ritmos de trabalho, e mudanças de atividade em grupos de acordo se dão de acordo com a necessidade. Assim, um grupo pode estar muito ativo durante um período de tempo e cair após certo ponto (p.ex., conclusão de projeto ou marco de projeto). Deste modo o sistema deve constantemente verificar a formação de novos grupos e mudanças em padrões de atividade. Esta atividade é refletida nos níveis de interação entre membros de grupo, que significa que um sistema deve capturar e processar dados constantemente, para manter um modelo atualizado das colaborações do usuário.

As esferas de colaboração do usuário não são definidas somente por um grupo de indivíduos, mas também pelo trabalho compartilhado que os une. Um método deve utilizar análise de conteúdo para diferenciar interações, definindo os mundos sociais como um conjunto de indivíduos com um tema compartilhado, meta ou projeto, e assim poder ser usado para determinar que atividades relacionar a cada mundo social. Os

recursos que acompanham interações são automaticamente atribuídos à esfera de colaboração, como incluem ação consciente e seleção da parte do usuário.

6.2.2 Determinação de Recursos e Temas

Depois de determinar que partes de uma rede do usuário estão ativas, passamos à determinação dos temas com os quais elas lidam e que recursos e atividades estão relacionados. Isto é realizado através de uma análise do conteúdo destas interações, e subsequente comparação destas a recursos sendo manipulados por participantes de grupo.

A determinação de que atividades estão relacionadas a uma esfera de colaboração é realizada utilizando análise de conteúdo e comparando interações a atividades em andamento. Nossa proposta é aplicar extração de palavras-chave e técnicas de indexação extrair mensagem e temas de atividade, utilizando os métodos de busca e recuperação de informação descritos anteriormente. Atividades em andamento com conteúdo semelhante com o conteúdo de interações em andamento (determinadas no passo prévio) são atribuídas às suas respectivas esferas de colaboração.

Para análise e comparação de texto, vetores de palavras-chave são construídos para representar interações e tarefas. Estes são construídos usando o algoritmo TFIDF, descrito em capítulo 2, que gera vetores de palavras-chave a partir de documentos de texto dados, e os compara usando o modelo vetorial, e a medida de co-seno de proximidade. Dado que a maioria das atividades são tarefas de processamento de informações que envolvem uma grande quantidade de informação textual, esta deve ser uma abordagem viável, que deve extrair atividades que são relacionados a conversas prévias. Sendo métodos estabelecidos de recuperação e comparação de informações, TFIDF e medidas de co-seno foram extensamente aplicados e testados, com resultados bons.

Os métodos de recuperação de informações fornecem uma avaliação que é usada para estabelecer quão semelhante uma esfera de colaboração é de recursos. Anexos e links incluídos em mensagens são tratados de forma diferente: como envolvem uma seleção ativa para distribuição de conteúdo, eles contêm “inteligência de usuário” e imediatamente são atribuídas à respectiva esfera de colaboração, o que significa que eles também serão levados em conta quando da indexação recuperação de informação, no próximo passo.

Arquivos de texto (pdfs, documentos de palavra, páginas de teia) são processados para extrair palavras-chave da mesma maneira que mensagens, e comparados com interações em andamento, fornecendo uma ligação entre tarefas individuais e os mundos sociais nos quais um usuário está inserido. Quanto mais alta a avaliação de semelhança, mais relacionado o recurso está à esfera de colaboração. Deste modo, a atribuição a esferas de colaboração inclui um “nível de relacionamento”, que é designado de acordo com quão bem relacionado o recurso está à esfera de colaboração. No caso de anexos e links, estes são unidos com força 1. Outros recursos são ligados de acordo com suas avaliações de semelhança. Assim, eles são estabelecidos de acordo com a semelhança. Um sistema ativo deve acompanhar as atividades do usuário, periodicamente recalculando as atividades e padrões de mensagem.

Nesta tese, nós só examinamos atribuição de recurso às esferas apropriadas de colaboração. Abaixo acontecimentos planos (e.g., editando um parágrafo de um arquivo) são partido para mais trabalho. Controle de mais fina granularidade, análise e exposição capacitariam o operador "perfurar para baixo" obter mais informação. No entanto, nesta fase, nós somos preocupados só com a constituição e atribuição das pessoas e recursos a esferas de colaboração, como estes primeiro devem ser descobertos para capacitar MAIS elicitação e análise de acontecimentos.

Une méthode pour la sélection et distribution d'informations d'activité basée sur théories d'attention a été présenté. La méthode groupe gens et objets dans une structure appelé sphère de collaboration, et détermine quelles sphères sont actives a un moment donné. Dans cette thèse, nous travaillons sur seulement l'allocation des ressources dans des sphères de collaboration. L'étude de quel niveau de détail est nécessaire est laisse pour un autre moment.

7 **Análise**

Dans ce chapitre nous présentons quelques analyses suivies pour vérifier la viabilité des méthodes décrits. Une validation complète nécessite une étude longue avec plusieurs utilisateurs et activités, ce qui n'est pas faisable dans le délai établi. Pourtant, nous avons conduit certaines analyses pour vérifier comment la méthode décrit fonctionnerait.

Neste capítulo apresentamos algumas análises empreendidas para avaliar a viabilidade dos métodos descritos. Uma plena validação exigiria um estudo de longo prazo envolvendo vários usuários e atividades, e seria difícil de realizá-la no timeframe disponível. Deste modo, projetamos algumas análises para ver quão bem o método descrito funcionaria.

7.1 Análise de Interações para Determinar Esferas de Colaboração

Ativas

Para a primeira análise foram utilizados os dados do caso Europeu descrito anteriormente. Antes de efetuar a análise, coletamos todas as interações, salvando mensagens individuais em uma base de dados, extraindo remetente, títulos, datas e anexos. A maioria dos anexos eram ponteiros (links web) ou documentos anexos, que foram manualmente convertidos em pdfs para simplificar o processamento. O conjunto de dados consiste em 1023 mensagens, subdivididas em 112 threads, enviados durante um período de aproximadamente 24 meses. A classificação em diferentes tipos de thread vista no Capítulo 3 não foi usada nesta análise, mas os threads técnicos foram anotados. Fazer uma referência cruzada entre os resultados e a classificação ficar para uma próxima análise. Por razões de privacidade, todos os nomes e filiações foram omitidos.

7.1.1 Análise I: Padrões de Interação no Dataset do Projeto UE

O primeiro passo do método proposto analisa padrões de interação para extrair esferas ativas de colaboração. Para verificar este passo, nós tomamos os dados do fórum para análise e verificamos a existência de padrões variáveis de interação, e extraímos diferentes grupos de trabalho de dentro do grupo maior. Algumas adaptações foram feitas ao método porque os dados vieram de um fórum grupal ao invés de uma coleção individual de email e o conjunto de conhecidos já era limitado. Como os participantes do fórum são todos de algum modo relacionados ao projeto, era difícil determinar se poderiam ser retirados por uma análise de mensagens, especialmente tendo em vista que as mensagens enviadas ao fórum estão abertas a todos. No entanto, só alguns respondem a cada colocação. Deste modo, contamos estes como os destinatários de fato da mensagem, visto que o assunto chamou sua atenção. Participação em um thread pode ser interpretada como interesse (representando um compromisso de atenção e tempo), e os subgrupos foram identificados nesta maneira. Os participantes de um mesmo thread formavam subgrupos lidando com assuntos particulares. Este meio também permite a determinação de threads que não conseguiram capturar o interesse do grupo ou não exigiu uma resposta (ex. anúncios ou mensagens “para sua informação/conhecimento”), aos quais ninguém envia qualquer resposta.

Nossa primeira hipótese neste estudo é: *se a intensidade troca de mensagens é representativa de colaboração, parceiros técnicos, que colaboram mais proximamente com B, terão níveis mais altos de interação (número de mensagens enviadas e recebidas).*

As mensagens foram contabilizadas da seguinte maneira: cada mensagem enviada a um thread foi contada como uma mensagem enviada a todos os participantes que contribuíram para este thread, o que significa que cada membro enviou ao menos uma mensagem a todos os outros membros, dado que alguns threads de organização extraíram uma alta taxa de resposta.

É possível observar a formação de sub-grupos, com alguns grupos de participantes se juntando em determinados pontos do tempo. É também possível observar os diferentes padrões de interação, que refletem o empenho dos participantes em conversas. A Figura 7-1 mostra um gráfico de interações para o grupo inteiro no período em questão (quase dois anos). Como um pode ser notado, há períodos de alta

atividade intercalados com outros de pouca ou nenhuma atividade (presumivelmente, feriados ou férias, como sugerido pelos períodos de agosto sem atividade). Salientamos este ponto porque os gráficos individuais de interação na Figura 7-4 foram baseados em períodos de maior atividade, para evitar períodos longos de inatividade que deixariam os gráficos difíceis de ler.

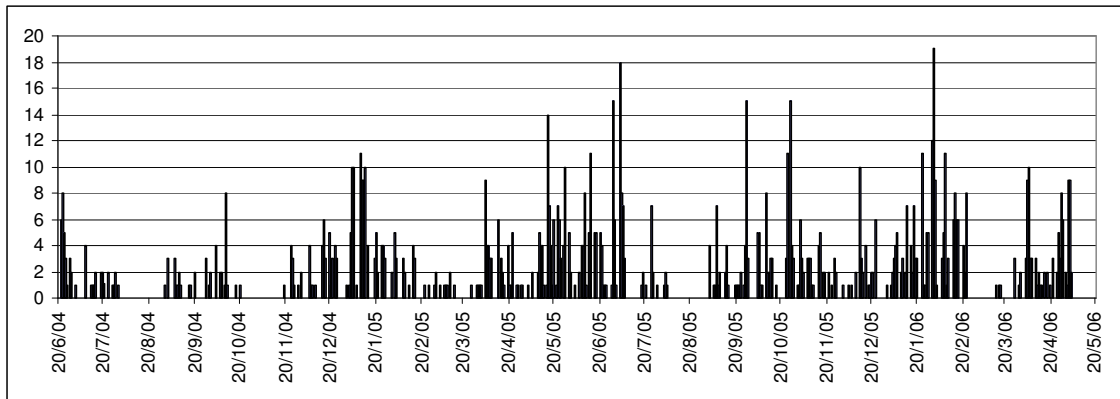


Figura 7-1: Mensagens totais por dia

Nos gráficos de interação apresentados mais tarde, o ponto de vista egocêntrico é usado, e ligações com os outros são construídos de acordo com o número de mensagens trocadas. Seleccionamos o participante B como ego, pois foi um participante claramente ativo, com 102 mensagens enviadas a seus pares no período analisado, e participação em 44 threads. Era um também um parceiro técnico, que acreditamos nos traria resultados mais interessantes.

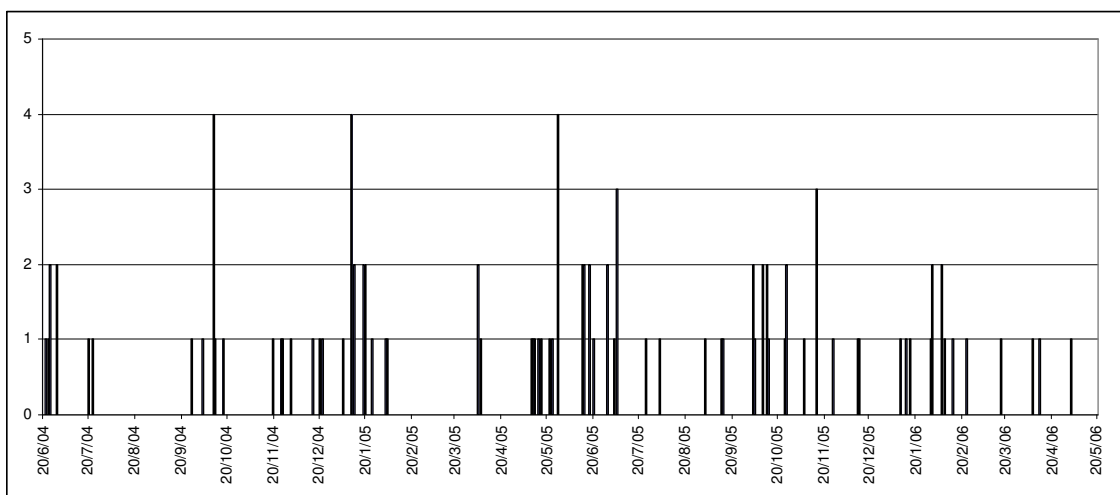


Figura 7-2: Mensagens enviadas por B por dia

Ao examinar o padrão de atividade de B dentro do grupo (Figura 7-2), podemos ver que há momentos de atividade intensa intercalados com outros de baixa atividade.

Os padrões de interação no tempo (mensagens enviadas a outros), mostram os ritmos diferentes em que participante B se engaja em conversas com os outros.

Uma análise de mensagens enviadas e recebidas pelo participante B a outros participantes mostra que a frequência de envio e recebimento varia muito entre participantes diferentes. Não surpreendentemente, o parceiro com o qual B interagiu mais frequentemente era outro parceiro técnico. Por usar o intercâmbio de mensagens como um indicador, nós esperamos chegar ao grupo de colaboradores técnicos (A, C, D, F, J, P).

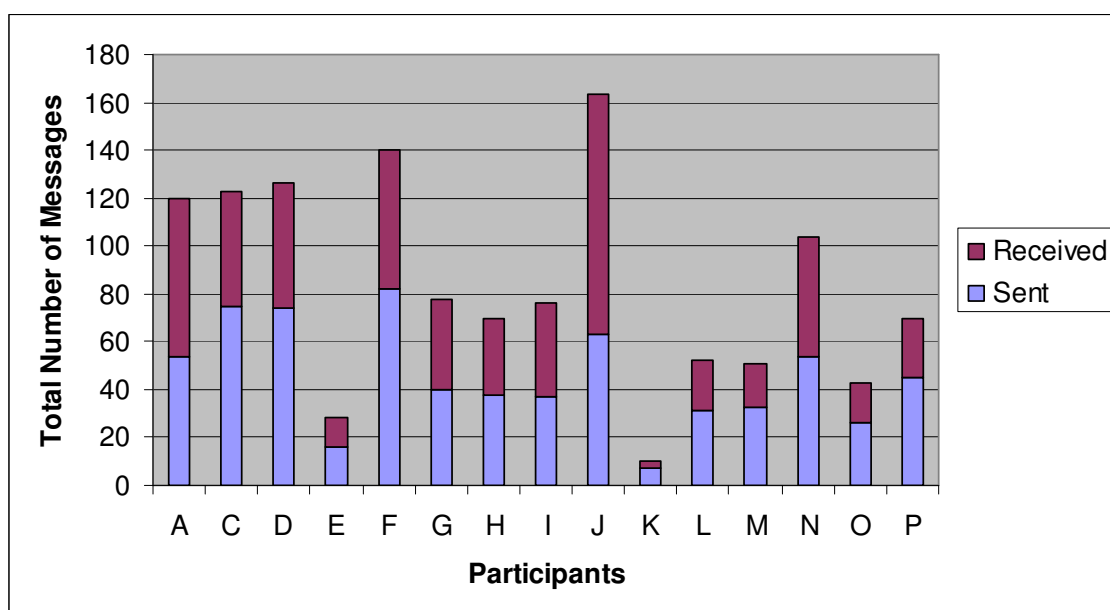
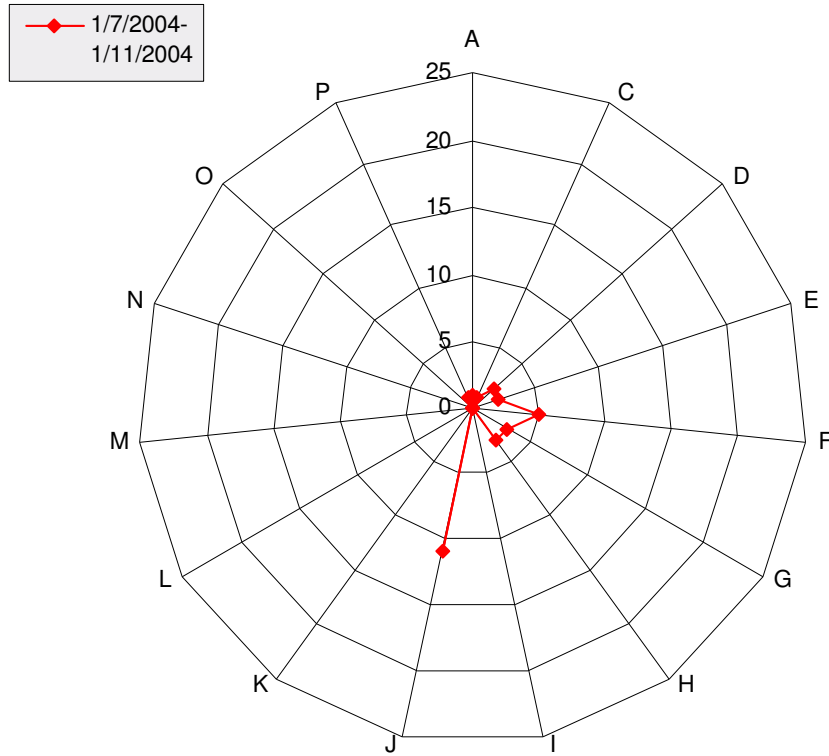


Figura 7-3: Mensagens trocadas entre membro B e outros membros do grupo

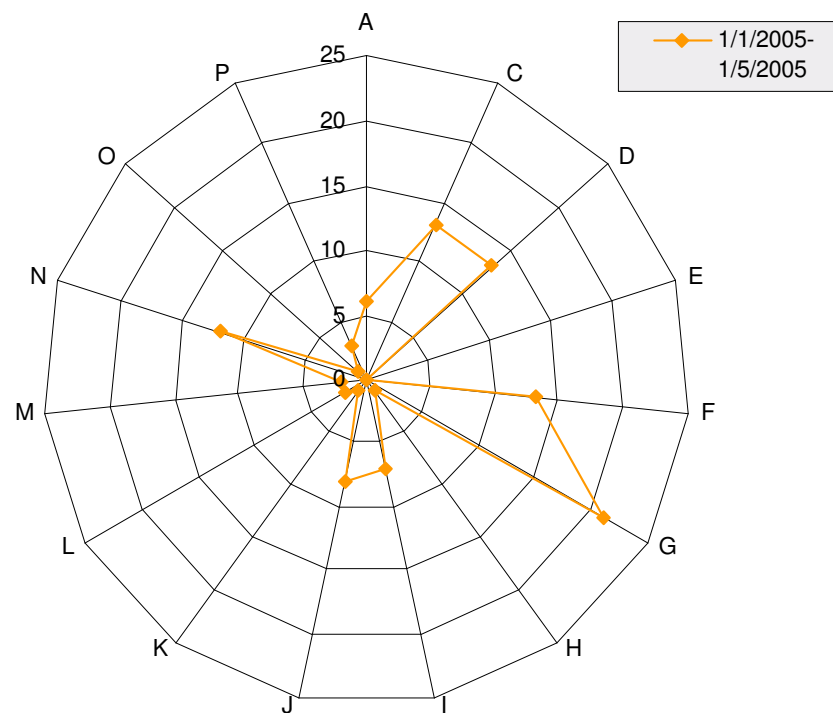
Após verificação, parceiros técnicos tinham intercâmbio de mensagens mais intenso. A única média baixa (inesperada) foi o membro P, que teve um histórico pobre de participação no grupo (participação em apenas 22 threads, a maioria para programar reuniões e confirmar presença). Uma contagem alta inesperada foi o membro N, que era um dos “clientes” para a aplicação sendo construída por B e seus colaboradores. Colaboração mais próxima com N era necessária como forneceram o caso sob estudo, equipando dados e requisitos para os módulos sob desenvolvimento. Isto fornece evidência que a colaboração está diretamente relacionada à interação.

Figura 7-4 mostra padrões de interação para usuário B em períodos diferentes, com diferentes parceiros (A, C, D, F), em termos de mensagens enviadas por dia (quanta atenção B dedica a cada parceiro), para ilustrar a ativação de diferentes laços (A-E). Ao analisar estes gráficos para cada laço, é possível imaginar relacionamentos tornando-se

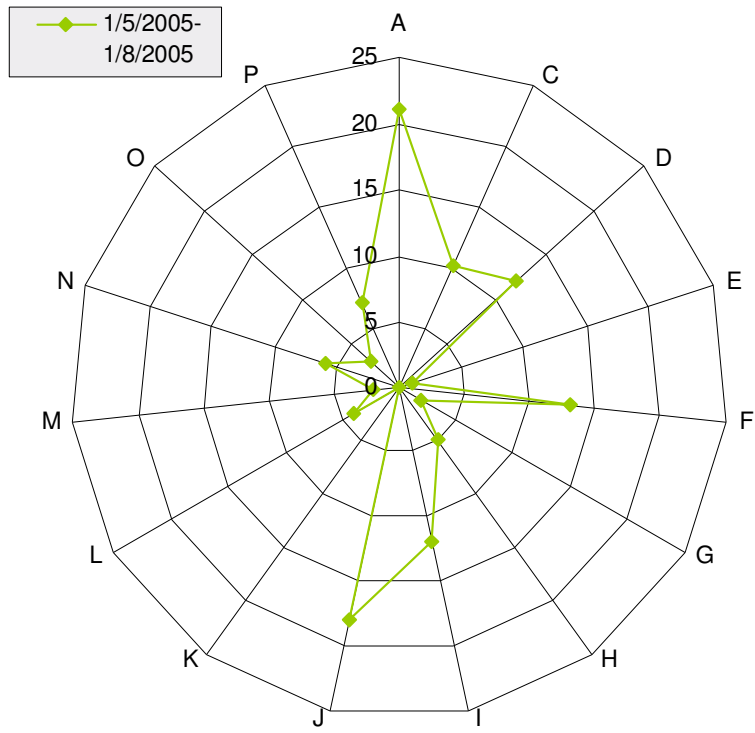
ativos através da comunicação, o que representa interesse ou trabalho compartilhado. Parceiros diferentes tornam-se ativos em momentos diferentes, mostrando assim que, mesmo no grupo pequeno como o analisado, membros se organizam em subgrupos conforme necessário.



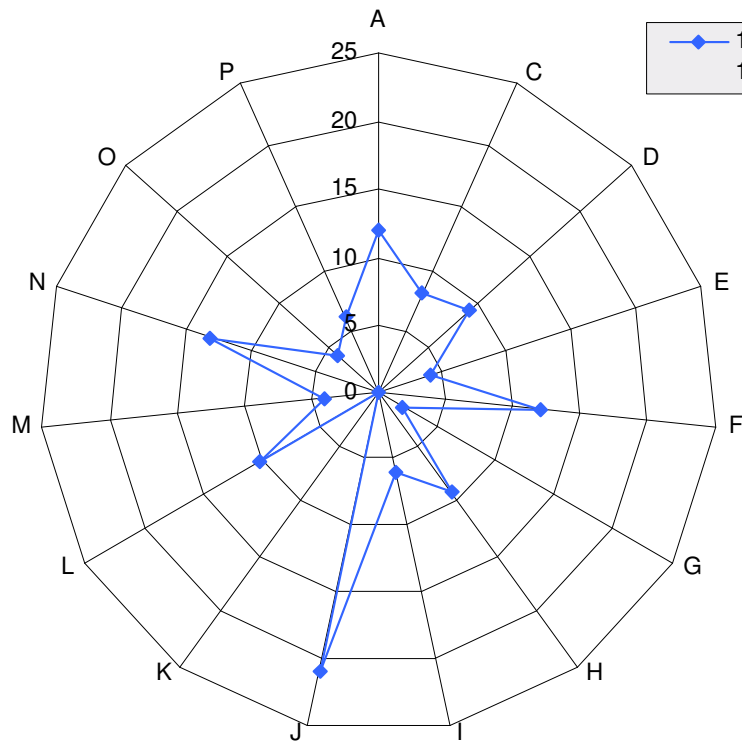
A



B



C



D

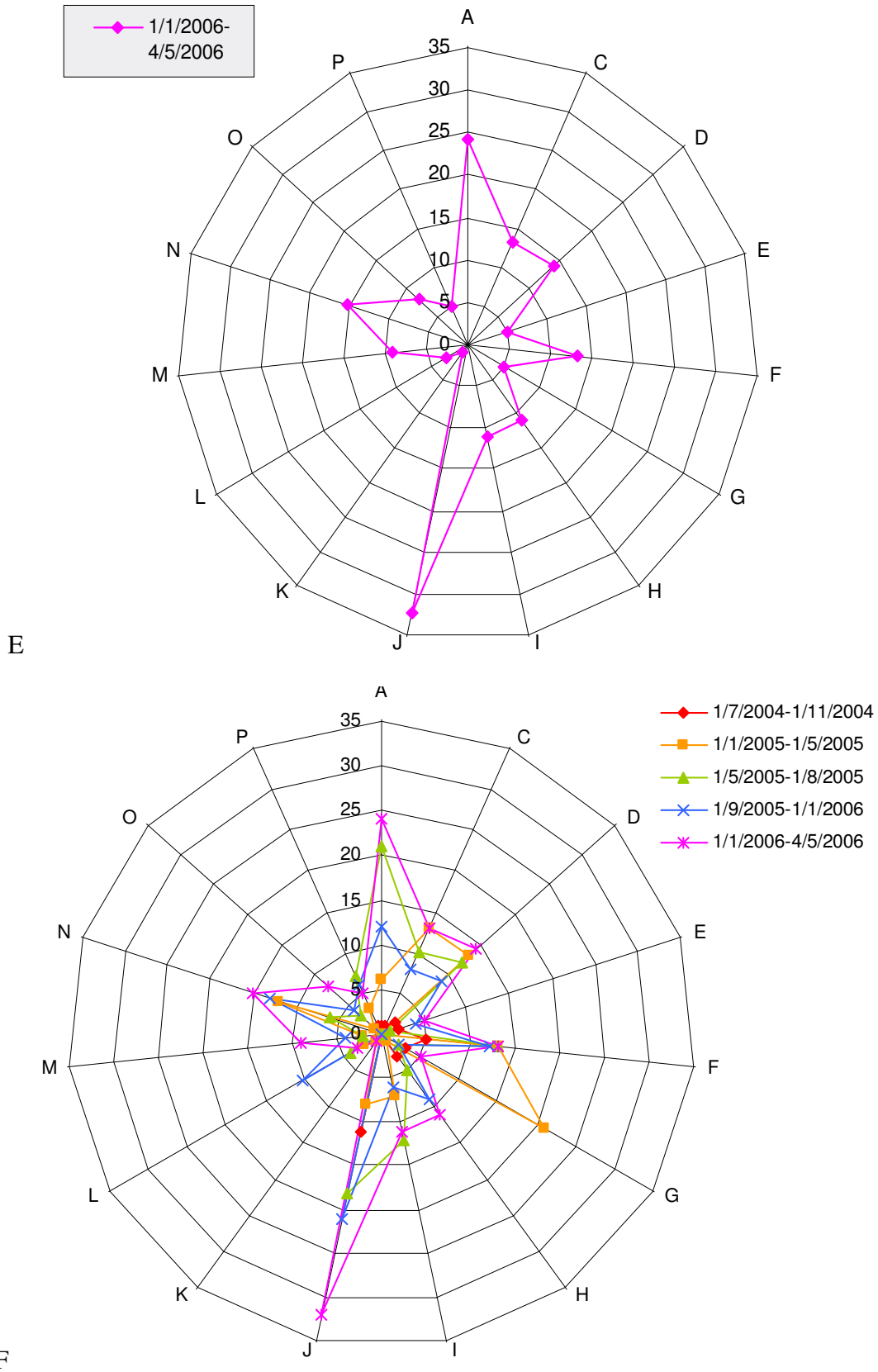


Figura 7-4 (A-F): Número de mensagens enviadas por dia de B a outros membros (A, C, D, F). O último gráfico mostra uma sobreposição dos anteriores

7.1.2 Análise II: Padrões de Interação no Dataset de Emails

Para uma segunda análise, uma ferramenta de visualização foi construída para viabilizar a verificação de outras algumas suposições envolvidas e obter retorno do usuário. Esta ferramenta executa análise estrutural de emails, extraindo remetentes e recipientes e constrói sociogramas, de acordo com os métodos descrito anteriormente. Os laços entre usuários são qualificados em termos de número de mensagens (mensagens enviadas e recebidas), mas nenhuma análise de conteúdo é executada.

Uma interface exibe o sociograma, com a possibilidade de exploração de diferentes intervalos de tempo e fontes, como visto na Figura 7-5. O sistema implementa uma visualização de grafo baseada em molas, usando o modelo de força de Fruchterman-Rheingold (Fruchterman e Reingold, 1991). Este algoritmo trata um grafo como um conjunto de nós que se repelem, mas são unidos por molas que os atraem. Ele gera assim um desenho que reflete proximidade entre nós enquanto tenta reduzir os cruzamentos de linha, aumentando a legibilidade. A visualização foi construída usando a linguagem de Java e a biblioteca de JUNG para construção e visualização de grafos. É importante notar que esta é uma interface de análise, e não será usada como uma interface para um sistema final. Ela foi usada para entrevistar usuários sobre seus mundos sociais e como estes se relacionam ao trabalho em andamento e necessidades de awareness, para obter uma análise preliminar e receber comentários dos usuários antes de prosseguir com uma implementação de sistema. Entrevistamos 4 usuários sobre os padrões de email vistos na visualização.

Para executar uma verificação da determinação de colaboradores, trabalhamos com três suposições:

- Esferas de colaboração estão refletidas em email;
- Participação numa esfera de colaboração indica interesse;
- Com o tempo, o nível de atividade entre esferas de colaboração muda.

Os históricos de email dos usuários foram processados, e estes usuários foram entrevistados com seu respectivos sociogramas à mão para inspeção. Nossos usuários são todos usuários experientes de email, com milhares de mensagens em suas caixas de correio. Um é um professor universitário, e os três outros são estudantes na mesma universidade. Eles freqüentemente se comunicam com estudantes e colegas via email,

trocando arquivos com seus colaboradores e recomendações em suas mensagens. A maioria de suas atividades consiste em redigir documentos (artigos, relatórios, propostas), estudar (lendo documentos de páginas web) ou codificação. Exploramos seu sociogramas com eles, examinando diferentes intervalos de tempo. Perguntamos aos usuários se:

1. os grupos (cliques) que eles identificaram em seu sociogramas estavam relacionados a projetos ou outras colaborações em andamento;
2. grupos diferentes tornavam-se ativos quando o alcance temporal era mudado;
3. padrões de troca de mensagens refletiam projetos;
4. mundos sociais em que o usuário não tinha participado (ou era simplesmente “observador”) eram de interesse quanto à awareness de tarefas dos participantes.

Todos os usuários foram capazes de relacionar cliques nos seus sociogramas com colaboração em andamento, mesmo se histórica. No entanto, nem todos estes mundos sociais estavam relacionados a trabalho. Em certas situações, eles representavam grupos com contexto compartilhado não relacionado a trabalho (e.g., estudantes no mesmo departamento), que não estavam em colaboração direta. Esta observação confirma que mundos sociais delineiam cliques, mas não segue que todos estes representam trabalho em conjunto.

Mais estudos são necessários para determinar como diferenciar entre interações de trabalho e aquelas puramente sociais. Porém, deve-se levar em conta que é difícil dissociar o aspecto pessoal do de trabalho quando lidando com relacionamentos interpessoais. Muitos relacionamentos têm uma natureza dupla, sendo parcialmente relacionados a trabalho e parcialmente pessoais, e estas duas facetas têm um efeito uma na outra.

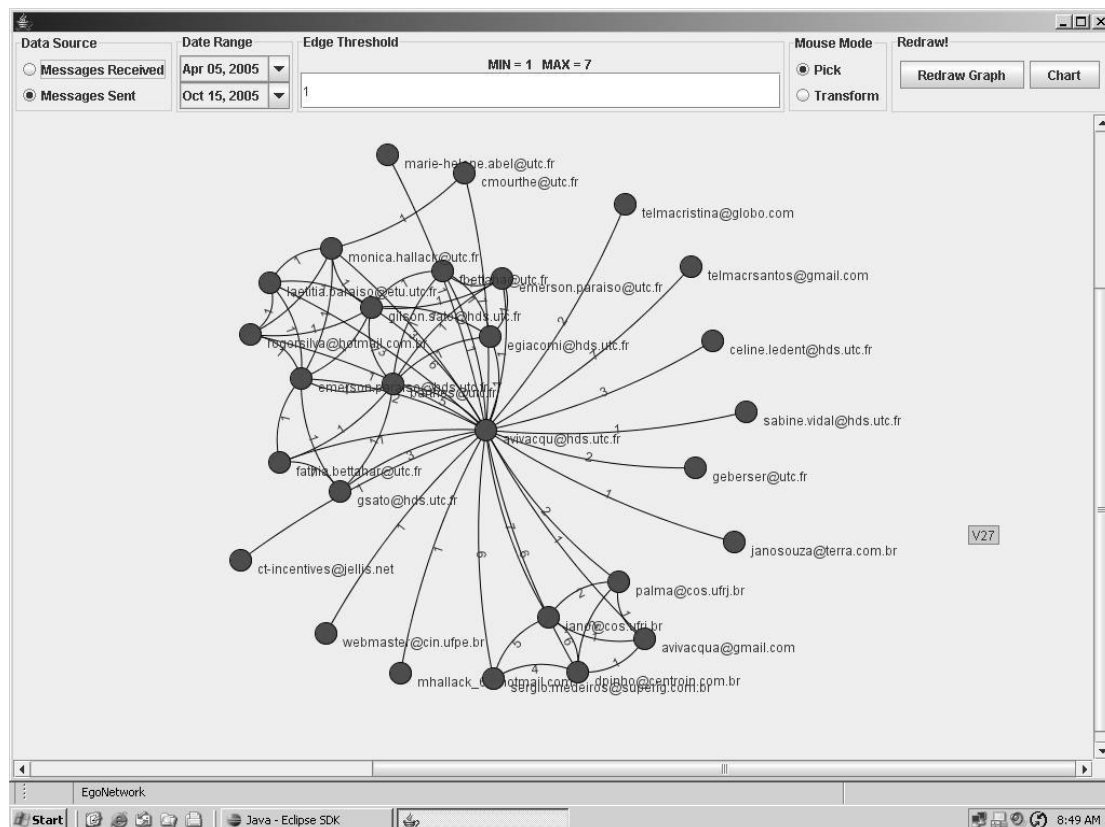


Figura 7-5: Screenshot de visualização, vários grupos estão visíveis.

Mudanças nas faixas temporais salientaram grupos diferentes, devido à mudanças em níveis de atividade. Mundos sociais tornaram-se mais ou menos ativos de acordo com a necessidade, e a participação do usuário neles também variou em intensidade, indicando interesse.

A utilização de intervalos mais curtos de tempo reduziu consideravelmente o número de mensagens e a visualização correspondente, tornando mais fácil a identificação de subgrupos. Isto confirma que níveis de atividade são refletidos em email. Ao inspecionar o gráfico temporal (visto na Figura 7-6), onde tempo é dividido em faixas de trocas diárias de email, mudanças no padrão de interação podem ser vistas. Um relacionamento dormente repentinamente se torna ativo, com emails sendo trocados diariamente (às vezes várias mensagens em um mesmo dia, dependendo da urgência), e então desaparece quando objetivos são alcançados. Isto confirma a suposição que padrões temporais são detectáveis por email. As mudanças frequentemente eram bastante bruscas, indo de nenhuma interação a quatro mensagens no dia de um dia para o outro. Enquanto isto era esperado, nós também esperávamos ver padrões mais suaves, onde interações gradualmente aumentam com tempo.

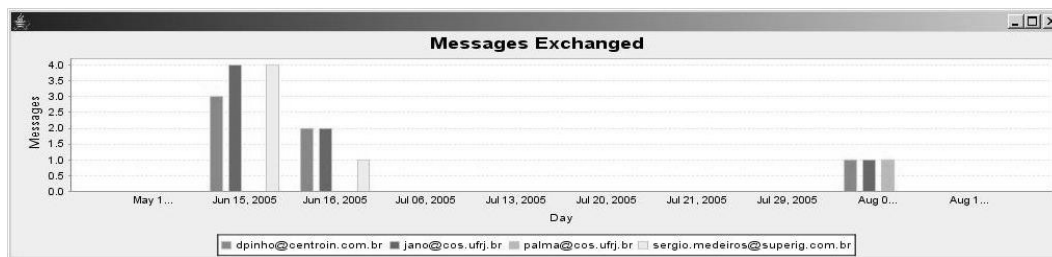


Figura 7-6: Gráfico de interação baseado no tempo

Todos os usuários confirmaram que mundos sociais em que eles ativamente não participaram não eram de interesse suficiente para levar à criação de uma esfera de colaboração. Uma análise dos dados disponíveis revelou que a maioria destes eram anúncios ou listas de email que não estavam relacionados a quaisquer projetos. Listas baseadas em projetos normalmente detinham o interesse dos usuários, e eles se empenhavam em participação ativa. Os usuários queriam estar conscientes de seus colaboradores mais próximos, e de atividades relacionadas a projetos urgentes, onde havia muita coordenação a ser feita, e não tinham nenhum desejo de estar cientes de todo o universo de trabalho, embora em alguns casos eles gostariam de permanecer superficialmente cientes do que estava ocorrendo em um grupo. Isto indica que, para propósitos de consciência de tarefa, threads em que o usuário não participou pode ser deixadas de fora. Em termos computacionais, isto reduz significativamente o tamanho do grafo, e, portanto, da memória necessária e tempo de processamento. Por exemplo, em um dado caso isto significava mais de 70% dos contatos totais existentes na caixa de emails do usuário. Quantias grandes de spam e listas de endereços de anúncio explicam a maioria dos emails retirados neste caso particular. No entanto, alguns falso positivos também foram observados, onde mensagens de indivíduos que possuíam múltiplos emails foram perdidas pois ego usava um só endereço. Esta dificuldade pode ser superada através da criação de alias que uma múltiplos endereços de email para representar um alter, e deste modo não perder mensagens relevantes.

Dentro dos emails, havia vários exemplos de mundos sociais relacionados a projetos, normalmente qualificados por interação intensa num período curto de tempo (semanas ou alguns meses). Isto sugere um meio de mais eficiente de selecionar grupos relacionados à atividade. A inspeção de mais dados mostrou que essa estrutura às vezes não era suficiente para separar atividades de grupos, especialmente quando existiam grupos sobrepostos. É necessário que estes sejam qualificados de acordo com os temas de interação, de modo que eles possam ser eficientemente diferenciados. Havia alguns mundos sociais sobrepostos, incluindo sobreposições temporais, onde um mesmo grupo

estava trabalhado junto em mais de um projeto ao mesmo tempo. Dentro de nossos conjuntos de dados, havia também alguns mundos sociais contidos em outros.

O número de mensagens enviado por ego provou ser um qualificador razoável para laços, mas como o número de mensagens enviado por um alter não. Alguns alters enviaram mais de 200 mensagens em um período de 6 meses e não eram colaboradores nem tinham qualquer interesse ao usuário. Mensagens enviadas pelo usuário refletiam sua participação e interesse em mundos sociais com maior precisão. Participar de conversas envolve um investimento de tempo e esforço que indica um nível de compromisso com o grupo. Assim, os algoritmos processam mensagens enviadas do usuário primeiro, determinando os alters relevantes e então considera mensagens recebidas destes.

7.2 Análise de Conteúdo Extraído de Recursos

7.2.1 Análise III: Análise de Conteúdo em Dataset do Projeto EU

Para verificar a viabilidade da segunda parte do nosso método, nós analisamos o uso de análise de conteúdo para ligar interações e recursos usando técnicas de recuperação de informações para medir a correlação entre documentos. Estas medidas de semelhança são usadas para associar recursos a mensagens, determinando quão inserido numa esfera de colaboração um recurso está. A medida primária é o nível de correlação entre dois documentos distintos: tanto mais semelhantes (segundo um cálculo de TFIDF, descrito anteriormente), o mais correlacionados serão. Deste modo, realizamos uma série de comparações de semelhança, para verificar quão bem diferentes variações de elementos funcionariam. O conjunto de arquivos continha arquivos enviados como anexos, arquivos colocados no fórum e arquivos externos, inseridos para ver se falsos positivos seriam encontrados. Índices foram criados para mensagens, threads e arquivos, e a semelhança entre eles foi calculada.

As comparações de conteúdo foram executadas tomando mensagens e as indexando, e então usando estes índices para recuperar documentos manipulados pelo usuário. Nós indexamos também threads completos, que eram focos de atividade. Depois de retirar stopwords e extrair radicais, extraímos as palavras-chave para cada um dos threads. Palavras-chave para 2 threads são mostrados em Tabela 7-1, para ilustrar a diversidade de temas envolvidos.

Tabela 7-1: Palavras-chave para threads

Thread	Palavras-chave mais freqüentes (com freqüências)
8	Activebpel: 5; codificam 15; contribuição 7; documentação 11; bibliotecas 12, módulo 11
11	quantia 5; deposita 5; encarrega 3; contrai 3; distribuição 4; arquiva 5
83	conferência 3; conversa 3; projetar 3; governo 5; impacta 3; sócios 6; apresentação 14

Comparamos mensagens aos anexos em pdf, para ver se documentos relacionados seriam recuperados. Os 10 melhores resultados para cada thread foram tomados para análise. A tabela 7-2 mostra um thread de exemplo, com o assunto e uma descrição dos documentos com seus resultados obtidos (pontuação). Esta pontuação seria usada como o indicador de pertinência de um artefato a uma esfera de colaboração. Esta pontuação é calculada tomando as palavras-chave dos threads e procurando-as nos documentos em questão.

Tabela 7-2: Resultados para um thread (10 primeiros mostrados)

Assunto: " código aberto e próximo encontro de revisão "	Pontuação
1. detalhes de uma aplicações piloto do parceiro F	0,495258
2. histórico da segunda conversa via chat	0,477502
3. artigo descrevendo aplicações piloto	0,472957
4. ontologia de piloto - sócio A	0,435104
5. histórico da primeira conversa via chat	0,414138
6. definição de métricas e pontos de partida de medidas	0,401056
7. armazenamento e detalhes de acesso	0,369296
8. armazenamento e detalhes de acesso, revisado	0,332887
9. aplicações piloto - sócio P	0,326288
10. versão de esboço do projeto para entrega	0,305564

Uma das questões é achar pontos de corte apropriados, onde documentos que não são mais relevantes à esfera de colaboração serão descartados. Discussões técnicas obtiveram os melhores resultados, com pontuações de semelhança altas com documentos técnicos entregues e com históricos de duas conversas online.

A primeira análise que conduzimos foi entre mensagens individuais e arquivos. Uma análise dos valores de semelhança entre eles resultou em pontuações pobres. As avaliações de semelhança eram baixas (Figura 7-7). Quase todas as pontuações mais altas eram falsos positivos, onde uma mensagem é ligada a recursos sem ligação com seu conteúdo. A maioria destes ocorreu porque mensagens eram curtas e continham

nomes próprios que também figuravam nos documentos analisados. Como havia poucas palavras nas mensagens, estes nomes ganharam peso, gerando falsos positivos. Algumas exceções ocorreram: correlações altas foram encontradas entre uma mensagem que continha itens de agenda para uma reunião, e documentos com itens de agenda (para aquela e outras reuniões). Outra correlação forte foi encontrada entre uma mensagem longa e altamente técnica sobre ontologias e documentos que discutiam o ciclo de vida da ontologia.

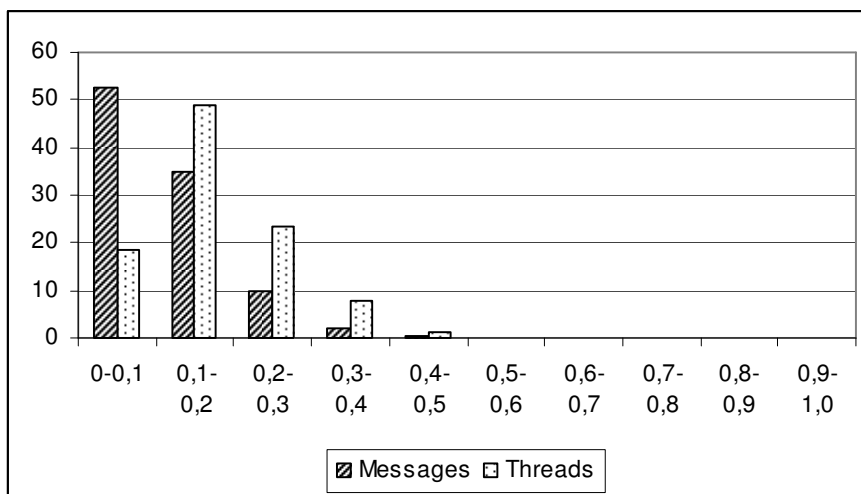


Figura 7-7: Distribuição de semelhança para 10 faixas de valores, para mensagens e threads, como uma porcentagem do total.

Os threads obtiveram pontuações mais altas que mensagens isoladas na comparação com recursos. Isto é compreensível, visto que uma série de mensagens no thread fornece conteúdo extra e estabelece uma conversação. Como notamos que as mensagens mais curtas levaram a resultados inconclusivos, selecionamos as 10 mensagens mais longas para analisar quanto à precisão, em nosso caso uma análise da relevância dos documentos buscados em conexão com as mensagens. As mensagens mais longas continham detalhes técnicos do projeto: algumas eram discussões técnicas, outras eram revisões de documentação questionando aspectos técnicos. Em todos os casos, os três primeiros resultados tinham ligação próxima com o conteúdo da mensagem, apesar de suas avaliações aparentemente baixas de semelhança (por volta de 0.3). Um procedimento semelhante foi executado para threads, com resultados semelhantes: entre os threads superiores, avaliações de semelhança eram aproximadamente 0.4, chegando a 0.5 em alguns casos. Poucos arquivos sem ligação ao projeto foram recuperados neste procedimento. Havia apenas 25 arquivos externos nos 1120 resultados para recursos relacionado a threads (aproximadamente 2%). Deve

ser notado que alguns arquivos externos realmente tinham relação com as conversas em andamento, pois versavam sobre ontologias, agentes e semântica. Recursos completamente sem ligação (ex., manuais de referência de LISP) não figuraram nos resultados.

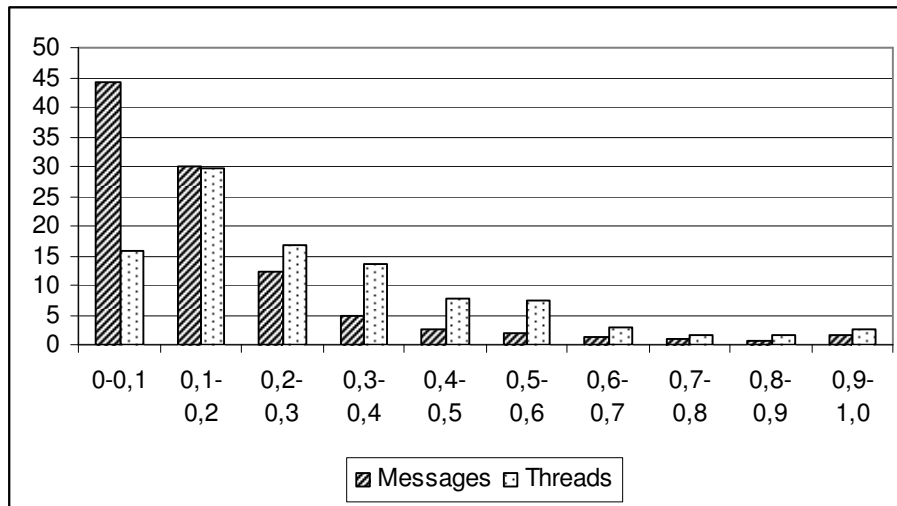


Figura 7-8: Distribuição de avaliação de semelhança para 10 faixas de valores, para mensagens e threads (com anexos incluídos), como uma porcentagem do total.

Desde o nosso estudo prévio do grupo do artigo para revista, passamos a acreditar que anexos e links forneceriam conteúdo adicional, talvez mais relacionados ao conteúdo do projeto que as mensagens em si, motivo pelo qual nós os alocamos automaticamente a esferas de colaboração. Para comprovar esta suposição, criamos índices que incorporavam o conteúdo de anexos ao corpo das mensagens. Esperamos estes fornecerão melhores resultados na comparação que mensagens individualmente. Neste caso, gostaríamos de verificar o seguinte pressuposto: *se anexos contêm conteúdo mais relevante que mensagens, então anexos terão melhores contagens de semelhança que mensagens por si.*

Esta pode ser verificado através de uma comparação de pontuação de semelhança entre mensagens onde o conteúdo do anexo foi incluído e mensagens sem estes. A figura 7-8 mostra um gráfico destas distribuições, para mensagens e threads. É fácil ver que as avaliações de semelhança são maiores quando anexos são considerados, visto que estes fornecem conteúdo adicional com o qual comparar documentos. Este resultado corrobora nossa hipótese de que anexos são uma fonte valiosa de dados relacionados ao projeto. Ao usar links e anexos enviados com mensagens, levamos em conta recomendações dos usuários, que já revisaram o conteúdo destes recursos e os

julgaram úteis ao grupo. Assim, associar um acessório a uma esfera de colaboração é uma operação correta.

Nossa última verificação funcionou de maneira reversa: selecionamos 30 arquivos (10 anexos, 10 entregas e 10 arquivos externos) e tentamos buscar threads a partir deles, para ver se seria possível ligar um artefato com interações que o cercam. Novamente, comparamos estes arquivos a threads e obtivemos avaliações baixas de semelhança (Figura 7-9). No entanto, houve uma distinção entre avaliações de arquivos externos e arquivos relacionados a projetos (sem anexos): arquivos externos nunca foram avaliados acima de 0.3, enquanto outros arquivos podiam chegar até 0.75. Com a inclusão de anexos, estas avaliações subiram consideravelmente, distinguindo arquivos externos de arquivos de projeto. Este é um ponto importante, que pode ajudar a estabelecer uma distinção entre recursos externos e internos.

Consideramos estes resultados bons. Esta análise evidencia que é possível combinar recursos a contextos sociais, e os 3 resultados superiores eram todos relevantes, na medida em que realmente estavam relacionados ao documento em questão (alta precisão). As melhores pontuações vieram de discussões técnicas, que foram relacionadas a documentos descrevendo aspectos técnicos dos projetos, mesmo sem os anexos para prover contexto adicional.

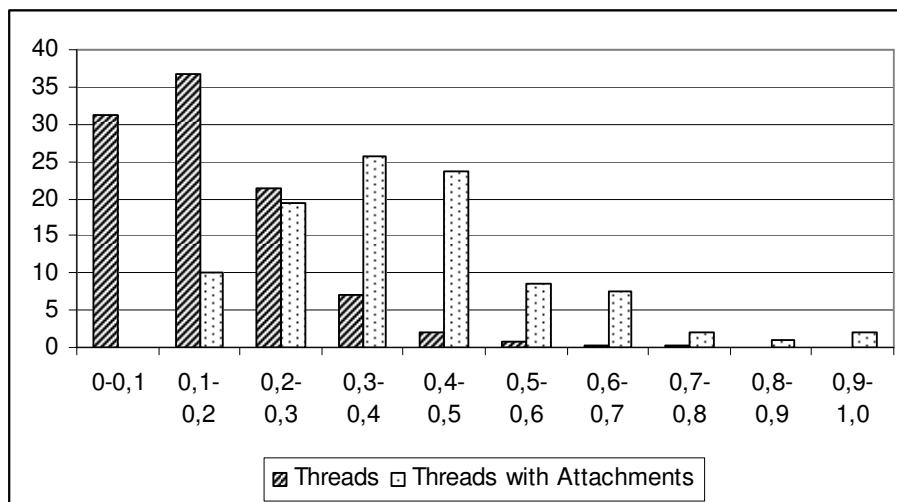


Figura 7-9:Distribuição de avaliação de semelhança para 10 faixas de valores, para comparações entre recursos (arquivos) e threads (com e sem anexos), como uma porcentagem do total.

Fomos surpreendidos ao obter pontuações de semelhança muito abaixo do que esperávamos. Um dos principais problemas foi o comprimento das mensagens: como havia muitas mensagens muito curtas, principalmente threads de programação ou preparação de reuniões (ex. “fulano estará presente representar companhia X” ou “aqui

está a contribuição de Y ao relatório” ou “por favor enviem o relatório trimestral de gerência”). O número grande de mensagens de planejamento e programação pode explicar os resultados pobres, visto que mensagens de organização e programação de reuniões e pedidos de contribuição normalmente geravam apenas mensagens curtas que têm pouco a ver com os temas de projeto em si. No entanto, quando as contribuições eram entregues como acessórios, elas contribuía para a definição de temas da esfera de colaboração.

Para medir quão bem as comparações funcionaram, nós analisamos manualmente os 3 arquivos primeiros recobrados para cada partida, ver se eles foram corretamente relacionados à mensagem ou thread em questão. As comparações foram avaliadas entre 0 e 3, onde 0 é sem ligação, 1 é relacionado, 2 é fortemente relacionado e 3 para os casos em que anexos foram corretamente designados a seus respectivos threads ou mensagens. Enquanto isto provavelmente teria tido resultados mais exatos tiveram sido marcado pelos proprietários dos dados, como eram documentos técnicos, os pesquisadores podia marcá-los sem qualquer problema.

Nós selecionamos 10 mensagens ao acaso e verificamos quão bem estavam correlacionadas aos arquivos: 4 de 30 arquivos eram relacionados e só 1 era fortemente relacionado (um histórico de conversa de IRC ligado a uma conversa técnica). Uma mensagem realmente levou a seu anexo. As mensagens eram muito curtas e portanto não forneceram bons resultados. O mesmo procedimento foi adotado para threads, e as pontuações foram melhores. Para 10 threads selecionadas ao acaso (30 partidas), uma foi ligada a seus arquivos (o thread continha a mensagem mencionado nele), havia 5 arquivos fortemente relacionados aos threads. Os arquivos restantes eram sem ligação. Isto parece indicar que, com exceção de situações específicas (ex. mensagens técnicas) o conteúdo de mensagens não se correlaciona bem com documentos. Isto já tinha sido inferido a partir da análise das mensagens no caso do artigo de revista, que forneceu a indicação inicial que anexos eram melhores fornecedores de conteúdo que mensagens em si.

Outra surpresa foi o número de resultados devido a nomes próprios: boa parte dos melhores resultados eram mensagens curtas que continham nomes próprios (e.g., “participantes A e B serão representados pela companhia X”) a estes foram ligados, por exemplo, relatórios de atividade que continham os nomes destas pessoas. Mesmo não sendo uma ligação particularmente útil, ela traz à tona uma possibilidade: a de dar

tratamento especial a nomes. Estes poderiam ser extraídos de campos de remetente ou assinaturas, e serem explicitamente procurados em mensagens ou documentos.

Nós julgamos mal o estilo de interação do grupo: esperávamos muitas conversas sobre o projeto, mas estas eram a minoria (embora quando aconteciam, forneceram conteúdo excelente para uma busca baseada em conteúdo). Nas reuniões face a face, participantes usavam o tempo para discutir detalhes técnicos e tomar decisões, e tentavam chegar a uma divisão do trabalho para reduzir a interação entre reuniões. Isto mostra que o grupo tem uma tendência a estruturar sua organização e distribuição de tarefas para reduzir a interdependência entre membros não co-localizados, reduzindo assim as necessidades de coordenação.

Entrevistas informais com membros do grupo revelaram que existiam subgrupos menores que interagiam mais frequentemente, fora do fórum (principalmente via email, mas também em IRC e no telefone quando necessário), e estes registros de interação não estavam disponíveis para este estudo. Eles também informaram que se visitavam entre reuniões do grupo quando necessário (um deles permaneceu após a reunião para trabalhar com o anfitrião). A grande diversidade de meios de comunicação adotados também poderia gerar um problema, pois seria difícil cobrir todos os diferentes canais de interação. No entanto, isto traz à tona novas oportunidades, visto que cada canal tem capacidades diferentes e será escolhido dependendo das necessidades específicas e desejos de cada usuário. Dada a escolha do usuário, seria possível inferir se a mensagem era urgente, por exemplo.

Ao juntar as duas partes da abordagem, os grupos ativos, indicados na primeira etapa da análise, teriam seu conteúdo transformado em vetores de palavras-chave e estes então poderiam ser ligados a documentos do usuário ou a recursos em uso (ex: página web). Uma vez a esfera de colaboração construída, informação poderia então ser distribuída como necessário. Por exemplo, quando o thread mostrado acima se tornasse ativo, o parceiro F seria informado de relacionamentos existentes com outros ao editar os detalhes do seu documento sobre o piloto. Alternativamente, parceiros participantes seriam notificados quando um deles editasse o artigo descrevendo as aplicações piloto.

Une fois la sphère de collaboration construite, les informations peuvent être distribués comme nécessaire. Les analyses et entrevîtes avec des utilisateurs montrent que l'analyse d'interactions est une bonne façon de découvrir les collaborateurs courants. Par contre, l'analyse du contenu des messages n'a pas donné des bon résultats, probablement parce que les utilisateurs échangeaient beaucoup plus de messages sur l'organisation que sur le travail et problèmes qu'ils trouvaient.

8 Discussão e Trabalhos Futuros

Cette thèse a présenté un approche pour la distribution d'informations através la découverte et construction des sphères de collaboration et une analyse de ce méthode. L'approche présentée fournisse une façon d'explorer les besoins d'information pour permettre auto-organisation dans groupes faiblement structurés.

Esta tese apresentou uma abordagem para disseminação de informação, através da descoberta de esferas de colaboração, e uma análise deste método. A abordagem apresentada fornece um meio de explorar necessidades de informação para viabilizar auto-organização em grupos fracamente estruturados.

Este é um passo em direção da descoberta e descrição de esferas de colaboração dentro das quais um usuário trabalha. Esta informação poderia ser útil para ajudar o usuário a automaticamente configurar grupos, lembrando-se de quem deve receber determinada informação, quem é impactado pelo trabalho num dado arquivo, quem mais trabalhou sobre ou acessou determinado recurso, o que outros estão fazendo que o usuário deva saber ou o que o usuário está fazendo que deve ser informado a outros.

Podemos antever algumas situações onde esferas de colaboração podem ser úteis: por exemplo, enquanto o usuário trabalha, recursos podem ser apresentados que relacionam este trabalho à esfera de colaboração devida, contextualizando o trabalho. Outra possibilidade é apresentar ao usuário informação sobre pares que estão relacionados à atividade em andamento, de modo que o usuário tenha noção de como esta impacta os outros (uma variação desta idéia é ter sistemas de disseminação automática de informação para fornecer consciência, que introduz um problema de privacidade). Um recurso também poderia ser rotulado com o contexto social ao qual ele pertence, de modo que este vá com ele quando ele for enviado ou recebido por outros. Isto beneficiaria, por exemplo, um recém-chegado ao projeto, que melhor entenderia os múltiplos papéis e indivíduos envolvidos em construir este artefato, além

de ficar a par de acertos, prazos e informação relacionada. A parte principal une artefatos produzidos individualmente com seus contextos sociais e colaborativos.

Esta técnica podia ser usada para construir sistemas para apoiar trabalho em rede distribuído e fracamente estruturado, por apoiar a alternância de indivíduos entre grupos diferentes e manter suas redes ativas através da comunicação. A compreensão e ajustes em resposta a acontecimentos imprevistos no trabalho é uma tarefa cotidiana, que poderia ser mais bem apoiada por sistemas com capacidade de realizar inferências automáticas. Isto envolve não só a manutenção da consciência do grupo, mas também a compreensão de como as pessoas reagem durante a construção de artefatos (mantendo uma perspectiva tanto presente quanto histórica), e que artefatos estão envolvidos em cada relacionamento.

Nardi e colegas (2002) dizem que dois processos são fundamentais no trabalho em rede: lembrança (as pessoas, seus interesses, colaborações continuadas, etc.) e comunicação (com os pares, estejam envolvidos em colaboração ou não). As Esferas de Colaboração poderiam ajudar os usuários a se lembrar de seus projetos em andamento ao relacionar trabalho individual e em grupo. Gonzalés e Mark (2005) também mencionam que um problema comum quando da alternância entre várias colaborações é lembrar-se de acordos feitos e não “deixar a peteca cair”. A distribuição de informação relacionando o indivíduo às colaborações em andamento ajuda os usuários a se manterem a par dos diferentes projetos, e ao mesmo tempo acompanhar o trabalho dos outros e descobrir oportunidades para interação e discussão. Adicionalmente, ao comparar interações e elementos de fora das esferas de colaboração, pode ser possível encontrar novos recursos que seriam de interesse dado o contexto.

Esta pesquisa contribui para área de Trabalho Cooperativo Apoiado por Computador, através da definição de um conceito (Esfera de Colaboração) para interligar áreas de trabalho individuais. Este conceito apóia a criação de espaços compartilhados, interligando indivíduos que trabalham a sós e trocam mensagens para colaborar. Também apresentamos um método para constituir estas esferas de trabalho, encontrando pessoas e recursos que pertençam a um mesmo contexto colaborativo. Este método é uma extensão do modelo de consciência de Rodden (1996), que determina o foco de atenção do usuário para distribuir informação. Os resultados desta pesquisa e de trabalhos relacionados conduzidos durante o desenvolvimento desta tese foram publicados em diferentes publicações, listadas no Anexo I.

8.1 Limitações

Nossos resultados são aplicáveis a quaisquer situações onde seria benéfico agrupar indivíduos com atividades compartilhadas e os respectivos recursos, tomando como base a atividade individual e construindo grupos informais. Um exemplo de um sistema que talvez se beneficie de uma abordagem como a apresentada é o ActivityExplorer (GEYER et al, 2006), que exige que indivíduos manualmente atribuam elementos (interações, artefatos, etc.) a suas atividades em andamento.

No entanto, acreditamos que este método não fornecerá boas soluções no caso de atividades tal como programação, apesar do fato que esta envolve manipulação de documentos em forma digital. Isto porque programação envolve uma linguagem diferente, que normalmente não está presente em conversas entre membros de grupos (salvo em situações quando problemas específicos de programação estão sendo discutidos, e não havia nenhum em nosso conjunto de mensagens de teste). Vemos isto, no entanto, como uma oportunidade para adaptar estes princípios para plataformas de desenvolvimento de software, de modo a integrar sistemas tal como Eclipse e CVS com comunicação entre colaboradores.

A abordagem encontra uma limitação séria em grupos co-localizados, mesmo que apenas parcialmente: o método não funciona bem em cenários onde participantes têm a oportunidade de ser encontrar pessoalmente, uma vez que normalmente as pessoas preferem trabalhar em interações face a face. Uma análise das seqüências de mensagem do caso do artigo de revista mostrou que a maioria das mensagens continha pouco ou nenhum conteúdo sobre o projeto em si, consistindo principalmente de passagens de documentos e programação de encontros (aproximadamente 70%; 22% eram conversas técnicas e o restante eram notificações sobre reuniões que tinham acontecido). Isto indica que esta abordagem funcionará melhor para grupos onde a maioria das conversas acontece por meio eletrônico, produzindo logs que servirão de input para o método.

Nossa abordagem funciona para grupos que interagem, coordenam e trocam idéias usando meios de comunicação textuais baseados em computador (e.g., email, messenger ou fóruns). Notadamente, não funcionaria com vídeo, áudio ou outros meios de comunicação, para os quais outros métodos de análise seriam necessários. Só é possível fornecer recomendações para atividades tais como composição de documentos,

leitura ou navegação web, onde documentos de texto são manipulados e podem ser analisados para correlação.

Uma das dificuldades encontradas ao tentar executar a análise de conteúdo foi o idioma: em nossos conjuntos iniciais de mensagens, as interações foram empreendidas em português, ao passo que os documentos foram escritos em inglês. Esta diferença de idioma pode se tornar uma grande barreira para qualquer sistema adotando esta estratégia. A abordagem também sofre de um problema de inicialização: para estabelecer padrões de interação, são necessários registros de interação prévia, de outro modo o sistema levaria muito tempo para aprender os padrões de trabalho do usuário.

8.2 Comparação com outros métodos/sistemas

Nesta seção apresentamos uma breve comparação com métodos adotados por alguns sistemas correlatos. Estes são sistemas de distribuição de informações relacionadas à consciência ou de apoio à organização do trabalho. Embora não exatamente iguais à nossa proposta, eles compartilham semelhanças, e por este motivo são listados aqui. Uma tabela comparativa é mostrada na Tabela 8-1.

Os seguintes sistemas foram selecionados (todos foram descritos no capítulo 3):

- PIÑAS (MORAN et al, 2002) é um arcabouço para apoiar colaboração potencial e real;
- CommunityBar (MCEWAN, GREENBERG, 2005) um sistema que fornece informação através de diferentes itens de mídia;
- ActivityExplorer (MULLER et al, 2004; GEYER et al, 2006), um sistema que apóia atividades colaborativas informalmente estruturadas ou oportunistas;
- ContactMap (NARDI et al, 2002), um sistema que permite aos usuários visualizar e organizar suas redes de contatos; e
- Elvin (FITZPATRICK et al, 2002), uma estrutura de notificação aplicada ao apoio à consciência e interação

A tabela mostra uma comparação ao longo de 8 eixos:

1. *Determinação de foco* de interesse é a capacidade do software determinar onde a atenção do usuário está sendo alocada;

2. *Considera interações* pergunta se o sistema leva interações em conta em seus cálculos;
3. *Criação de esferas de colaboração* pergunta como esferas de colaboração são criadas, se automaticamente ou manualmente (pelo usuário).
4. *Alocação em esferas de colaboração* preocupa-se com a forma de alocação de um usuário a uma esfera de colaboração;
5. *Seleção de eventos* pergunta como eventos são selecionados para distribuição;
6. *Adaptação a mudanças* preocupa-se com a capacidade dos sistemas se adaptarem a mudanças (novos grupos se formando, dissolução de grupos antigos, etc.);
7. *Servidor central* determina se o sistema ou não é centralizado;
8. *Privacidade* questiona se o sistema se preocupa com privacidade dos usuários.

Tabela 8-1: Análise comparativa entre esferas de colaboração e outros sistemas

	PIÑAS	Community Bar	Activity Explorer	Contact Map	Elvin	Esferas Col.
<i>Determinação de foco</i>	Manual	Manual	Manual	Automático (um tempo)	O manual (regras)	Automático
<i>Considera interações</i>	Não	Não	Não	Sim	Não	Sim
<i>Criação de esferas de colaboração</i>	Manual	Manual	Manual	Manual	Manual	Automático
<i>Alocação em Esferas</i>	Manual	Manual	Manual	Não	Manual	Automático
<i>Seleção de eventos</i>	Automático	Automático	Automático	Não	Automático (regras)	Automático
<i>Adaptação a mudanças</i>	Não	Não	Não	Não	Não	Sim
<i>Servidor central</i>	Não	Não	Sim	Sim	Sim	Não
<i>Privacidade</i>	Não	Sim	Não	Não	Não	Não

8.3 *Trabalhos futuros*

São necessários mais trabalhos no sentido de agregar mais informações à esfera de colaboração, concernente a papéis e hierarquias, designações de tarefa, arranjos, prazos, etc. Análises adicionais das diferentes estruturas encontradas nas redes talvez forneça informações sobre hierarquia e posicionamento dentro do grupo. Outras pesquisas indicam que existem padrões de interação que denotam atribuição de papéis ou organização de equipes, e que análises baseadas em emails podem ser usadas para extraí-los (Fisher, Dourish, 2004). Esperamos que papéis diferentes e padrões de organização tenham necessidades diferentes de informação (e.g., membros centrais e periféricos diferem em termos de natureza, quantidade e profundidade da informação desejado), mas mais pesquisas são necessárias para esclarecer este ponto, estabelecendo os tipos e nível de granularidade da informação a ser distribuída a cada membro de grupo.

Um usuário provavelmente não estará interessado em manter controle próximo das atividades de todos os membros de grupos maiores com quem ele tem pouco em comum, mas pode querer receber resumos periódicos ou relatórios de como o trabalho tem progredido dentro do grupo. Assim, a consciência pode ser vista como um intervalo contínuo, com necessidades amarradas a níveis de participação dos usuários nos grupos. O usuário talvez deseje receber mais ou menos informação (no que diz respeito tanto à profundidade quanto à frequência) sobre seus colegas, dependendo do seu envolvimento. Maiores pesquisas são necessárias neste tema, para melhor determinar a profundidade e nível de detalhe da informação desejada.

Trabalhos futuros também podem entrar em melhores técnicas de análise de texto que levem em conta a semântica das interações. Um das possibilidades é a criação de uma ontologia, relacionando-a a termos para extrair detalhes sobre contratos estabelecidos entre indivíduos (por exemplo, promessas de entrega de parte do trabalho, ou a expectativa que algo estará pronto até uma certa data, existência de estruturas hierárquicas, etc.) Outra possibilidade é usar teorias de atos de fala, como sugerido por Cohen e colegas (2004), para classificar mensagens. Outra abordagem que poderia ser incorporada é apresentada por Raposo e colegas (2004), que usa clichês para extrair compromissos entre sócios de conversas online. Os clichês são máquinas de estado que controlam a seqüência de um diálogo. Em domínios bem conhecidos, pode ser possível aplicar esta técnica para determinar que contratos estão sendo estabelecidos entre

participantes. O uso de uma ontologia ou dicionário de palavras para categorizar ou inferir significado adicional de mensagens poderia ajudar na inferência de papéis e contratos feitos entre partes.

Como esperado, a análise de recursos indica que estes são bons indicativos de conteúdo relevante, uma vez que eles estão diretamente relacionados às decisões dos usuários (selecionando o recurso e os recipientes). Outros exemplos deste tipo de situação (e.g. reenvio de mensagens) também ajudariam a criar uma melhor definição de métodos de distribuição. Os resultados ruins encontrados quando da comparação entre mensagens e recursos leva às possibilidades de analisar os textos de recursos para extrair conteúdo e temas de colaboração, deixando mensagens de fora.

Através do acompanhamento do intercâmbio de recursos, seria possível extrair um padrão de contribuições ao artefato sob construção, seguir mudanças feitas por diferentes membros do grupo (uma análise mais profunda do documento), e mesmo projetar um método para calcular a autoria do documento através da análise da quantidade de trabalho que cada participante contribui (uma questão complexa, que necessita maior reflexão).

Uma análise cruzada de dados estatísticos com a classificação por tipos de thread (e.g., programação, conversas técnicas, preparação de reuniões, etc.) também deve encontrar padrões diferentes de comunicação para diferentes tipos de thread. Estes mais tarde poderiam ser usados para gerar melhores expectativas de tempos e níveis de resposta.

Outra dificuldade se encontra em separar interações pessoais daquelas relacionadas ao trabalho. No entanto, isto pode não ser um problema, visto que relacionamentos normalmente têm uma dualidade inerente, sendo simultaneamente profissionais e pessoais. Deste modo, um laço pode ser fortalecido com informação não relacionada ao trabalho e isto possivelmente irá beneficiar o indivíduo posteriormente. Nardi e colegas (2002) apontam para o fato de que essas atividades de manutenção de rede frequentemente envolvem lembrar de aniversários, ocasiões especiais ou preferências pessoais e enviar cartões, convites ou presentes para fortalecer estes laços. Mesmo que a comunicação pessoal leve a falsos positivos, no caso de um usuário dirigir tanta atenção e tempo a colaborações de uma natureza pessoal, estas também devem ser apoiadas.

Para desenvolver um mapeamento mais completo entre níveis de interação e necessidades de awareness, outras variáveis devem ser levadas em conta, tais como tempo de resposta e comprimento de mensagem. Estes podem indicar prioridades ou papéis (e.g., escrever emails mais longos para o chefe), por exemplo. Tratamento adicional poderia ser incluído para levar em conta tamanhos de mensagem. Mensagens mais longas carregariam mais peso ao definir o conteúdo, enquanto mais curtas teriam peso menor. Um estudo interessante seria o de correlacionar comprimento de mensagem com tipos de thread (organizacional, técnico, etc.), que forneceria uma indicação de com que tipos de mensagens se está lidando e viabilizariam indexação apropriada.

8.3.1 A Questão da Privacidade

Sempre que informações são automaticamente coletadas ou distribuídas, a privacidade dos usuários torna-se um problema. A gestão automática do nimbus do usuário é uma questão que demanda maiores pesquisas. A escolha de que informação disponibilizar pode partir do usuário, ao se adotar um esquema simples em três níveis de privacidade, por exemplo, onde o usuário define se a informação na sua esfera de colaboração é pública (qualquer um pode vê-la), protegida (pares selecionados podem vê-la) ou privada (ninguém pode vê-la). Através deste esquema, o usuário poderia selecionar pares ou recursos e encaixá-los em cada uma das faixas, e determinar quem tem a cesso a que no nível protegido. Quando um evento ocorre que deva ser propagado para outros pares, este seria verificado contra as restrições especificadas para ver em que faixa de privacidade ele se enquadraria e se poderia ser enviado ao agente que o solicitou. O mesmo raciocínio poderia ser aplicado a níveis diferentes de granularidade de informação.

Atividades de usuários normalmente recaem em um das seguintes categorias: manipulação de objetos compartilhados, manipulação de objetos não-compartilhados e interação entre membros. Supomos que todos os objetos compartilhados e interações pertencentes a uma esfera de colaboração possam ser divulgados a membros dessa esfera de colaboração. Os agentes poderiam analisar e remeter um arquivo já enviado a alguns outros como um anexo ou uma conversa entre outros membros da esfera de colaboração.

Organizações modernas são compostas de redes de atores interativos, onde AS relações entre estes estão sujeitas à renegociação constante (Bernoux, 1995). Em geral, conhecimento é trocado e trabalho é realizado através destas relações informais entre trabalhadores, em redes que atravessam limites departamentais, funcionais e organizacionais. Assim, as organizações modernas exigem coordenação e integração de atividades que ultrapassem estes limites, e sistemas de informação devem fornecer apoio para coordenação distribuída e tomada de decisão, ao mesmo tempo em que ajudam o usuário individual na administração das múltiplas colaborações nas quais ele está inserido. Configurações em rede dão origem a grupos fracamente estruturados, onde membros têm a liberdade de se reorganizar conforme necessário para realizar o trabalho.

Nossa abordagem de apoio a grupos fracamente estruturados é baseada em análise egocêntrica de interações baseadas em email. Esta é uma linha promissora de pesquisa, que guarda muitas possibilidades para trabalhos futuros. Muitos estudos vêm aplicando análise de redes sociais para descobrir relações entre as pessoas e padrões de interação, mas poucos usam estes padrões como base para um sistema que ajude o usuário ativamente. Com este trabalho, nós expandimos a teoria de Foco e Nimbus, adicionando métodos para determinação automática do foco de um usuário. A relação entre interação e necessidades de informação precisa ser mais profundamente explorada, e acreditamos que este é um passo relevante na construção de sistemas de apoio ao trabalho colaborativo, especialmente na medida em que o trabalho se torna mais descentralizado e menos estruturado, com atores ganhando autonomia e liberdade para revisar seus arranjos de trabalho.

L'approche présentée est basée sur l'analyse egocentrique d'interactions. Cette direction de recherche peut apporter bons résultats et a plusieurs possibilités pour travaux futurs. Plusieurs recherches ont utilisé l'analyse des réseaux sociaux pour étudier les relations sociales, mais aucune n'a utilisé cette méthode pour faire des inférences dans un système actif. Nous utilisons comme base la théorie focus-nimbus et proposons une extension avec l'addition de méthodes pour la détermination automatique du focus de l'utilisateur.

9 Referencias Bibliográficas

- ALDUNATE, R. NUSSBAUM, M., GONZALÉZ, R. (2002) An Agent Based Middleware for Supporting Spontaneous Collaboration Among Co-Located, Mobile, and not necessarily Known People. In Workshop on Ad hoc Communications and Collaboration in Ubiquitous Computing Environments at the ACM 2002 Conference on Computer Supported Cooperative Work. ACM Press.
- ANDRIESSEN, J.H.E., 2003. *Working with Groupware: Understanding and Evaluating Collaboration Technology*. London: Springer-Verlag.
- BAEZA-YATES, R., RIBEIRO-NETO, B., 1999. *Modern Information Retrieval*. Addison Wesley.
- BERGENTI, F., GARIJO, M., POGGI, A., SOMACHER, M., VELASCO J.R., 2002. Enhancing Collaborative Work through Agents. VIII Convegno dell'Associazione Italiana per l'Intelligenza Artificiale
- BERNOUX, P., 1985, *La Sociologie des Organisations*. 5e ed., Paris, Éditions du Seuil.
- BERNOUX, P., 1995, *La Sociologie des Entreprises*. 2e ed., Paris, Éditions du Seuil.
- BORGES, MRS, BRÉZILLON, P. PINO, JA, POMEROL, J-C., 2004. Bringing Context to CSCW. In Proceedings of the 8th International conference on computer supported cooperative Work in Design (CSCWD'04). IEEE Press.
- CARSTENSEN, P., SCHMIDT, K. (2002). Self Governing Production Groups: Towards Requirements for IT Support. In *5th IFIP International Conference on Information Technology in Manufacturing and Services (BASYS'02)* (pp. 49-60). Netherlands: Kluwer Academic Publishers.
- CASTELLS, M., 1996, *A Sociedade em Rede*, 8ed. São Paulo, Paz e Terra. [portuguese translation of *The Rise of the Network Society*]
- CASTILLA, E.J., HWANG, H. GRANOVETTER, E., GRANOVETTER, M., 2000, Social Networks in Silicon Valley. In Chong-Moon Lee, William F. Miller, Marguerite Gong Hancock, Henry S. Rowen (Eds.) *The Silicon Valley Edge: A Habitat for Innovation and Entrepreneurship*. California, Stanford Business Books.
- COHEN, W.W., CARVALHO, V.R., MITCHELL, T.M. (2004) Learning to Classify Email into Speech Acts. In *Proc. 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.

- CSIKSZENTMIHALYI, M. (1997) *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books.
- DAVENPORT, T.H, BECK, J.C., 2001. *The Attention Economy: Understanding the New Currency of Business*. Boston, MA, Harvard Business School.
- DEERWESTER, S., DUMAIS, S., FURNAS, G. W., LANDAUER, T. K. & HARSHMAN, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the Society for Information Science* 41 (6) pp.391-407. John Wiley and Sons.
- DIX, A., FINLAY, J., ABOWD, G., BEALE, R., 1998, *Human Computer Interaction*. Prentice Hall Europe.
- DOURISH, P., BLY, S., 1992. Portholes: Supporting Awareness in distributed Work Group. In P. Bauersfeld, J. Bennett, & G. Lynch (Eds.) *Proceedings of the 1992 ACM Conference on Human Factors in Computing Systems (CHI'92)* (pp 541-547). New York: ACM Press.
- DOURISH, P., BELLOTTI, V., 1992, Awareness and Coordination in Shared Workspaces (1992). In M. Mantel, & R. Baecker (Eds.) *Proceedings of the 1992 ACM Conference on Computer Supported Cooperative Work (CSCW'92)* (pp 107-114) New York, ACM Press.
- EDWARDS, K. MYNATT, E. (1997). Timewarp: Techniques for Autonomous Collaboration. In *Proceedings of the 1997 Conference on Human Factors in Computing Systems (CHI 1997)*. New York: ACM Press. Retrieved on December 1st, 2006 from <http://acm.org/sigchi/chi97/proceedings/paper/wke.htm>
- ELLIS, C.A., BARTHELMESS, P., QUAN, B., WAINER, J. 2002. NEEM: An Agent Based Meeting Augmentation System. Technical Report CU-CS-937-02, University of Colorado at Boulder, Computer Science Department
- ELLIS, C.A., WAINER, J., (1999). Groupware and Computer Supported Cooperative Work. In Weiss, G. (Ed.) *Multiagent, Systems*, MIT Press
- ENDSLEY, M.R., (2000). Theoretical underpinnings of situation awareness: a critical review. In: Endsley, M.R., Garland, D.J. (Eds.), *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum, Mahwah, NJ, pp. 3–31.
- ENEMBRECK, F., BARTHÈS, J. P. 2002. Personal Assistant to Improve CSCW. *Proceedings of the 7th International CSCWD*, Rio de Janeiro, pp 329 – 335.
- ENGESTRÖM, Y. (2005) Knotworking to Create Collaborative Intentionality Capital in Fluid Organizational Fields. In *Collaborative Capital: Creating Intangible*

- Value. *Advances in Interdisciplinary Studies of Work Teams*, Vol. 11, pp.307-336. Elsevier.
- FISHER, D., DOURISH, P. (2004). Social and Temporal Structures in Everyday Collaboration. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI 2004)* (pp. 551-558). New York: ACM Press.
- FITZPATRICK, G. (1998). *The Locales Framework: Understanding and Designing for Cooperative Work*. Unpublished PhD Thesis, University of Queensland, Australia.
- FITZPATRICK, G., KAPLAN, S., MANSFIELD, T. (1998). Applying the Locales Framework to Understanding and Designing. In *Proceedings of the 1998 Australasian Computer Human Interaction Conference (OzCHI 1998)* (pp. 122-129) Washington: IEEE Computer Society.
- FITZPATRICK, G., KAPLAN, S., MANSFIELD, T., DAVID, A. SEGALL, B. (2002) Supporting Public Availability and Accessibility with Elvin: Experiences and Reflections. In *Computer Supported Cooperative Work* Vol. 11, Issue 3, pp. 447-474, Norwell, MA, Kluwer Academic Publishers.
- FITZPATRICK, G., TOLONE, W., KAPLAN, S. (1995). Work, Locales and Distributed Social Worlds. In *Proceedings of the 1995 European Conference on Computer Supported Cooperative Work (ECSCW 1995)* (pp. 1-16). Berlin: Springer.
- FONER, L. Yenta: a multi-agent, referral-based matchmaking system, *Proceedings of the first international conference on Autonomous agents*, p.301-307. Marina del Rey, California, United States, February 05-08, 1997.
- FREEMAN, L.C. 1979. Centrality in Social Networks: a Conceptual Clarification. *Social Networks*, 1 (1978/79) 215-239. Netherlands, Elsevier Sequoia S.A.
- FRUCHTERMAN, T. REINGOLD, E. (1991). Graph drawing by force-directed placement. In *Software: Practice and Experience*, 21 (11) pp.1129–1164. John Wiley and Sons.
- GABARRO, J.J., 1990. The Development of Working Relationships. In Galleguer, J and Kraut, R.E., Egidio, C. (Eds.), *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*. New Jersey, Lawrence Erlbaum.
- GARTON, L., HAYTHORNTHWAITE, C., WELLMAN, B., 1999, Studying Online Social Networks, in S. Jones (Ed.) *Doing Internet Research*. California, Sage.

- GEYER, W., MULLER, M.J., MOORE, M., WILCOX, E., CHENG, L., BROWNHOLTZ, B., HILL, C.R., MILLEN, D.R., ActivityExplorer: Activity-Centric Collaboration from Research to Product. *IBM Systems Journal – Special Issue on Business Collaboration*, October/November 2006
- GONZALÉS, V.M., MARK, G., 2004 Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2004)* pp. 113-120. Vienna Austria. ACM Press.
- GONZALÉS, V.M., MARK, G., 2005, Managing Currents of Work: Multi-tasking Among Multiple Collaborations. In H. Gellersen et al. (eds.), *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work*, pp. 143–162, Netherlands, Springer
- GRANOVETTER, M., 1973. The Strength of Weak Ties. *The American Journal of Sociology* Vol 78 (6), pp. 1360-1380.
- GRANOVETTER, M., 1983, The Strength of Weak Ties: A Network Theory Revisited. In *Sociological Theory*, Vol. 1, pp.201-233. Michigan, American Sociological Association.
- GREENBERG, S., JOHNSON, B. (1997) Studying Awareness in Contact Facilitation. In *CHI 97 Workshop on Awareness and Collaborative Systems*. Atlanta Georgia.
- GROSS, T. (1999) Supporting Awareness and Cooperation in Digital Information Environments. (1999) In *Basic Research Symposium at the Conference on Human Factors in Computing Systems - CHI'99* (May 16-17, Pittsburgh, PA).
- GROSS, T., OEMIG, C. (2006) COBRA: A Constraint-Based Awareness Management Framework. Technical Report # BUW-CSCW-2006-04, Computer-Supported Cooperative Work Group, Faculty of Media, Bauhaus-University Weimar, Bauhausstr. 11, 99423 Weimar, Germany, Apr. 2006.
- GROTH, K., (2003) Using Social Networks for Knowledge Management. Workshop on Moving From Analysis to Design: Social Networks in the CSCW Context. Held at the 2003 European Conference on Computer Supported Cooperative Work (ECSCW 2003) Helsinki, Finland, 2003.
- GRUDIN, J., 1994 Computer-Supported Cooperative Work: History and Focus. In *IEEE Computer*, number 5 in vol 27, pp. 19–26.

- GUTWIN, C., GREENBERG, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. In *Computer Supported Cooperative Work 11* (pp. 411-446). Netherlands: Kluwer Academic Publishers.
- GUTWIN, C., GREENBERG, S. (2004). The Importance of Awareness for Team Cognition in Distributed Collaboration. In E.Salas and M.Fiore (Eds) *Team Cognition: Understanding the Factors that that Drive Process and Performance* (pp 177-201) Washington: APA Press.
- GUTWIN, C., GREENBERG, S., BLUM, R. DYCK, J., (2005), Supporting Informal Collaboration in Shared-Workspace Groupware. The Interaction Lab Technical Report HCI-TR-2005-01, Canada, University of Saskatchewan.
- HEER, J. BOYD, D. (2005) Vizster: Visualizing Online Social Networks. In InfoVis 2005, IEEE Symposium on Information Visualization.
- HERTEL, G., GEISTER, S., KONRADT, U., 2005, Managing Virtual Teams: A review of current empirical research. In *Human Resource Management Review* 15 pp 69-95. Elsevier.
- ISAACS, E.A., TANG, J.C., MORRIS, T. (1996). Piazza: A desktop Environment Supporting Impromptu and Planned Interactions. In M. Ackerman (Ed.) *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW'96)* (pp. 315-324). New York NY: ACM Press.
- JACOB - <http://danadler.com/jacob/>
- JENNINGS, N.R. (2001) An Agent Based Approach for Building Complex Software Systems. In Communications of the ACM Vol. 44 No. 4. ACM Press.
- JUN-FENG, S., WEI-MING, Z., WEI-DONG, X., GUO-HUI, L., ZHEN-NING, X. (2005) Ontology-based information retrieval model for the semantic Web. In *Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'05)* pp.152 – 155.
- KAUTZ, H, MILEWSKI, A., SELMAN B. (1995) Agent Amplified Communication, AAAI '95 Spring Symposium Workshop Notes on Information Gathering in Distributed, Heterogeneous Environments, Stanford, CA.
- KAUTZ, H., SELMAN, B., SHAH, M. (1997). ReferralWeb: Combining Social Networks and Collaborative Filtering. In Communications of the ACM, Vol. 40, Issue 3 (pp.63-65). New York: ACM Press.
- KIRSCH-PINHEIRO, M., DE LIMA, J.V., BORGES, M.R.S. (2002) A Framework for Awareness Support in Groupware Systems. Proceedings of the 7th

- International Conference on Computer Supported Cooperative Work in Design – CSCWD'2002, Rio de Janeiro, Brazil, September 2002, pp.13-18
- KIRSCH-PINHEIRO, M., DE LIMA, J.V., BORGES, M.R.S. (2003) A framework for awareness support in groupware systems In *Computers in Industry* V 52 (1) Special issue: Knowledge sharing in collaborative design environments pp. 47 – 57, Amsterdam, Netherlands, Elsevier Science Publishers
- KRAUT R.E., EGIDO C., GALEGHER J. (1990a) Patterns of contact and communication in scientific team collaboration. In J. Galegher and R. Kraut (Eds.) *Intellectual Teamwork*. NJ, Erlbaum. pp.149-171.
- KRAUT, R., FISH, R., ROOT, B., CHALFONTE, B. (1990b). Informal communication in organizations: Form, function and technology. In S. Oskamp and S. Spacapan (Eds.), *People's reactions to technology in factories, offices and aerospace*, The Claremont Symposium on Applied Social Psychology, Sage Publications pp. 145-199.
- LASHKARI, Y., METRAL, M., MAES, P. (1994) Collaborative Interface Agents, *Proceedings of the Twelfth National Conference on Artificial Intelligence*, MIT Press, Cambridge, MA, 1994.
- LATOUR, B. (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Harvard University Press.
- LUCENE - <http://lucene.apache.org/java/docs/>
- MAES, P. (1994) Agents that Reduce Work and Information Overload. In *Communications of the ACM* July 1994 Vol. 37, No. 7, 31-40
- MCARTHUR, R., BRUZA, P. (2003) *Discovery of Social Networks and Knowledge in Social Networks by Analysis of Email Utterances*. In *Workshop on Moving From Analysis to Design: Social Networks in the CSCW Context*. Held at the 2003 European Conference on Computer Supported Cooperative Work (ECSCW 2003). Helsinki, Finland.
- MCEWAN, G., GREENBERG, S. (2005a). Community Bar: Designing for Awareness and Interaction. In *ACM CHI Workshop on Awareness Systems: Known Results, Theory, Concepts and Future Challenges*. Held at the 2005 Conference on Human Factors in Computing Systems (CHI 2005). Portland, Oregon.
- MCEWAN, G., GREENBERG, S. (2005b) Community Bar: Places for Collaboration. In Luigina Ciolfi, Geraldine Fitzpatrick and Liam Bannon (Eds) *Workshop*

- Proceedings Settings for Collaboration: The Role of Place, held in conjunction with ECSCW'2005, Paris, Sept 18.
- MCEWAN, G., GREENBERG, S. (2005c) Supporting Social Worlds with the CommunityBar. In Proceedings of GROUP 2005. New York: ACM Press, 2005
- MERCKLÉ, P., 2004, *Sociologie des Réseaux Sociaux*. Paris, La Découverte.
- MIRANDA, M.; XEXÉO, G. B.; SOUZA, J. M. Building Tools for Emergent Design with COPPEER. (2006) In: 10th International Conference o Computer Supported Cooperative Work in Design (CSCWD) Nanjing, 2006.
- MORÁN, A. L., DECOUCHANT, D., FAVELA, J., MARTÍNEZ-ENRÍQUEZ, A. M., GONZÁLEZ-BELTRÁN, B. AND MENDOZA, S. (2002a) PIÑAS: Supporting a Community of Authors on the Web. Proceedings of Fourth International Conference on Distributed Communities on the Web, Sydney, Australia, April 2002.
- MORÁN, A. L., FAVELA, J., MARTÍNEZ-ENRÍQUEZ, A. M., DECOUCHANT, D. (2002b) Before Getting There: Potential and Actual Collaboration. In: Haake, J. M. and Pino, J. A. (Eds.) *Proceedings of the 2002 International Workshop in Groupware (CRIWG 2002) CRIWG 2002*, LNCS 2440, pp. 147 – 167, Springer-Verlag, 2002
- MORAN, T.P. (2005) Unified Activity Management: Explicitly Representing Activity in Work Support Systems. In *Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW 2005) Workshop on Activity: From a Theoretical to a Computational Construct*. Paris.
- MULLER, M.J., GEYER, W., BROWNHOLTZ, B., WILCOX, E., MILLEN, D.R. (2004) One-Hundred Days in an Activity-centric Collaboration Environment based on Shared Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 375-382), New York, ACM Press
- NARDI, B., ENGESTRÖM, Y. (1999) AWeb on the Wind: The Structure of Invisible Work. *Computer-Supported Cooperative Work*, vol. 8, nos. 1–2. Special issue, pp. 1–8.
- NARDI, B., WHITTAKER, S., ISAACS, E., CREECH, M., JOHNSON, J., HAINSWORTH, J., (2002a). Integrating communication and information through ContactMap. In *Communications of the ACM* (1), Volume 45, Number 4, pp. 89-95. New York, ACM Press.

- NARDI, B., WHITTAKER, S., SCHWARZ, H., (2002b). NetWORKers and their activity in Intensional Networks. In *Computer Supported Cooperative Work 11*, pp205-242. Netherlands, Kluwer Academic Publishers.
- NARINE, T. LEGANCHUK, A., MANTEI, M., BUXTON, W. (1997). Collaboration Awareness and its use to consolidate a Disperse Group. In *Proceedings of TC13 International Conference on Human-Computer Interaction (Interact 1997)*. Chapman & Hall.
- NUNAMAKER, J., BRIGGS, R., VREEDE, G.J. (2001) From Information Technology to Value Creation Technology. In Dickinson G.W., DeSanctis, G. (eds.) *Information Technology and the Future Enterprise: new Models for Managers*. New York: Prentice Hall
- OH, A., LOOK, G., HORTON, T., VAN KLEEK, M. Awareness Agents: A Distributed System for Group Awareness. *Proceedings of Student Oxygen Workshop*. 2004
- OLSON, G.M., OLSON, J.S. (2000) Distance Matters. In *Human Computer Interaction*, 2000, vol 15 pp. 139-178. Lawrence Erlbaum Associates.
- OLSON, J. TEASLEY, S. (1996) Groupware in the Wild: Lessons Learned from a Year of Virtual Collocation. In *Computer Supported Cooperative Work'96*, Cambridge, MA, 1996
- PARALIC, J., KOSTIAL, I. (2003) Ontology-based Information Retrieval. In *Proceedings of the 14th International Conference on Information and Intelligent systems*, IIS 2003, Varazdin, Croatia, September 2003, pp. 23 - 28.
- PAYTON, D., DAILY, M., MARTIN, K. (1999) Dynamic collaborator discovery in information intensive environments. In *ACM Computing Surveys (CSUR) Volume 31, Issue 2es* New York, ACM Press.
- PDFBOX - <http://www.pdfbox.org/>
- PERER, A., SHNEIDERMAN, B., OARD, D.W. (2005). Using Rhythms of Relationships to Understand Email Archives. Technical Report TR 2005-82, Institute for Systems Research, University of Maryland. Maryland, USA. Last retrieved December 1st from http://techreports.isr.umd.edu/ARCHIVE/dsp_details.php?isrNum=82&year=2005&type=TR¢er=ISR

- PINELLE, D., GUTWIN, C. (2005). A Groupware Design Framework for Loosely Coupled Groups. In *Proceedings of the 2005 European Conference on Computer Supported Cooperative Work (ECSCW 2005)*. Berlin: Springer.
- PRINZ, W. (1999) NESSIE: An Awareness Environment for Cooperative Settings. In S. Bødker, M. Kyng, and K. Schmidt (eds.) *Proceedings of The Sixth European Conference on Computer Supported Cooperative Work (ECSCW'99)* pp. 391-410 Netherlands, Kluwer Academic Publishers.
- RAMDUNY, D., DIX, A., RODDEN, T. (1998) Exploring the Design Space for Notification Servers. In S. Poltrock, & J. Grudin (Eds.) *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work (CSCW'98)* (pp. 227-235). New York NY: ACM Press.
- RAPOSO, AB, GEROSA, MA, FUKS, H. (2004) Combining Communication and Coordination toward Articulation of Collaborative Activities. In Vreede, GJ, Guerrero, LA, Raventos, GM. (Eds.) *Groupware: Design, Implementation and Use. 10th International Workshop in Groupware (CRIWG 2004) LNCS 3198*, pp.121-136. Berlin, Springer
- RESNICK, P., VARIAN, R.H., (1997) Recommender Systems, *Communications of the ACM*, Vol. 40. No. 3, March 1997.
- RODDEN, T. (1996) Populating the Application: A Model of Awareness for Cooperative Applications. In *Proceedings of the 1996 Conference on Computer Supported Cooperative Work (CSCW 1996)* (pp. 87-96). New York: ACM Press.
- ROMERO, N., MCEWAN, G., GREENBERG, S. (2006) A Field Study of Community Bar: (Mis)-matches between Theory and Practice. Report 2006-826-19, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, T2N 1N4.
- ROOT, R. (1988) Design of a Multi-Media Vehicle for Social Browsing. In I. Greif (Ed.) *Proceedings of the 1988 ACM Conference on Computer Supported Cooperative Work (CSCW'88)* (pp. 25-38). New York, NY: ACM Press.
- RUSSELL, S., NORVIG, P. (1995) *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, Prentice Hall.
- SALTON, G., MCGILL, M.J. (1997) The SMART and SIRE Experimental Retrieval Systems. In *Readings in Information Retrieval*. Karen Jones and Peter Willet, eds. Morgan Kaufmann, San Francisco.

- SALTON, G. (1988) *Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer*. Addison-Wesley Publishing.
- SAXENIAN, A. (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press
- SCHMIDT, K., BANNON, L. (1992) Taking CSCW Seriously: Supporting Articulation Work. *Computer Supported Cooperative Work: An International Journal*, 1992 (1) vol. 1 pp. 7-40
- SCHWARZ, H., NARDI, B., WHITTAKER, S. (1999) The Hidden Work in Virtual Work. Proceedings Critical Management Conference, Manchester, July.
- SCOTT, J. (1991) *Social Network Analysis: A Handbook*. London, Sage Publication.
- SHARDANAND, U., MAES, P. (1995) Social Information Filtering: Algorithms for Automating 'Word of Mouth,' Proceedings of the CHI '95 Conference, 1995.
- SIMMEL, G. (1955) The Web of Group Affiliations, in *Conflict and the Web of Group Affiliations*. Glencoe, Free Press.
- SOHLENKAMP, M. (1998) Supporting Group Awareness in Multi-User Environments through Perceptualization. GMD Research Series Report, No 6, 1999. Paderborn, Fachbereich Mathematik-Informatik der Universität – Gesamthochschule.
- STRAUSS, A. (1993) *Continual Permutations of Action*. New York, Aldine de Gruyter.
- SUCHMAN, L., (1996). Constituting Shared Workspaces. In Engeström, Y. and Middleton, D. (Eds.) *Cognition and Communication at Work*. Cambridge: Cambridge University Press.
- TACLA, C. A.; BARTHÈS, J-P. (2002). A Multi-agent Architecture for Knowledge Management Systems. In *Second IEEE International Symposium on Advanced Distributed Computing Systems, 2002, Guadalajara. ISADS 2002, 2002. p. 1-12.*
- TACLA, C., BARTHÈS, J-P. (2003). A Multi-Agent System for Acquiring and Sharing Lessons Learned. *Computers in Industry*, v. 52, p. 5-16, 2003.
- Tacla, C.A., Enembreck, F. Perception of Centers of Interest. In *Computer Supported Cooperative Work in Design II. 9th International Conference, CSCWD 2005. LNCS 3865. Berlin: Springer Verlag, 2005*
- TAM, J., GREENBERG, S. (2006), A framework for asynchronous change awareness in collaborative documents and workspaces. In *International Journal of Human-Computer Studies* 64 (2006) pp. 583–598. Elsevier

- VANDENBERGHE, F. (2001), *La Sociologie de Georg Simmel*. 1^e ed. Paris, La Découverte.
- VERTEGAAL, R. (2003) Attentive User Interfaces. In *Communications of the ACM*, March 2003, Vol. 46 No. 3, pp. 47- 51. ACM Press, 2003.
- VIVACQUA, A.S., MELLO, C.R., SOUZA, D.K., MENEZES, J.A., MARQUES, L.C., FERREIRA, M.S., SOUZA, J.M. (2007). Time Based Activity Profiles to Recommend Partnership in a P2P Network. In *Proceedings of the 11th Conference on Computer Supported Cooperative Work in Design (CSCWD'07)*. Melbourne, Australia. In Press.
- VIVACQUA, A.S., MORENO, M. SOUZA, J.M. (2005). Using Agents to Detect Opportunities for Collaboration. In *Computer Supported Cooperative Work in Design II. 9th International Conference, CSCWD 2005. LNCS 3865*. Berlin: Springer Verlag.
- WAINER, J., ELLIS, C.A. (1998) Agents in Groupware Systems. *Proceedings CRIWG*, Búzios, RJ, Brasil, 1998.
- WANG, A., CONRADI, R., LIU, C. (1999). A Multi-Agent Architecture for Cooperative Software Engineering *Proceedings of the Third International Conference on Autonomous Agents*, 1999
- WASSERMAN, S., FAUST, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge, Cambridge University Press.
- WELLMAN, B. GULIA, M. (1999) Netsurfers don't Ride Alone: Virtual Communities as Communities. In B. Wellman (org.) *Networks in the Global Village*. Boulder, CO, Westview Press, pp.331-366.
- WELLMAN, B. (2005) Community: From Neighborhood to Network. In *Communications of the ACM*, Vol.48, No 10 (pp.53-55). New York, ACM Press.
- WENGER, E. (1998) *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press, Cambridge, UK.
- WENGER, E., MCDERMOTT, R., SNYDER, W. (2002) *Cultivating Communities of Practice*. Harvard Business School Press, Boston, MA. 2002
- WOOLRIDGE, M. (1997) Agent Based Software Engineering. In *IEEE Proceedings of Software Engineering 144*, pp.26-37. IEEE Press.

ZACKLAD, M. (2003). *Communities of Action: a Cognitive and Social Approach to the Design of CSCW Systems*. In *Proceedings of GROUP'03*, New York, ACM Press.

ZAGER, D. (2000) *Collaboration as Activity*. In *Computer Supported Cooperative Work*, vol 11, Nos 1-2 pp.180-204. 2000

Anexo I

Este apêndice apresenta uma lista de publicações advindas desta tese e de outras pesquisas correlatas, conduzidas durante o desenvolvimento desta.

Journals

- VIVACQUA, A.S., DE SOUZA, J.M. (2008) The Vineyard Approach: A Computational Model for Determination of Awareness Foci in E-Mail-Based Collaboration In International Journal of e-Collaboration, 4(1), 41-59, January-March 2008. *To appear*
- PINHO, D.; VIVACQUA, A. S. ; MEDEIROS, S. ; SOUZA, J. M. (2006) SYMBAD Similarity based agents for design. Expert Systems with Applications, v. 31, p. 728-733, 2006.
- TORNAGHI, A.; VIVACQUA, A. S. ; SOUZA, J. M. (2005) Creating educator communities. International Journal of Web Based Communities, v. 1, p. 296-307, 2005.
- XEXEO, G. B.; VIVACQUA, A. S. ; SOUZA, J. M. ; BRAGA, B. ; DALMEIDA JR, J. N. ; ALMENTERO, B. K. ; CASTILHO, R. ; MIRANDA, B. (2005) COE: A collaborative ontology editor based on a peer-to-peer framework. Advanced Engineering Informatics, v. 19, p. 113-121, 2005.

Post-Proceedings

- VIVACQUA, A. S. ; MORENO, M. ; SOUZA, J. M. . Using Agents to Detect Opportunities for Collaboration. In: Weiming Shen; Kuo-Ming Chao; Zongkai Lin; Jean-Paul A. Barthès; Anne James. (Org.). Computer Supported Cooperative Work in Design II. Berlin/Heidelberg: Springer-Verlag, 2006, v. 3865, p. 244-253.
- PINHO, D. ; VIVACQUA, A. S. ; MEDEIROS, S. ; SOUZA, J. M. . Similarity Based Agents for Design. In: Weiming Shen; Zongkai Lin; Jean-Paul A. Barthès; Tangqiu Li. (Org.). Computer Supported Cooperative Work in Design I. Berlin/Heidelberg: Springer-Verlag, 2005, v. 3168, p. 117-126.
- PINHO, D. ; VIVACQUA, A. S. ; MEDEIROS, S. ; SOUZA, J. M. . Agents to Foster Conscious Design and Reuse in Architecture. In: Paolo Bresciani. (Org.).

Agent-Oriented Information Systems II (LNCS). Berlin/Heidelberg: Springer, 2005, v. 3508, p. 212-226.

Conferências

- VIVACQUA, A. ; PINHEIRO, W. A. ; Barros, R. ; Mattos, A.S. ; Cianni, N.M. ; Monteiro Jr., P.C.L. ; MARTINO, R. ; Marques, V. ; XEXEO, G. B. ; SOUZA, J. M. ; SCHNEIDER, D. . Dynaflow: Agent-Based Dynamic Workflow Management for P2P Environments. In: ICEIS, 2007. Proceedings of 9th International Conference on Enterprise Information Systems, 2007.
- VIVACQUA, A. ; MELLO, C. ; SOUZA, D. K. ; Menezes, J.A.A. ; Marques, L.C. ; Ferreira, M.S. ; SOUZA, J. M. . Time Based Activity Profiles to Recommend Partnership in a P2P Network. In: CSCWD, 2007, Melbourne. Proceedings of 11th International Conference on Computer Supported cooperative Work in Design, 2007.
- CARVALHO, G. ; VIVACQUA, A. ; SOUZA, J. M. ; MEDEIROS, S. P. J. . Large Scale Decision Making in Participatory Environmental Design. In: CSCWD, 2007, Melbourne. Proceedings of 11th International Conference on Computer Supported Cooperative Work in Design, 2007.
- VIVACQUA, A. S. ; RODRIGUES NETO, J. A. ; SOUZA, J. M. . FoxPeer: Navigating the Web with Community Recommendations. In: IADIS Conference on Web Based Communities (WBC'07), 2007, Salamanca, Spain. Proceedings IADIS Conference on Web Based Communities (WBC'07), 2007.
- VIVACQUA, A. S. ; SOUZA, J. M. . Using Email-Based Network Analysis to Determine Awareness Foci. In: International Workshop in Groupware CRIWG 2006, 2006, Medina del Campo, Spain. Groupware: Design, Implementation, and Use (LNCS). Berlin/Heidelberg : Springer, 2006. v. 4154. p. 78-93.
- PINHO, D. ; VIVACQUA, A. S. ; MEDEIROS, S. ; SOUZA, J. M. . Ontology Based Design for Similarity Discovery in SYMBAD Project. In: 10th International Conference o Computer Supported Cooperative Work in Design, 2006, Nanjing, China. 10 th International Conference o Computer Supported Cooperative Work in Design, 2006. v. 1. p. 477-482.
- VIVACQUA, A. S. ; BARTHES, J. P. ; SOUZA, J. M. . A Framework to Support Self-Governing Software Design Groups. In: 10th International Conference o Computer Supported Cooperative Work in Design, 2006, Nanjing, China. 10th

- International Conference o Computer Supported Cooperative Work in Design, 2006. v. 1. p. 375-380.
- PINHO, D. ; VIVACQUA, A. S. ; MEDEIROS, S. ; SOUZA, J. M. . Similarity Based Agents for Design. In: International Conference on Computer Supported Cooperative Work in Design, 2005, Coventry, England. 9th International Conference on Computer Supported Cooperative Work in Design. Coventry : Coventry University Press, 2005. v. 1. p. 417-422.
- VIVACQUA, A. S. ; MORENO, M. ; SOUZA, J. M. . CUMBIA: An Agent Framework to Detect Opportunities for Collaboration. In: International Conference on Computer Supported Cooperative Work in Design, 2005, Coventry. 9th International Conference on Computer Supported Cooperative Work in Design. Coventry : Coventry University Press, 2005. v. 1. p. 434-439.
- VIVACQUA, A. S. ; BARTHES, J. P. ; SOUZA, J. M. . Towards a Model of Cooperation. In: International Workshop in Groupware, 2005, Porto de Galinhas, Brazil. Groupware: Design, Implementation, and Use (LNCS). Berlin/Heidelberg : Springer, 2005. v. 3706. p. 359-366.
- VIVACQUA, A. S. ; SOUZA, J. M. . Cooperation for Creativity in Design. In: 8th International Conference on CSCW in Design, 2004, Xiamen, China. International Conference on CSCW in Design, 2004.
- VIVACQUA, A. S. ; PINHO, D. ; SOUZA, J. M. ; MEDEIROS, S. . Improving Project Design through Process Awareness Agents. In: 8th International Conference on CSCW in Design, 2004, Xiamen, China. International Conference on CSCW in Design, 2004.
- XEXEO, G. B. ; SOUZA, J. M. ; VIVACQUA, A. S. ; MIRANDA, B. ; BRAGA, B. ; ALMENTERO, B. K. ; DALMEIDA JR, J. N. ; CASTILHO, R. . Peer to Peer Collaborative Editing of Ontologies. In: 8th International Conference on CSCW in Design, 2004, Xiamen, China. International Conference on CSCW in Design, 2004.
- PINHO, D. ; VIVACQUA, A. S. ; SOUZA, J. M. . Agents to Foster Conscious Design and Reuse in Architecture. In: Sixth International Bi-Conference Workshop on Agent Oriented Information Systems (AOIS 2004) at the 16th Conference on Advanced Information Systems Engineering, 2004, Riga, Latvia. 16th Conference on Advanced Information Systems Engineering Workshop

- Proceedings. Riga : Faculty of Computer Science and Information Technology - Riga Technical University, 2004. v. 2. p. 355-369.
- TORNAGHI, A. ; VIVACQUA, A. S. ; SOUZA, J. M. . Creating Educator Communities. In: IADIS International Conference Web Based Communities, 2004, Lisbon, Portugal. IADIS International Conference Web Based Communities 2004, 2004. p. 123-130
- VIVACQUA, A. S. ; SOUZA, J. M. . CreEx: A Framework for Creativity in Cooperative Problem Solving. In: 10th International Workshop in Groupware (CRIWG'04), 2004, San Carlos, Costa Rica. Groupware: Design, Implementation and Use (LNCS). Berlin/Heidelberg : Springer, 2004. v. 3198. p. 166-173.
- VIVACQUA, A. S. ; SOUZA, J. M. . Fostering Creativity in Cooperative Design. In: Cooperative Design Visualization and Engineering, 2004, Palma de Maiorca, Spain. Cooperative Design, Visualization, and Engineering (LNCS). Berlin/Heidelberg : Springer, 2004. v. 3190. p. 115-122.
- OLIVEIRA, J. ; SOUZA, J. M. ; VIVACQUA, A. S. . Uma Abordagem de Criação e Edição Colaborativa de Ontologias via Web. In: Simposio Internacional da Gestao do Conhecimento, 2003, Curitiba. ISKM 2003, 2003.
- MORENO, M. ; VIVACQUA, A. S. ; SOUZA, J. M. . An Agent Framework to Support Opportunistic Collaboration. In: 9th International Workshop in Groupware (CRIWG'03), 2003, Autrans, France. Groupware: Design, Implementation, and Use (LNCS). Berlin/Heidelberg : Springer, 2003. v. 2806. p. 224-231.
- VIVACQUA, A. S. ; MATTOS, F. ; TORNAGHI, A. ; SOUZA, J. M. ; CUKIERMAN, H. . Perspectives on Creativity in Web Learning. In: International Conference in Web-Based Learning (ICWL'03), 2003, Melbourne, Australia. Advances in Web-Based Learning (LNCS). Berlin/Heidelberg : Springer, 2003. v. 2783. p. 145-156.
- VIVACQUA, A. S. ; MORENO, M. ; SOUZA, J. M. . Profiling and Matchmaking Strategies in Support of Opportunistic Collaboration. In: Cooperative Information Systems (COOPIS'03), 2003, Catania, Italy. On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE (LNCS). Berlin/Heidelberg : Springer, 2003. v. 2883. p. 162-177.

Relatórios Técnicos

- MELLO, C. E. R. ; SOUZA, D. K. ; MENEZES, J. A. A. ; MARQUES, L. C. ; FERREIRA, M. S. ; VIVACQUA, A. S. ; SOUZA, J. M. . Navigator: uma Ferramenta para Colaboração Oportunística (ES-704/06). Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2006 (Relatório Técnico).
- VIVACQUA, A. S. ; SOUZA, J. M. . CreEx: A Framework for Creativity in Cooperative Problem Solving (ES-642/04). Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2004 (Relatório Técnico)
- MORENO, M. ; VIVACQUA, A. S. ; SOUZA, J. M. . CUMBIA: Una Arquitectura para Soportar Colaboración Oportunística y Espontánea (ES-600/03). Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2003 (Relatório Técnico).
- MORENO, M. ; VIVACQUA, A. S. ; SOUZA, J. M. . An Agent Framework to Support Opportunistic Collaboration (ES-599/03). Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2003 (Relatório Técnico).
- VIVACQUA, A. S. ; MATTOS, F. ; TORNAGHI, A. ; SOUZA, J. M. ; CUKIERMAN, H. . Perspectives on Creativity in Web Learning (ES-595/03). Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2003 (Relatório Técnico).

Anexo II: Versão da Tese em Inglês

**An Approach for Information Dissemination
in Collaborative Environments**

I. Introduction

The wide dissemination of computer technology and of the internet has led to an increase in computer mediated personal and working relationships (WELLMAN, 2005), as evidenced by the success of networking sites such as Orkut (www.orkut.com) or Friendster (www.friendster.com). Studies show increased organizational adoption of virtual work teams, in which members are geographically dispersed and communicate and coordinate mainly via electronic communication tools (HERTEL et al., 1995). The widespread adoption of communication and networking tools has effectively interconnected individuals, who now use email, messenger and other electronic media to meet, interact and get work done. As a result, co-workers who used to be collocated are now spread throughout the world, and proximity is no longer an excluding factor in the establishment of relationships or in the formation of communities.

Adoption of networking technologies, distribution of organizations and subcontracting have also led to a decentralization of work: workers now participate in several collaborative projects at a time, and may be involved with different individuals who, in turn, are also engaged in other projects (GONZALÉS, MARK, 2005). Just as organizations bring people together to work on different projects, people participate in several projects at the same time, dividing their time and attention accordingly (MORAN, 2005).

Motivated by increased complexity in the environment and worker specialization, organizations often constitute multidisciplinary problem solving teams. Thus, modern knowledge and information workers frequently find themselves engaged in collaborative situations. Projects often involve multiple people, who must work together to accomplish joint goals. Even in apparently solitary environments, such as research and academia, there are strong social elements that influence the outcomes of the work in progress (LATOURET, 1987). Other studies have emphasized the importance of social networks and how they contribute towards the creation of a dynamic, creative environment (SAXENIAN, 1994; CASTILLA et al., 2000).

The rise of networking technologies has changed the way people work, and networks have become commonplace (CASTELLS, 1996), with individuals navigating personal social networks (WELLMAN, GULIA, 1999) and creating “intentional networks” to get work done (NARDI et al., 2002b). To function in this networked

structure, individuals need to manage their relations to others and their roles and tasks in the projects they are active in. They must organize themselves and their work to accomplish different tasks, very often with different collaborators. Participation in different groups usually means that, depending on the situation, an actor might have different roles and obligations, perform different tasks and work towards different goals, which must be managed so they do not conflict with each other.

Even in collaborative situations, much work is still undertaken individually, and individuals alternate working alone with close collaboration and information sharing (GUTWIN, GREENBERG, 2002; GONZALÉS, MARK, 2005). The interdependency between group activities leads to a need for articulation of efforts (SCHMIDT, BANNON, 1992): actors need to divide, allocate, coordinate, schedule and interrelate their activities. To be able to articulate activities to get work done, group members need to exchange information regarding aspects of the work in progress, inform others of their individual actions, discuss the state of the joint project or consult with others (SCHMIDT, BANNON, 1992, p.13). Knowledge of other people's capabilities, activities and availability, as well as about the overall progress of the work, competencies required, unforeseen alterations or corrections is necessary for the group to better work together, as it allows members to keep track of the joint project and of each others' activities, manage coupling levels (i.e., decide when to move into closer cooperation and when to part into individual work) and reorganize when necessary. This knowledge is commonly called workspace awareness (GUTWIN, GREENBERG, 2002).

Self-governing groups are groups in which actors often have control over job allocation and day-to-day operations. They emerge out of a need to handle unpredictable events or contingencies, sometimes deviate from pre-established norms and rules (CARSTENSEN, SCHMIDT, 2002). Even in controlled environments such as shop-floor operation, teams have been found to have a large amount of independence and control over their own work (CARSTENSEN, SCHMIDT, 2002, p.50). Information is necessary to enable these adjustments and self-governing structures.

A recurring concept in group work, awareness is an understanding of activities of others, which provides a context for one's own activity (DOURISH, BELLOTTI, 1992). This involves information about others, their tasks and the work environment as a whole. Awareness is obtained through the perception of information gathered from the

environment, collected and transmitted through interactions with the environment, and it must be maintained as the environment changes (GUTWIN, GREENBERG, 2002). This information is important for the group to engage in their activities more easily, allowing individuals to adjust their behaviors according to the situation, and generating synergy within the group (GUTWIN, GREENBERG, 2004). In the absence of strictly defined (and followed) roles and hierarchies, this information becomes even more important, as independent actors have to negotiate with each other regarding their work in order to come to their final goal. Decisions are often the result of arrangements between group members, as is the work that finally gets done (BERNOUX, 1995).

In collocated environments, much of the necessary information can easily be disseminated and perceived with little additional effort (KRAUT et al., 1990a). This picture changes when the group moves to a distributed environment. While staying aware of others is taken for granted in everyday interactions, it is a complex problem in computational environments, because communication and interaction resources are poorer than those in face-to-face situations (GUTWIN, GREENBERG, 2002). Furthermore, distance also has a negative effect on awareness maintenance, as individuals have difficulty keeping abreast of each other when they are not in the same physical environment (KRAUT et al., 1990a). As awareness is achieved through the gathering and processing of information, the process of maintaining awareness necessarily involves providing information to users. An individual's awareness of the group endeavor will be as good as the information he or she has access to.

The large amount of information available means that users must choose between the many items that request their attention, as, to a large extent, information about the activities of others is irrelevant and can sometimes hinder an individual's work (SOHLENKAMP, 1998). Deciding what should receive one's attention has become a frequent task in information workers' environments (DAVENPORT, BECK, 2001). This creates a need for the careful control of information flow, in order to provide relevant information and minimize disruption. Information provided by automated systems must be filtered down to that which is of interest. The issue of what information is pertinent to a user is important and must be addressed. This is the main thrust of this research, as presented in the following section.

1.1 Philosophical Basis

Organizational sociology views modern organizations as networks of interacting actors, where relations between these actors are subject to constant renegotiation (BERNOUX, 1995). These ties go beyond organizational structures, protocols and hierarchies, sometimes surpassing them. More often than not, knowledge is exchanged and work is undertaken through these informal relations between workers (for an early study, see GRANOVETTER, 1983). They cut across departmental, functional and organizational boundaries, and constitute a new way of working to which the organization must adapt. The networks influence actors' behavior in organizing and conducting work, their relations with each other are also strongly affected by the network as a whole (SIMMELL, 1955).

It has been noted by researchers that, when people collaborate, they often shift back and forth between individual and shared work, and between loosely and tightly coupled collaboration, especially when individual tasks have low interdependence (GUTWIN, GREENBERG, 2002). Gutwin and colleagues (2005) argue that a sensible approach in these cases is to support individual work and provide tools to enable a shift into collaboration when necessary. As the individual becomes the nexus of his or her collaborative network, it seems reasonable to adopt a user-centric approach to assist the user with the management of the many arrangements and interdependencies established with others. Schmidt and Bannon (1992) conceptualize CSCW as a field that should be concerned with cooperative work arrangements, and make a case for systems that support articulation by workers themselves, so that they can adapt their work division and procedures to compensate for unforeseen events.

Following these considerations, we depart from perspectives that analyze the group as a whole and pursue an approach that moves from the individual to the group, using individual relations to others as the basic element of analysis. This is consistent with the position of the individual as a node in a network of ties (and the central node in a personal network), and provides a direct way to manage one's multiple relations and involvement in different groups. Juggling multiple roles and assignments has become a constant in information workers' lives: these individuals must multitask between multiple *working spheres* (work units tied to particular collaborative structures), shifting their attention from one project to another (GONZALÉS, MARK, 2004). Our goal is to design systems that help users manage their multiple commitments and relationships to

other individuals and groups and keep in step with the multiple activities happening within the different working spheres. We take a step towards bridging the gap between individual and joint work, by helping each person understand how their current work relates to others' work and maintain awareness of the collaborative project.

1.2 Research Statement and Goals

Multitasking among different working spheres involves constant remembering of people and work states within the network and communicating to create and maintain ties or get work done (NARDI et al., 2002). This implies a need to keep track of the state of the network and of which contacts are active and in which contexts (performing which activities/projects).

The main theme of this research is the provision of computational support for the distribution of awareness information to an individual enmeshed in multiple working spheres. We approach this issue through the analysis of activity within the network, which entails the determination of which parts of one's network are active. Starting from an individual viewpoint, we introduce the concept of collaboration sphere, to bring several working spheres together. Constituting a collaboration sphere involves determining which individuals are working together, which resources are involved, and maintaining an overview of the working sphere involves determining when individual work falls within these collaboration spheres. This type of information could facilitate the coordination of the group, enabling spontaneous adjustments when necessary (SCHMIDT, BANNON, 1992).

Increasing group members' awareness of the group's activities means tracking and disseminating activity data. Distributing activity data involves creating a system that can figure out who is related to certain collaboration spheres and what activities relate to an ongoing collaborative project. Our objects of study are the individuals that compose workgroups, performing individual and cooperative tasks related to a collaborative project; the interactions they engage in and the artifacts they manipulate.

In a computational environment, a large amount of data is available for collection, processing and distribution. Data such as time spent on certain tasks, files worked on, file sizes and timestamps, interaction logs, sites visited, contributions to fora or participation in mailing lists can be automatically collected from the systems and made available to users without their intervention. However, more data does not mean

more information, as not all data is useful to an individual in a given moment (ENDSLEY, 2000). Handling the large amount of information individuals are exposed to on a daily basis has become a hard task. Information consumes attention, and the increase in the available information creates a demand for attention that is hard to meet (VERTEGAAL, 2003).

We present an agent-based architecture to support automatic collection and distribution of data from users' individual workspaces. Agents create user models based on each user's interactions, where ties to others are established and change according to the situation. These models are dynamic in that they are automatically adjusted with time. As not all data is relevant, we present a method to filter relevant information. This combination should enable a computational system to proactively analyze and disseminate information that a user might be interested in without overwhelming the user with too much information.

Our approach is applicable in the context of knowledge-intensive activities that involve information manipulation in digital form. It is meant for distributed groups, which interact online using text based media, thus providing subsidies for analysis. To reach our goals, we have:

- reviewed bibliography on collaborative work and self organization in distributed groups, relating case studies with philosophical and sociological observations;
- drawn from the theoretical points of view a technological approach to support networked collaboration;
- performed observations of working groups, both in distributed and collocated interaction, to uncover their needs and requirements for systems support;
- developed an architecture/framework to support automatic capture and dissemination of information, as part of a technical solution;
- developed a method to detect active links in a user's intensional network;
- developed a method to relate activities and people to their collaboration spheres, which can be used in a collaborative setting;
- constructed proof of concept prototypes to verify and refine our ideas.

These are further explained throughout this document.

I.1.1 Applicability and Limitations

This approach was designed to address certain characteristics. Group activities are often classified according to participants' physical location and temporal participation in the group (DIX et al., 1998). Our research focuses on distributed groups, where individuals aren't necessarily in the same place when working, regardless of time (that is, they may be working simultaneously or not). In our investigations, we found that the approach functions better when there are few opportunities for face to face interactions.

Table I-1: Time x Space Matrix (adapted from DIX et al., 1988 and GRUDIN, 1994). Non-computational technology noted in italics.

	Same Place	Different Place
Same Time	<i>Face to face conversation</i> Meeting facilitation	<i>Telephone</i> Chats
Different Time	<i>Post-it note</i> Team rooms	<i>Letter</i> Threaded Forum

Another categorization defines three types of collaborative effort to which groupware systems can be directed (NUNAMAKER, et al. 2001). These may be:

- collective (where each individual works alone and the pieces they produce are put together);
- coordinated (where group members make individual efforts and need to coordinate to achieve their goal);
- concerted, where all members must make a synchronous effort towards a shared goal.

We focus on coordinated efforts, especially in groups where individuals undertake certain tasks individually and must then coordinate with others in order to

accomplish joint goals. In this case, productivity is dependent on the level of individual effort and on the coordination between individuals.

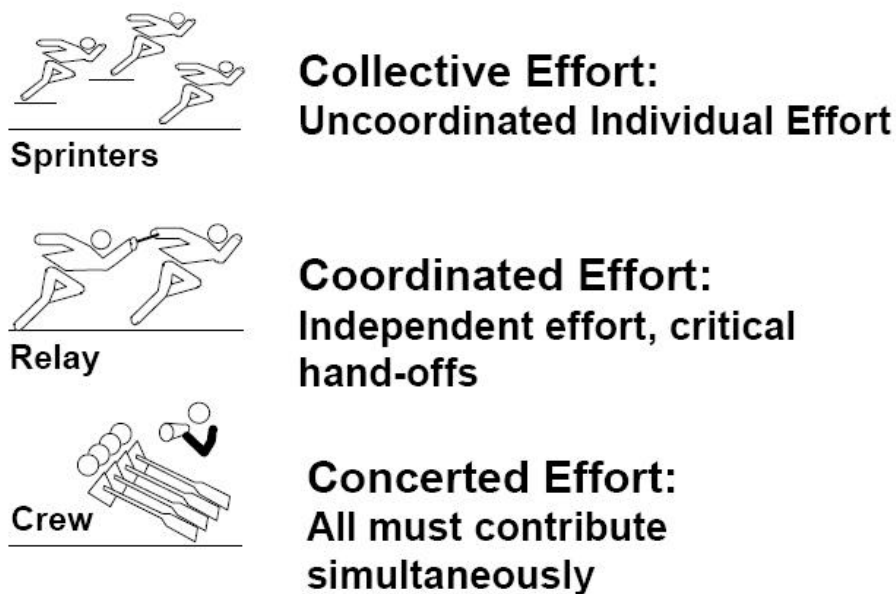


Figure I-1: Different types of effort (NUNAMAKER et al., 2001).

I.1.2 Technical Approach

One of the techniques used by sociologists to analyze group and individual interrelations, Social Network Analysis maps groups of people as networks, where members are tied through relationships that represent aspects under study (SCOTT, 1991; WASSERMAN, FAUST, 1994). This technique employs graph operations and specific calculations to detect a node's importance, tie strength and importance within the group. Analyses can be conducted based on the whole network (*sociocentric* analysis) or on an individual point of view (*egocentric* analysis) (GARTON et al., 1999).

Given that the individual is the central node in a personal network and handles different collaborations, we have adopted egocentric social network and interaction analysis as our basis. This enables us to emphasize the central role of the individual in the management of his or her ties and his or her allocation of attention while working. Our approach is to analyze individual ties to others to elicit collaboration patterns and activities that pertain to these. This analysis takes into account each individual's allocation of attention to determine interest and urgency, as peers who receive more attention are probably the ones a user is currently collaborating with. Additionally,

content analysis is used to relate individual activities to interpersonal ties, in order to determine an artifact's social context and other individuals involved in it.

For our prototype system, an agent-based design was used to implement a peer-to-peer architecture. We develop an awareness scenario, where agents gather and distribute information to their users, according to their analysis of links to others and of information manipulation. Our studies and analyses, which will be presented in later sections, indicate that this is a feasible technique.

Awareness has been the object of many studies, and different group support systems have been proposed to provide awareness-enabling information to their users. To be useful to a user, however, information needs to be provided at the correct time, depending on activities and overall context. In this work, we consider awareness information distribution as it fits into a collaborative context. Thus, maintaining awareness in an intensional network situation involves discovering links to others and what they refer to. Our proposed information filtering method applies interaction and social network analysis: subnets to which the user devotes the most attention are the currently active ones, which demand coordination and synchronization. Activities performed by members of these subnets that relate to the topics of discussion between them are those the user might need to know about.

This document is organized as follows: we start out by providing motivation and a general introduction, setting the stage and presenting the problem and our goals (Section 1). This is followed by a theoretical exploration into the concepts and theories involved, which should help the reader understand our line of thought, the sociological and organizational concepts and ideas involved, and how these develop into an approach and technical solution to the problem under study (Section 2). This conceptual framework is followed in Section 3 by an investigation and summarization of related technical solutions and approaches to, if not exactly the same problem, similar or related ones, including recommender and matchmaking systems and awareness support systems. The discussion is then shifted towards some case studies undertaken to better understand loosely structured work in Section 4. A proposed architecture is presented in Section 5 and followed by a method for determination of collaborative contexts in Section 6. This is then followed by analyses in Section 7 and a discussion of applicability of the proposed solution and future avenues for work in Section 8.

II. Theoretical Background

This chapter presents the theoretical underpinnings of our research, to pave the way for the presentation of our conceptual approach and technical solution in the following chapters. We present concepts and techniques from sociology and organizational theory that discuss individuals and organizations, and how people associate to conduct work. We also present a framework developed for design of group support tools and social network analysis as a technique, and prior work on effects of distance and information that influenced ours.

I.3 Sociological and Organizational Basis

The dissemination of network technology has brought about organizational changes leading to new, networked structures. The network way of functioning, in which elements are interconnected and the intelligence is on the network itself, is now applicable to any type of activity, segment or location that can be connected electronically, and has been described extensively in (CASTELLS, 1996). This network form is well adapted to the growing complexity of the environment and provides added flexibility to organizations (BERNOUX, 1985).

Networks exist in a global level, where enterprises connect and interact with other organizations, governments and press. In this network economy, actors are strongly interdependent, as organizations (especially small and medium sized ones) form networks to attain their objectives (CASTELLS, 1996, p.218), subcontracting and collaborating in different parts of the production cycle to gain market penetration. Strategic alliances have become a common way for organizations to function in the networked world.

Networks are also found within organizations, as new structural forms are adopted: enterprises seem to have shifted from vertical bureaucracies to horizontal structures, aiming to streamline production, gain performance and cut costs (CASTELLS, 1996, p.221). Technology and increased competitiveness in the environment have changed the way people work, driving enterprise restructuring and reinforcing the individualization of work within work processes. In the new organization, administration is decentralized, work is individualized and markets are personalized (CASTELLS, 1996, p.330). Information technology simultaneously

enables task decentralization and coordination in real time interactive communication networks, global or local. As a rule, the traditional forms of work are slowly being substituted for more flexible and individualized contracts (CASTELLS, 1996, p.337).

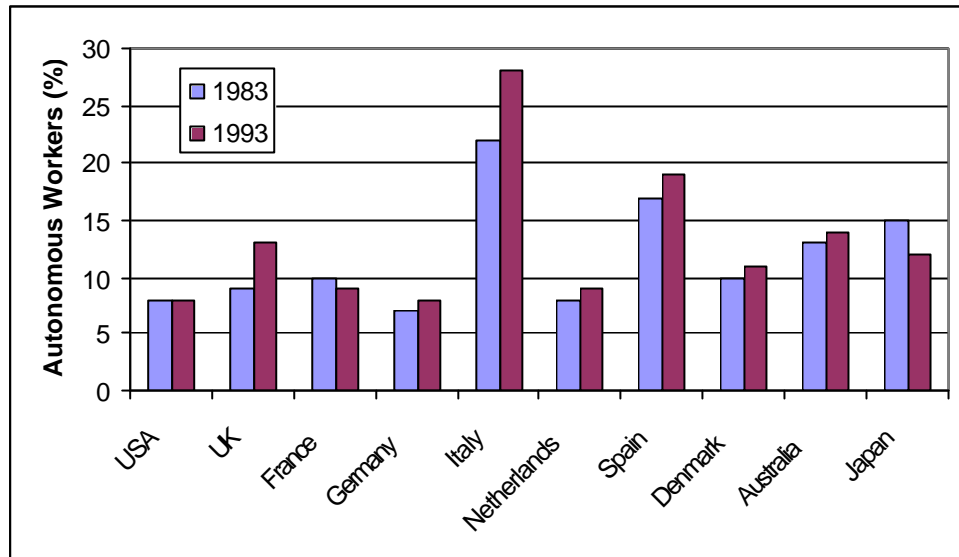


Figure II-1: Autonomous workers as a percentage of the workforce (from Castells, 1996, p332)

Consistent with this view, Bernoux (1995) describes organizations as networks of interacting actors. These actors are autonomous and each has a personal view of the organization, according to his or her objectives and needs. Associations with others are governed by contracts and rules of their own making, subject to constant negotiation and redefinition as the situation demands. The organization becomes a system of social ties, where actors create the system, the rules, the network and their own interpretations of it (BERNOUX, 1995, p.173, BERNOUX, 1985). Not only do individuals define the group, setting rules, methods and protocols, but the group also influences the individual, through these same rules and protocols (SIMMEL, 1955). Thus, an individual changes/defines and is constrained by the groups in which he or she participates.

More often than not, knowledge is exchanged and work is undertaken through informal relations between workers, who organize themselves in communities of practice (WENGER et al., 2002, WENGER, 1998) or use their personal networks to elicit information or get work done (GRANOVETTER, 1973). Thus, an organization provides structures and regulations, but functions as much as a result of actors' interactions and relations as of the organizational structure provided (BERNOUX, 1995).

An organization can thus be seen as a coalition of individuals motivated by individual interests (BERNOUX, 1985). The formal organization becomes a contractual governance structure that influences the behavior of individuals and collectives according to the interests of the other members of the organization and external stakeholders. Work is conducted through the establishment of cooperative work arrangements, which emerge to handle specific requirements and dissolve when they are no longer necessary. These arrangements are organized according to specific needs, following established patterns (SCHMIDT, BANNON, 1992).

It is these networks that enable the organization to quickly adapt to changes in the environment (BERNOUX, 1995). They cut across departmental, functional and sometimes even organizational boundaries, and constitute a new way of working to which the organization must adapt. Therefore, modern organizations require coordination and integration of activities across functional and professional boundaries, and individuals within these organizations need tools to better understand and work in this networked configuration. Traditional group support systems have focused on the development of effective models of the structures and processes relevant to the organization's work and adequate ways to present them, but haven't provided much in the way of systems to support these informal networks or to enable them to work more effectively (SCHMIDT, BANNON, 1992).

Even in rigid environments such as shop-floor situations, there has been an increase in worker autonomy, with ensuing performance growth in large organizations (CASTELLS, 1996, p.310). Carstensen and Schmidt (2002), describe self-governing groups as those in which actors have control over job allocation, day-to-day planning and control. These groups emerge out of a need to handle unpredictable events or contingencies, and enable an organization to quickly adapt to new demands generated by the environment, sometimes deviating from pre-established norms and rules. In many cases, groups are composed of peers, where there is no formal hierarchical structure. In this scenario, many of the decisions are the result of arrangements between peers, as is the work that finally gets done (BERNOUX, 1995).

Due to the underlying interdependence between their tasks, workers have to articulate (i.e., divide, allocate, coordinate, schedule, interrelate, etc.) their activities (SCHMIDT, BANNON, 1992). The organization provides a relatively stable pattern of cooperative arrangements and a governance structure to regulate the diverse interests

within the cooperative ensemble, but new collaborative work groups emerge because actors could not accomplish certain tasks as well if they were to do it individually (SCHMIDT, BANNON, 1992). In most situations, individuals have some level of control over their work, and are capable of adjusting for errors or unusual circumstances, handling unpredictable events or contingencies.

Information is paramount for the functioning and organization of elements in the network structure. Castells (1996, p.119) underlines the importance of information in the new economy, as productivity and competitiveness of units or agents in this economy are strongly dependent on their capacity to efficiently generate, process and apply knowledge based information.

Personal and Intensional Networks

Individuals may occupy different positions in the organizational structure, but have three main roles within the networks: active workers, who establish connections on their own and navigate the network as needs dictate; passive workers, who are connected to others but do not decide when, how, to whom or why and unconnected workers are those stuck with their specific tasks, defined by unilateral, non-interactive instructions (CASTELLS, 1996). This presents new challenges for active workers, who must manage ties to others, establishing new connections when necessary.

Over a series of studies, Wellman has detected the existence of personal communities in advanced societies: individual social networks that exist both offline and on. In this case, communities are substituted by personal networks, wherein ties are specialized and diversified, as people create their own “personal portfolios” (WELLMAN, GULIA, 1999).

In recent studies, Nardi and colleagues (2002) document the rise of personal networks in the workplace. They call these *intensional networks* and claim that the most important unit of analysis for computer supported cooperative work is not at the group level, but at the individual level in this type of environment. She also states that relationships outside the organization (e.g., government and press) are critical to many businesses, and that many corporations now operate in an increasingly distributed way, where workers, contractors, consultants and other contacts in different locations. Within organizations, constant reorganizations mean workers’ responsibilities, colleagues and reporting relationships change frequently.

Intensional networks are personal social networks workers draw from to get work done. Intensional networks are both emergent (in that they can be called into existence to accomplish particular work) and historic (in that work is based on known relationships and shared experiences, and each joint project adds to the relationship history, altering it). Different sections of the networks become active (or “live”) depending on ongoing work. When a set of contacts is live, the relations that keep it going need constant renewal through acts of communication (NARDI, 2002, p.221). The live substructure is a result of human interaction, happening due to ongoing group work. Intensional networks pulsate as activity ebbs and flows, and different parts of the network come to life. Working in intensional networks introduces additional complications for the participants, in remembering other participants, roles and responsibilities (NARDI, 2002, p.213). Live subnets are the parts of a network that keep a worker most occupied, but dormant contacts also need to be tended and remembered for future joint work.

Creating and maintaining relationships which can be relied upon for work purposes requires deliberate and careful work from network owners (NARDI, ENGESTROM, 1999; SCHWARZ et al., 1999). Work groups are put together through the assemblage of people found through intensional networks rather than being teams created through organizational planning and structuring. In turn, much of the work is undertaken via interpersonal communication tools, such as email, messenger or the phone, to coordinate, manage interdependencies and keep abreast of ongoing work.

Other similar forms of association have been described in recent work: *Coalitions* are groups of people working on a same problem from different ends, usually unaware of each other (ZAGER, 2000). Coalitions are loosely bound, temporary and fluid. Individuals are bound by shared interests, and usually do not communicate with each other, being part of different organizational substructures. As individuals are usually not in contact, work is sometimes replicated. *Knots*, on the other hand, are small groups put together by the organization with an explicit purpose (ENGESTROM, 2005). In knots, combinations of people, resources and tasks are unique and the group isn't long lived. Knots are situation driven, in that individuals come together to solve a specific situation, and usually draws heavily on role assignment, designed by the encompassing organization.

Communities of Practice are groups of people with shared interests that come together to exchange information and share experiences (Wenger, 1998). These groups form spontaneously within organizations, and congregate practitioners who must often face similar problems. Communities of practice are central to learning and knowledge dissemination within organizations, as they provide a favorable environment for exchanging lessons learned. Structurally, there is usually a set of core members, who dedicate a significant amount of time to the community, a larger set of active members, who participate actively, and a set of peripheral, members, who have engage very little, but sit in as observers and learn as a result. Participation is fluid as individuals become more or less involved as time and interests dictate.

Zacklad (2003) proposes the study of individuals in structurally open situations, where it is possible to change the social structures that tie individuals and the technical configurations within which they work. In these so-called '*Communities of Action*', collective activity is oriented towards goals but the means of designing and attaining these goals are neither completely formalized nor standardized. Actors have a significant amount of autonomy and are free to define their modalities of coordination and to adapt themselves to emergent situations.

In communities of action, coordination is achieved through intellectual transactions, or linguistic communications in which actors exchange knowledge and establish commitments. The most basic level of coordination is achieved through mutual perception, in which members coordinate by collecting and exchanging information about the activities of partners and their fields of operation (ZACKLAD, 2003).

A large number of organizational theories describe groups where actors have little opportunity to transform the organizational framework, change the group's structure, alter member roles or functional characteristics of available tools (ZACKLAD, 2003). The aforementioned theories emphasize looser situations, where individuals have freedom to alter configurations and improvise as needed. The common unifying thread is the possibility to negotiate change with other team members. Communities of action are the focus of our research: groups where individuals have the freedom to reorganize to adjust to changes, and coordinate by mutual perception, gathering and processing information about others' activities. Our main research question is whether it is possible to automatically provide such information and, if so, how to discover it.

Distance and Collaboration

Even in tightly knit teams, individuals still perform the majority of activities individually (ANDRIESSEN, 2003). When individuals collaborate, they often shift back and forth between individual and shared work, and between loosely and tightly coupled collaboration (GUTWIN, GREENBERG, 2002). This is especially true when there is low interdependence between them. In these cases it is beneficial to add collaboration support to individual work tools, enabling the shift into close collaboration when necessary (GUTWIN et al, 2005). Awareness of current and past efforts becomes necessary, since one individual might work on a shared artifact for a while and another may pick it up later (EDWARDS, MYNATT, 1997).

Globalization has created a trend towards geographically dispersed (virtual) teams, where individuals' work is mediated by computer technology (HERTEL, 2005). The main reasons why virtual teams arise are: expected reduction of costs, increase in flexibility, easily accessed technologies, which also facilitate information capture and dispersal and adoption of organizational networks (ANDRIESSEN, 2003). Distributed work, however, hampers communication. Physical proximity minimizes a number of problems that arise while collaborating, such as meeting partners, defining problems, planning, supervising co-workers and subordinates, etc. Studies have shown that ease of contact plays an important role in the partner choice, and distance creates a barrier between individuals, as detected by Kraut and colleagues (1990a), who showed that frequency of communication and collaboration decrease with distance.

In fact, collocated synchronous interactions have a few characteristics that are hard to reproduce in computational environments, such as rapid feedback, multiple channels, nuanced information, shared local context, informal "hall" time, implicit cues (OLSON, OLSON, 2000), and argue that most of these characteristics are poorly supported by today's technologies. The list is shown in Table II-1.

Table II-1: Characteristics of collocated synchronous interactions and their support by today's technologies (adapted from Olson and Olson, 2000).

Characteristic	Description	Implication	Supported by today's technologies?
Rapid feedback	As interactions flow, feedback is as rapid as it can be	Quick corrections possible when there are noticed misunderstandings or disagreements	Not supported
Multiple channels	Information among participants flows in many channels—voice, facial expressions, gesture, body posture, and so on	There are many ways to convey a subtle or complex message; also provides redundancy	Poorly supported
Personal information	The identity of contributors to conversation is usually known	The characteristics of the source can be taken into account	Poorly supported
Nuanced information	The kind of information that flows is often analog or continuous, with many subtle dimensions (e.g., gestures)	Very small differences in meaning can be conveyed; information can easily be modulated	Poorly supported
Shared local context	Participants have a similar situation (time of day, local events)	A shared frame on the activities; allows for easy socializing as well as mutual understanding about what is on each others' minds	Not supported
Informal "hall" time before and after	Impromptu interactions take place among subsets of participants on arrival and departure	Opportunistic information exchanges take place, and important social bonding occurs	Poorly supported
Coreference	Ease of establishing joint reference to objects	Gaze and gesture can easily identify the referent of deictic terms	Not supported
Individual control	Each participant can freely choose what to attend to and change the focus of attention easily	Rich, flexible monitoring of how all of the participants are reacting to whatever is going on	Not supported
Implicit cues	A variety of cues as to what is going on are available in the periphery	Natural operations of human attention provide access to important contextual information	Not supported
Spatiality of reference	People and work objects are located in space	Both people and ideas can be referred to spatially; "air boards"	Not supported

A looser structure and distance may lead to a decrease in involvement and interaction. As a consequence, individuals miss opportunities for collaboration, and sometimes end up performing overlapping tasks or duplicating work because they are unaware of each other's activities. Distributed groups display a tendency to change organizational structure and task distribution in order to reduce coupling of non-collocated members, thereby reducing coordination needs (OLSON, TEASLEY, 1996).

This leads to fragmentation, as members of different teams may not be up-to-date of the latest evolution in other teams' designs, and the group as a whole becomes fragmented.

Awareness in Collaborative Work

Situation awareness research focuses on each individual's capacity to perceive elements and the cognitive processes involved in maintaining awareness of the environment. Endsley (2000) defines situation awareness (SA) as the process of perceiving environmental cues, interpreting their meaning and projecting their status in the near future. This information is used as a basis for individual decision making while working. A number of factor influence this perception, both personal (e.g., abilities or training) and inherent to task or system (.eg., complexity or interface design). Situation awareness influences decision making, which in turn influences the performance of actions that affect the environment (these effects will also be perceived and acted upon, restarting the cycle). Endlsey's SA model is shown in Figure II-2.

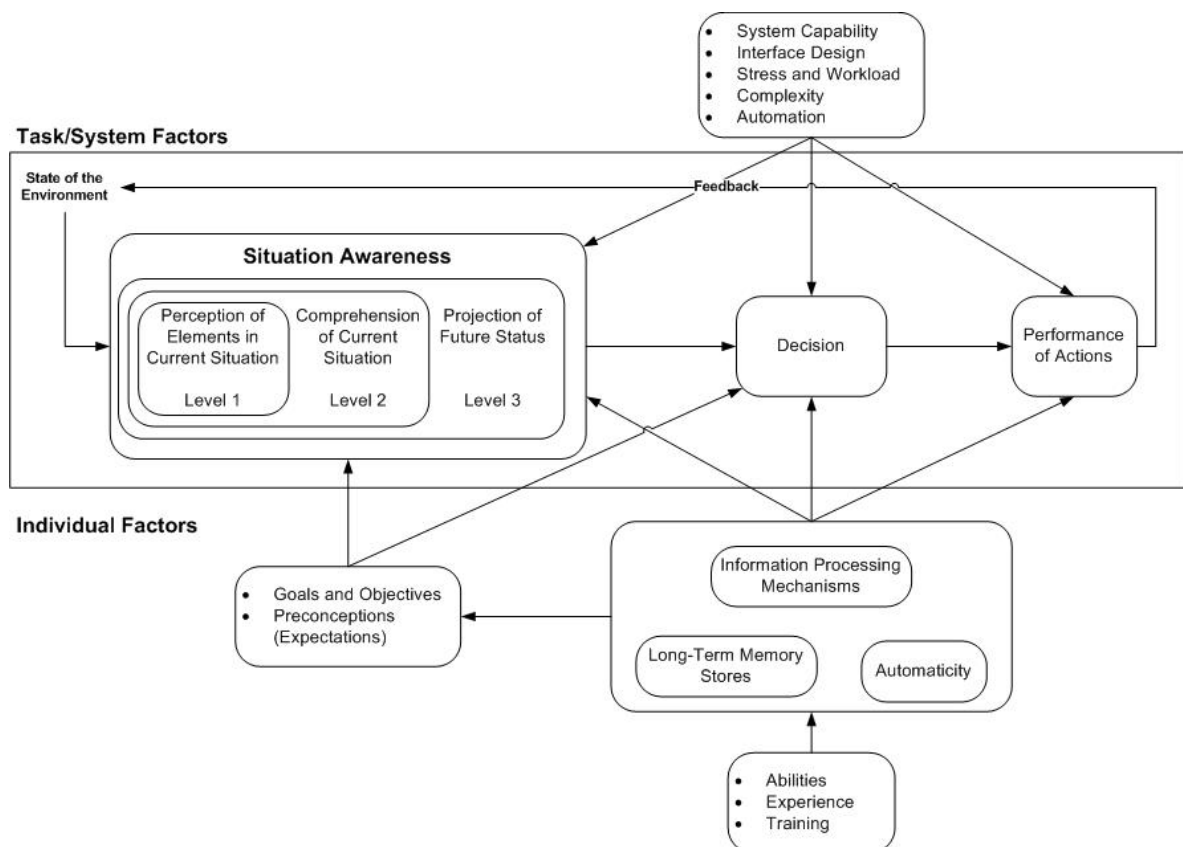


Figure II-2: Functioning of situation awareness (from Endsley, 2000)

Dourish and Bellotti (1992) define awareness as an understanding of activities of others, which provides a context for one's own activity. This information ensures that individual contributions are relevant to the group's activities, and enables individuals to

assess others' actions with respect to group goals and work progress, which in turn allows individuals to adjust their behavior or take action according to the situation. Elaborating further on those concepts, Gutwin and Greenberg (2002) define workspace awareness as up to the moment understanding of another person's interaction with the shared workspace. It is knowledge about the group's working environment, which creates an understanding of people within a workspace. These two definitions specialize the SA definition by defining the environment as a collaborative workspace, where information about other's activities and status is an important asset.

Rodden's (1996) model of awareness considers a set of objects in space, which interact based on their levels of awareness. Awareness is manipulated via focus and nimbus, subspaces within which an object directs its presence or attention. Awareness is the overlap between nimbus and focus, where:

- Nimbus is the information given out by each element in space that can be perceived by others; and
- Focus describes the elements at which a user directs his or her attention.

Thus, the awareness of individual A towards individual B is the intersection between the information being given out by B (B's nimbus) and the information A is interested in (A's focus), as shown in Figure II-3. This model details how information to maintain awareness is obtained and shows that attention is a key aspect in the process, as awareness is affected by each individual's focus and the information provided by each element (nimbus).

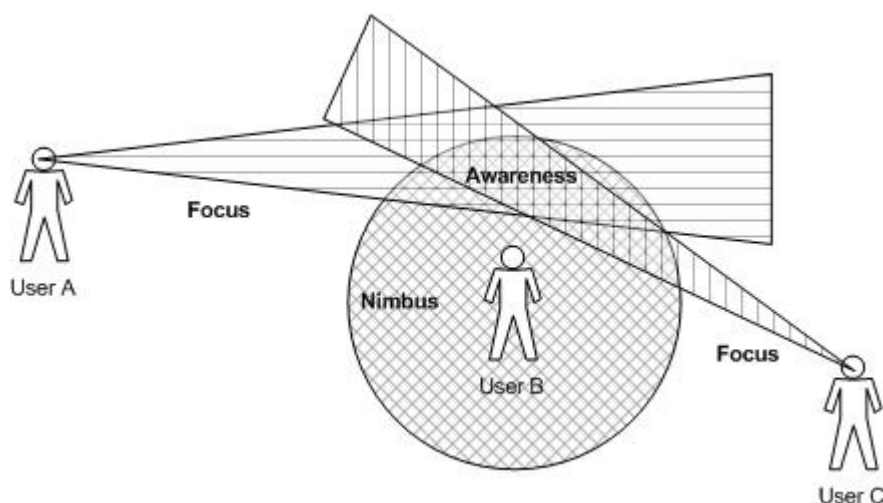


Figure II-3. Awareness as the overlap between focus and nimbus

In a collocated environment, individuals give out a large amount of information while working, which can be picked up by anyone paying attention to it. In computational settings, users give out information via the applications they interact with and the operating system, which is normally not relayed to others.

In a collaborative environment, awareness involves knowledge about the people one is collaborating with (presence, identity and authorship), the activities they are working on (actions, intentions and artifacts manipulated) and where (location of work, gaze direction, view and individual reach). Historical awareness information also includes action, artifact and event history, and should be provided in asynchronous work situations (Gutwin and Greenberg, 2002). This framework provides a point of reference for designers to consider awareness elements and what information to provide in given situations (based on which questions they should answer), and is summarized in Table II-2.

Table II-2: Workspace awareness elements and the questions they address (adapted from Gutwin, Greenberg, 2002)

		<i>Present</i>	<i>Past</i>
Category	Element	Specific Questions	Specific Questions
Who	Presence	Is anyone in the workspace?	Who was here and when?
	Identity	Who is participating? Who is that?	
	Authorship	Who is doing that?	
What	Action	What are they doing?	What has a person been doing?
	Intention	What goal is that action part of?	
	Artifact	What object are they working on?	
Where	Location	Where are they working?	Where has a person been?
	Gaze	Where are they looking?	
	View	Where can they see?	
	Reach	Where can they reach?	
How	Action History		How did that operation happen?
	Artifact History		How did this artifact come to be in this state?
When	Event History		When did that event happen?

In later work, Tam and Greenberg (2006) expanded upon the concepts involved in asynchronous change awareness. They develop a view of each awareness element according to three different perspectives (artifact, person and workspace), which

correspond to the three ways a user may need to view the shared workspace at any given point in time. These are summarized in Table II-3. When multitasking, this information becomes more important, as an individual must be able to quickly switch back and forth between working spheres, and needs to be up-to-date with the current status of the collaborative project to be able to integrate with other participants.

Table II-3: Asynchronous change awareness elements according to different perspectives, and questions that should be addressed (adapted from Tam and Greenberg, 2006)

Category	Information Elements	Specific Questions		
		Artifact-based View	Person-based view	Workspace-based view
Where	Location history	Where was this artifact (when I left)? Where is the artifact now?	Where in the workspace has a person visited?	Where have people been in the workspace? Where were artifacts in the workspace?
	Gaze history		Where in the workspace has a person looked at?	Which parts of the workspace have people looked at?
	Edit history	Where has this artifact been during the time that I have been away?	Where in the workspace has a person made changes?	Which parts of the workspace have people made changes in?
Who	Presence history	Who has looked at this artifact?	Who has this person interacted with?	Who has been in the workspace?
	Identity	Who has changed this artifact?		
	Readership history		Who made changes with this person?	Who has looked at the workspace?
	Authorship history			Who has made changes to the workspace?
What	Action history	What changes have been made to the artifact?	What artifacts has a person looked at? What artifacts has a person changed? What activities has a person engaged in?	What changes have occurred in the workspace? What artifacts were viewed? What artifacts were changed?
How	Process history Outcome history	How has this artifact changed?	How has a person changed things?	How has the workspace changed?
When	Event history	When was this artifact changed? When was a particular change to this artifact made? In what order were changes made to this artifact?	When did a person make changes? When did a person make a particular change? In what order did this person make changes?	When were changes made to the workspace? When did a particular change in the workspace occur? In what order did changes to the workspace occur?
Why	Cognitive history Motivational history	Why was this artifact changed?	Why did a person make that change?	Why was that change made in the workspace?

Parallel research introduces the notion of context in collaborative environments. Context is a complex description of knowledge on several levels (e.g., physical, social, historical, etc.) within which action and events occur (Borges et al., 2004). Contextual knowledge is used when individuals execute tasks, as they reason about the surrounding context to decide on courses of action. In group work, context may be seen as information falling into one of 5 categories: people and groups, scheduled tasks, relationships between people and tasks, environments where the interaction takes place and tasks and activities already concluded (Borges et al., 2004).

Kraut and colleagues (1990b) define *unplanned interactions* as interactions that occur without prior planning by either interactant. These interactions occur during semi-spontaneous encounters (e.g., hallway encounters or coffee breaks), because there is some common context between the parties or a possibility for information exchange. Collaboration is based on interactions between people. Kraut (1990b) defines four interaction categories, described below:

- Planned interactions are scheduled, formal reunions.
- Intentional interactions happen when a person explicitly searches for the others.
- Opportunistic interactions happen when one party means to talk to someone, but they meet by chance.
- Spontaneous interactions are not anticipated by any of the parties and happen completely by chance.

Intentional, opportunistic and spontaneous interactions are considered unplanned, but happen on a daily basis when a group shares the same physical workspace. These interactions happen more frequently than planned ones, and contribute to increase the awareness of what is happening in the workspace and of what others are working on (KRAUT et al, 1990a; KRAUT et al, 1990b). This knowledge generates a greater understanding of the context in which people are inserted, increasing chances of new collaborative projects between peers (MCEWAN, GREENBERG, 2005). However, if the group does not share the same workspace, these informal interactions stop happening and the chances of new, spontaneous collaboration happening also drop.

In collocated environments, information that helps maintain awareness is gathered mainly through (Gutwin and Greenberg, 2004):

1. intentional communication, i.e., communication intended by the sender, such as conversation and gestures;
2. consequential communication, or information transfer that happens as a consequence of the individual's activity within the environment, obtained by observing others' actions or body positions; and
3. feedthrough, which is the mechanism of determining a person's actions through cues given by the artifacts they interact with in the environment, such as position, orientation or movement).

In distributed collaboration, intentional communication is frequently the only way to maintain awareness of the environment, which creates a significant overhead, as individuals must remember to inform others of important events and send messages to that effect. As it involves knowledge about a dynamic environment, awareness must be maintained as the environment changes. It is maintained through perceptual information gathered from the environment (through interaction with the environment); and it is generally secondary to other goals. While staying aware of others is taken for granted in everyday interactions, it becomes hard in distributed systems, where communication and interaction resources are poor (GUTWIN, GREENBERG, 2002).

Ethnographic studies have determined that awareness allows group members to manage the process of working together, and is necessary for coordination of group activities (Dourish and Bellotti, 1992). Being aware of others' activities in a workspace allows participants to better understand the boundaries of their actions, which in turn helps them fit their own actions into the collaborative activity stream. This also enables groups to better manage coupling levels between their activities, helping individuals decide who they need to work with and when to make the transitions from looser to tighter coupling (Heath and Luff, 1991), and deciding when to establish contact (GREENBERG, JOHNSON, 1997). It is this awareness of the working space that will enable individuals in loosely structured situations to adjust their course of action to compensate for unforeseen events.

Furthermore, awareness simplifies communication by allowing individuals to reference the shared environment and elements within it: when discussing shared

artifacts, the workspace can be used as a communication prop (Brinck and Gomez, 1992). This makes awareness an important building block for the construction of team cognition (Gutwin and Greenberg, 2004), as it helps create and maintain shared context. It is also an enabler of shared understanding that allows individuals to get a better sense of the work that is being performed by others (Gutwin et al, 2005), besides contributing to group integration (NARINE et al, 1997).

User Attention and Focus

Rodden's (1996) definition of awareness refers to another important concept, that of attention. One of the hardest issues individuals face nowadays is handling the increasing amount of information they are exposed to on a daily basis. Deciding what should receive one's attention has become a frequent task in information workers' environments, as information consumes attention, and the increase in the available information creates a demand for attention that is hard to meet (VERTEGAAL, 2003). In recent years, researchers have acknowledged that attention is in short supply among organizations and individuals.

Every email sent is a demand for attention, which must be addressed by the recipient. In this scenario, each individual functions as an information provider, trying to attract attention to his or her memos, e-mails, projects, etc (DAVENPORT, BECK, 2001). Controlling one's attention thus becomes an important part of everyday life, and a determinant of productivity, as information reaches consciousness only when it is attended to (Csikszentmihalyi, 1997, apud Davenport and Beck, 2001).

Attention is paramount for decision making, as individuals need to attend to issues and problems to make good decisions. Attention is focused mental engagement on a particular item of information (DAVENPORT, BECK, 2001). It occurs between a narrowing phase, where much of the sensory input is discarded, and a decision phase, in which the individual decides to act on the information attended to. There is a causal relationship between awareness, attention and action, as shown in Figure II-4. Attention is a selective, cognitive process, through which individuals absorb selected information.

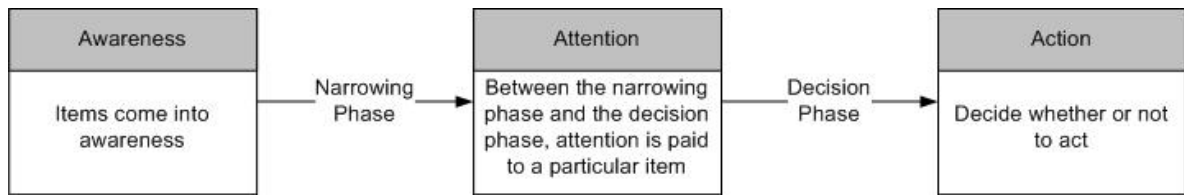


Figure II-4: Relationship between awareness, attention and action (from Davenport and Beck, 2001, p.21)

Appropriate allocation of attention is fundamental to the individual, and a person will usually choose among the many attention demands according to the as expected payoffs (e.g., learning something new, fixing something that is broken or reinforcing a relationship). This attention, once spent, does not come back, so it is reasonable to assume that individuals will prioritize their allocation of attention according to their goals and to the contracts they have established with others. In loosely structured configurations, workers often multitask between several collaborations, dividing their time and attention according to the urgency of the tasks at hand (GONZALÉS, MARK, 2005)

The two main problems encountered when working in loose, reconfigurable groups are *remembering* people and activities and *communicating* with others (NARDI, 2002). A recurring problem is managing attention and remembering which arrangements have been made and what tasks are due in order “not to let anything fall through the cracks” (GONZALÉS, MARK, 2005). A *working sphere* groups together a set of events involving a particular group of people, a common motive and a unique set of resources (GONZALÉS, MARK, 2004).

Support systems for workers engaged in multiple virtual teams must help individuals manage their multiple working spheres. One way to do this is by helping them understand the social context of the work they are performing: who is involved with this project and in what capacity? What other resources are relevant? The determination of collaborators and identification of collaboration themes is a step towards that goal. A collaboration sphere is composed not only of a set of people, resources and tools but also of the interrelations between them, which affects ongoing work by defining how people relate, organize themselves and get work done. The group’s context, informally constructed and dynamic, contains hierarchies, roles and task assignments. Discovery of these contexts is a first step towards the creation of systems to support workers enmeshed in them.

Information workers are typically involved in multiple collaborations at the same time, and must divide their time accordingly (GONZALÉS, MARK, 2005). This means there may be several different subsections of one's intensional network active simultaneously, and the user needs to shift attention between the different working spheres. One of the difficulties in this scenario is remembering the different contexts and managing transitions between them. Workers use a variety of methods to remember things, people and the current context, so they can pick up where they left off. Practices include dropping in on collaborators, constantly checking email, keeping to-do lists or writing notes to oneself (GONZALÉS, MARK, 2005).

Attention goes to people and projects in which a user has the most interest (or higher stakes). Interaction analysis elicits the user's priorities and enables a system to maintain an updated collaboration profile. The allocation of attention between acquaintances provides a user's priorities regarding ongoing collaboration. If each email received is a demand for attention, each email responded to is a clear action on the information contained therein. This send-receive dynamics lasts as long as there is a joint goal to be reached or a common focus of attention. Continued interactions show joint commitment to a collaborative project.

Thus, an analysis of interactions should enable the discovery of how attention is allocated between a user's many acquaintances. Prior research has indicated that a long term historical analysis could elicit patterns that mirror a user's changing interests (PERER et al, 2005). The same study indicated an intensification in interactions during collaborative work.

The Locales Framework

The Locales Framework (FITZPATRICK, 1998) provides a set of abstractions to support the design and analysis of collaborative work. It seeks to bridge the space between the social and technical perspectives, and was conceived as a tool for analysis and design of collaborative work support systems. It is based on Anselm Strauss' Theory of Action (STRAUSS, 1993 apud FITZPATRICK et al., 1995), according to which individual action is always embedded in interactions, which are carried out by one or more interactants. It is based primarily on the notion of continually evolving action and of *Social Worlds*. According to the Theory of Action, Action takes place in

the context of structural conditions, which facilitate or hinder the course of interactions, and which may change with time.

Social Worlds are groups of people who share a commitment to collective action, and form the prime structuring mechanism for interaction (STRAUSS, 1993 apud FITZPATRICK et al., 1995). A social world is an interactive unit that comes to life when a number of individuals decide to act collectively (which often require coordination of separate perspectives and sharing of resources). These individuals committed to collective action work together via sites and employ means to do so. Sites are the spaces (e.g. shared file systems) and means are objects contained in these spaces (e.g. the files and documents stored in this file system). The system's role in many cases is only to provide a setting, but not to structure collaboration.

This is in consonance with Suchman's (1996) observations of how collocated workers use their available environment as a backdrop for their work, and reorganize themselves and the space according to the needs. Her study of an operations room crew while dealing with an aerial incident showed how individuals moved into closer collaboration when necessary, physically moving to different locations or turning around to change focus. Participants became engaged in solving a joint problem through their partial awareness of the shared situation, and joined the group as needed. While the incident occurred in a short period of time, it shows how, through manipulation of the shared space and mobilization of resources, the group was able to move from individual activities to a concerted course of action. The shared workspace was collaboratively constructed by the members of the group.

A *Locale* is an abstract concept that arises from the use of space and resources by a group. It maps the relationship between a Social World (and its interaction needs) and the *sites* and *means* its members use to meet those needs. (FITZPATRICK et al., 1998).

Given that individuals are usually involved in multiple social worlds at a time, different social worlds will be interconnected through the individuals that compose them, and actions in one social world may reflect in another. Further, each individual is typically engaged in multiple tasks or activities, spanning more than one social world. This is in accordance with observations by Gonzalés and Mark (2005), who describe the processes involved in multitasking between several collaborations.

An automated analysis of the locales used by a group should yield patterns that reflect the group's work. Which social worlds are active, what the prioritization is between them and how they connect to each other.

Social Network and Interaction Analysis

Interpersonal associations are the elements through which individuals form groups and societies (VANDENBERGHE, 2001). George Simmel states that interactions between individuals are the most basic element of groups, serving simultaneously as the means of association and of socialization between members (ibid, p.50), and should therefore be taken as the objects of study.

Simmel was more concerned with the forms rather than the content of the interaction, directing his work towards a science of the structures of social relations (MERCKLÉ, 2004). It is possible to study social relations because they are regular and stable, and they figure in different times and domains of social life: typical relationships, such as domination, competition, imitation or conflict can be found by studying these structures. Simmel's work provided early inspiration for the development of Social Network Analysis (SCOTT, 1991; VANDENBERGHE, 2001).

J.L. Moreno developed sociometry as a way to map ties between individuals and establish their strength. It is an instrument to study social structures in light of attractions and repulsions that manifest within a group (MORENO 1954, apud MERCKLÉ, 2004). The resulting sociogram is, at the same time, a means of exposure and exploration of the group, which allows the determination of the position each member occupies or the relations of acceptance or rejection established between individuals.

Social network analysis (SNA) takes stock of the relationships between individuals and their regularities to analyze individual and group behavior, thus bringing into individual behavior the complexity or the social systems within which they operate (MERCKLÉ, 2004). This reflects that fact that most activities are inserted in social contexts, and that the network within which they are inserted is bound by rules/practices defined by its members.

Social network analysis is used widely in the social and behavioral sciences, as well as economics. It concerns the study of social entities and their relationships:

communication among individuals, trade between businesses or treaties between nations. The basic elements of a network are nodes and the ties that bind them. In a social network context, nodes are individual actors, and the ties are relationships between these individuals, which may be hierarchical, familial, friendship, work, etc. (SCOTT, 1991). It considers structures such as the sociogram, a graph that represents individuals and the relations between them (WASSERMAN, FAUST, 1994). These relations can be of diverse nature (communication, party attendance, information diffusion, etc.), and are usually expressed as graphs and matrixes (see Figure II-5); upon which network analysis can be performed (SCOTT, 1991). SNA uses graph theory and linear algebra techniques to determine structural properties of individuals within each network. Thus, network attributes such as centrality (how many ties link to a certain node (FREEMAN, 1979)) are mapped to a sociological interpretation (importance in the group). Social network analysts look at the world in terms of patterns or regularities in relationships between actors.

As the actor participates in different groups (GONZALÉS, MARK, 2005; SIMMEL, 1955), he or she constructs a personal network around him/herself, which is never exactly the same as anyone else's (WELLMAN, GULIA, 1999). Ties that bind pairs of actors are constructed individually, as relationships progress (GABARRO, 1990). To establish new ties to others, actors commonly use a part of the network that surrounds him/her (FERRAND, 1997, apud MERCKLÉ, 2004).

-	A	B	C	D	E	F	G	H	I
A	-	1	0	1	0	0	0	0	0
B	1	-	0	1	0	0	0	0	0
C	0	0	-	1	0	0	0	0	0
D	0	1	0	-	0	0	1	1	1
E	0	0	0	0	-	0	0	0	1
F	0	0	1	1	0	-	1	0	0
G	0	0	0	0	0	1	-	0	0
H	0	0	0	1	0	0	0	-	0
I	0	0	0	0	0	0	0	0	-

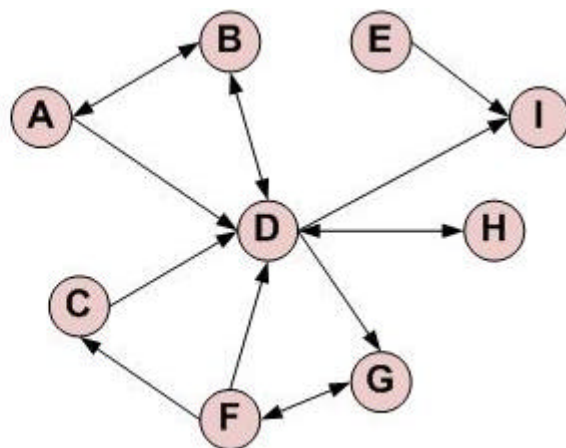


Figure II-5: Sample sociogram and corresponding table of links between elements.

An actor's network is subject to his or her interpretation and understanding of it, which will affect the ways in which he or she uses/navigates the network. The extent of a network is variable, with some researchers saying that all of society is enmeshed in a set of communicating individuals and groups (LÉVI-STRAUSS, 1952, apud

MERCKLE, 2004). Research has shown the importance of personal networks, and how strong and weak ties play an important role in an individual's life, as they use it to find jobs and form new associations (GRANOVETTER, 1973). Later studies departed from small group studies and documented the existence and importance of these networks in large scale situations, such as Silicon Valley (SAXENIAN, 1994; CASTILLA et al., 2000), where members meet informally, exchange information, discuss ideas and change companies frequently, creating an innovative environment.

Two types of analysis are possible: while *sociocentric* network analysis looks at relationship structures from a global perspective (e.g., a graph of the communication between all members of a department or group), *egocentric* network analysis, on the other hand, focuses on the individual (*ego*), and analyzes his or her relations to a set of others (*alters*) (GARTON et al., 1999). This type of network has been used to study the social environment surrounding individuals or families, or social support structures (WASSERMAN, FAUST, 1994).

In distributed groups, individuals usually interact electronically, and electronic interactions leave traces, such as email, fora or messenger logs. These interactions display certain rhythms that correspond to an individual's work patterns (PERER et al., 2005), and can be used, for instance, to study the evolution of an individual's interests (as he or she relates to others in different domains) or collaborative endeavors: for example, intense message exchange usually accompanies cooperative work. Individual patterns of email exchange can also indicate hierarchy and positioning in a group (FISHER, 2004).

In most cases, social network and interaction analysis have been used as analysis tools only. In this thesis, we seek to apply social network and interaction analysis in an active way, to enable reasoning by an intelligent entity and system adaptation. We construct an egocentric network based on records of electronic communication. This network is then searched to discover ongoing collaboration and information related to joint projects.

Information Retrieval

In general, information retrieval techniques represent documents as keyword vectors, calculated based on their contents (BAEZA-YATES, RIBEIRO-NETO, 1999). These keywords are extracted from document bodies and weighed according to their

frequency in each document. This extraction process can be refined through stopword removal (removal of those words that have high frequency but add little to the documents, such as “the”, “a”, “them”, etc.) and stemming (a process through which suffixes are removed, leaving only the stems to be processed as keywords). Each word’s importance (in defining the documents contents) is calculated according to its frequency in the document and its frequency in all documents, using the TFiDF (Term Frequency, inverse Document Frequency) formula (SALTON, MCGILL, 1997): a word’s relevance for a particular document equals its frequency in the document divided by its frequency in the document set. Thus, words that appear very frequently in the document set lose their distinguishing power, and documents are associated to word vectors containing the keywords that best represent it.

Matching can be performed by comparing the keyword vectors representing the different documents to calculate their similarity. The cosine measure is a frequently used one, where the cosine of the angle between the two vectors is calculated and, the closer the vectors, the more similar the documents (BAEZA-YATES, RIBEIRO-NETO, 1999).

Collaboration Spheres as Bridges between Group Members

To enable self-coordination and group work in loosely structured settings, information about ongoing work must be made available to participants. This information concerns events relating to people and resources involved in each work setting.

Given that intensional networks and working spheres represent individual points of view towards collaborations, we depart from sociocentric approaches and instead focus on the individual and his or her relations to others. We adopt an egocentric approach to assist with management of multiple tasks and awareness maintenance. While an individual works alone, on a given activity belonging to a certain working sphere, other working spheres may be incurring action from other members. This information should be detected and disseminated to enable easier transition to and from working spheres. Based on this sociological point of view, we adopt a peer-to-peer architecture to enable maximum flexibility (individuals may link to others as desired) and we use agent technology to create a proactive system that keeps track of activities and collaborations.

In loosely structured settings such as the ones described, individuals often work without centralized support, using email and other electronic communication media to coordinate and exchange information and tasks (NARDI et al., 2002). Different groups may select different tools or have different resources available to them (GONZALÉS, MARK, 2005). Determining collaborators, resources and information exchanged, in a distributed situation where a variety of tools are used, is a complex yet necessary task to facilitate collaboration.

Since subnets become active through communication (NARDI et al, 2002) and work is often undertaken through linguistic exchanges (ZACKLAD, 2003), we propose a method to for the determination of collaborators (active working spheres) and of activities that belong to different working spheres based on analysis of interaction records and content. With this, we expect to be able to determine active collaborator groups in a network and what their activity concerns. These can then be matched to individual activity to discover which activities belong to which working spheres, and keep track of the evolution of the joint workspace.

To create a unified view of the group, bridging the gap between individual and joint work, and tie multiple working spheres together, we introduce the concept of collaboration spheres. These group together people and resources, along with contextual elements related to the joint work. Events occurring within a collaboration sphere are pertinent to members in this sphere and can be relayed to other members. In the next chapters, we further elaborate on and present a method for constituting collaboration spheres.

III. Related Work

This section presents a review of systems that deal with information organization and dissemination. This includes recommendation systems, awareness support and social networking systems, as well as related work on personalization.

1.1 Work Organization

Many social network based systems have the goal of finding experts. ReferralWeb (KAUTZ et al, 1997) was a prototype to locate expertise within a company, based on references provided by users about who they might call regarding certain topics. Agents follow these recommendation chains to find users who can help with given problems. The system also processed users' files, extracting keywords from messages exchanged with others, to build the user's view of each of his or her contacts. McArthur and Bruza (2003) present a similar method to construct networks of people and keywords from email data. Groth (2003), used networks to locate individuals with a certain expertise and availability through an analysis of their activities and tasks. In the aforementioned approaches, the emphasis is on finding experts, and navigating the social network to discover who knows what.

ContactMap (NARDI et al, 2002) is a system that enables users to visualize and organize their contact networks. It is an attempt at supporting work through intensional networks, and was built to help people remember others in their networks, tasks they were involved in, the connections between different acquaintances, documents exchanges and one's own work status and deadlines. It builds networks through email analysis, constructing a personal social network, from which the user can choose which contacts to display. The user can then lay out contacts, grouping them as appropriate. It is an organizational tool that helps users remember certain contacts and group them according to their interrelations. It functions as a personal contact manager, which enables individuals to easily maintain an updated view of their personal social networks, based on email records.

Vizster (HEER, BOYD, 2005) enables users to visualize and access information on their personal social networks. The visualization contains a wealth of information about user's interests extracted from Social networking site Friendster (www.friendster.com). This system is especially useful for the exploration of one's

social network, grouping members and finding out about interests and activities, which enables one to tap into the network as necessary.

Working with ties involves more than knowing who is in the network and who can do what. It involves remembering ongoing projects and associated outstanding arrangements, contracts and roles or task division and interdependency. Community Bar (MCEWAN, GREENBERG, 2005) was designed following the Locales Framework (FITZPATRICK, 1998), and is a configurable interface that enables users to keep abreast of group members' activities through media items (e.g., webcams, sticky notes, chat windows) providing identity, presence and some activity information. The system allows users to create links to others by selecting the different media items through which they will receive information. Users will be informed of any information that goes through that particular channel. Contacts and media items can be grouped to represent different working groups, and focus and nimbus can be established manually (the user can determine how much he or she wants to receive through each particular channel, and how much he or she lets other others see through this channel). It provides an easy establishment of interaction, as chats can be started with any member.

Activity Explorer (MULLER et al, 2004; GEYER et al, 2006) was conceived to fill the gap between formal, structured collaborative applications and informal, easy-to-initiate collaborative tools. It supports lightweight, informally structured opportunistic activities relating to different shared objects with dynamic membership by controlling shared objects. It is based on the notion of shared objects, and enables sharing of resources in a fine grained way. Shared objects contain persistent information and define a list of people who have access to this content (e.g., files, messages, chat logs, etc.). Each object supports real-time notification of the events that happen to it to the users involved with it. This provides fine grained awareness of who is working on what, which may serve as a trigger for opportunistic collaboration. As collaboration proceeds, users manually combine and aggregate heterogeneous shared objects into structured collections. The Activity Explorer (AE) provides an integrated view of ongoing activities and who is related to them. Interviews with users showed that the system was useful as a coordination tool and to get a better sense of what others were doing. User indicated that they substituted email and messenger for AE, as it integrated additional features and also enabled sending email and engaging in chat.

1.2 Recommendation Systems

Recommendation systems assist and support social processes through information filtering algorithms. A typical recommendation system receives and aggregates recommendation data and directs results to the right people (Resnick, Varian, 1997). These results should help people make decisions, based on the information provided.

Maes (1994) proposes the use of agents to support cooperation. She describes a system that allows users to find subjects that might be of interest to another user by analyzing their profiles and looking for topics that may be of interest to both parties. Foner (1997) presents a multi-agent system to bring together people that share interests. To achieve this end, the system performs an analysis of the user's documents (e-mails, text files, papers, etc.), building keyword based user profiles and matching these against other user's profiles.

Two different approaches can be used for multi-agent tools and matchmaking. The first one is to have a central server in charge of the matchmaking, such as used in Webhound and Webdoggie (SHARDANAND, MAES, 1995) and HOMR, Ringo or Firefly (LASHKARI et al, 1994). These systems build profiles by observing their users' actions, and saving content based situation-action pairs for each user (e.g., saving emails about genetic algorithms into an AI folder.)

The other approach is to distribute the system: Kautz, Milewski and Selman (1995) describe a prototype to locate expertise within a company based on references provided by users about others who they might call regarding certain topics. Agents follow this chain of recommendations to find users that can help with a problem. In this system, users need to explicitly supply the system with the data it needs to work. For each user, the system keeps a list of contacts and a list of domains of expertise for these contacts. Each user's model is built using standard keyword-vector techniques on the user's files, while each contact's profile is built using the same technique on messages sent between the user and his or her contacts.

Payton and colleagues (1999) describe a system where user interests are inferred from access to information. These interests accumulate with repeated access and decay at varying rates. Users can then use a specialized interface, which displays users as

nodes in a social network to locate potential collaborators. Matches are made based on information access history.

1.3 Awareness Systems

Maintenance of awareness is facilitated by physical proximity, as it is possible to perceive a large amount of information simply by walking around the office, overhearing others or engaging in brief coffee break conversations (Kraut et al., 2002).

Providing awareness-related information in e-collaboration environments is far from simple, especially when participants are distant from one another. Input and output devices generate less information than a face-to-face situation and users' interactions with computational workspaces also generate less information than a physical environment (Gutwin and Greenberg, 2004).

Group support tools should address the decrease in awareness caused by distance between participants and by the move to computer mediated media. One way this can be done is by supplying information about what others are doing and identifying possible collaboration opportunities. A number of applications have been designed to provide awareness-enabling information in e-collaboration applications.

Early awareness work was heavily geared towards audiovisual interfaces and the use of video to support personal awareness and informal interactions: CRUISER (Root, 1988) was a virtual environment using audio and video channels to support informal social interaction and social browsing. The possibility of easily engaging peers generated a number of spontaneous interactions which would otherwise not have happened without the system. Through the system, users explored the virtual workplace, strengthening their relationships to co-workers.

Polyscope and Portholes (DOURISH, BLY, 1992) are systems that use a shared view of the workspace, to allow users to be more aware of their environment through the use of a shared media space. Polyscope captures live images of the workspace and presents them in a user's computer screen. These images show activity in public areas and offices. Portholes followed Polyscope, extending the shared media space across greater distance and connecting Xerox's PARC to EuroPARC.

Portholes (Dourish and Bly, 1992) was a media space designed to support shared awareness between remote sites. Portholes enabled informal visualization of others,

such as late night sightings, snippets of songs being played in remote locations or individuals working long hours. This information created a sense of awareness of the work group as a whole, strengthening the sense of community between individuals, who reported feeling connected to their remote co-workers. Negative reports included unreliability of images (due to technological shortcomings) and the fact that events took a while to happen: the slow information turnaround led to loss of motivation to use the systems. These early systems also helped people identify the right moment for starting a conversation by checking on their counterparts' availability, an important issue in non-collocated environments.

Most voice and video tools have a serious limitation: although many of these tools provide the users with enough information to lead to interaction, most of them don't offer efficient means to move from vision into interaction, which compromises the users' capability to take advantage of opportunities that may arise. Additionally, these tools have no profiling capability, indiscriminately distributing anything that comes through the channel.

Later research focused on providing information to users about others performing similar tasks at a given moment: Piazza (Isaacs et al, 1996), for example, provides awareness information about others who are working on similar tasks when using their computers, exposing opportunities for interaction or cooperation. It allows a user, while working on his assignments, to glance at what other users, who are working on similar tasks, are doing. Their measure of similarity is based on three dimensions: data being manipulated, when it is manipulated and by what application. In this manner, people using the same data through different applications are considered to be close, and are displayed to each other.

Other approaches focus on a single object of the collaborative work, such as a document that is being written by several people simultaneously. Quilt, PREP and GROVE (DOURISH, BELLOTTI, 1992) follow this approach: these systems function as regular text editors, but have additional communication and synchronization functionalities built in. These allow users to become aware, through the document itself, of what is being changed.

PIÑAS (MORAN et al, 2002) is a framework to provide support for potential and actual collaboration by grouping together people and resources, tailored specifically for collaborative writing on the Web. To that end, PIÑAS defines collaborative spaces

(PICO_S) to integrate and structure correlated services. The services allow users to join a PICO_S that is shared among other users and resources, become peripherally aware of other users and resources that belong to the same PICO_S, send messages to users and documents in this PICO_S and launch other applications as necessary. Users must login to a shared environment to have access to the services, although the architecture is decentralized, linking distributed users together through an architectural layer. Users must also join the relevant PICO_S and subscribing to environments they want to know about. The system keeps track of activities in these shared spaces and informs the user about them through a messenger-like interface.

Community Bar (MCEWAN, GREENBERG, 2005), described above, is a tool that displays visual images (photos or videos) and names of the people that one is working with and provides manually entered profiles with contact information. A user can stay aware of what his co-workers are doing and initiate interaction if needed.

The CSCW3 system (Gross, 1999) adds a host of collaborative functionalities to a web browser, enabling users to view who is currently browsing the same page or has done it previously, move into interaction (via chat) when desired and exchange information in different forms (bookmark or history lists). Users reported they liked seeing who else was on the same page and rarely closed the visualization.

The Nessie Environment provides information based on interest profiles. It provides information on other users who are logged on and on events that have happened since their last login that match their interest profiles. These profiles contain a list of events relevant for this user, and have to be manually specified by the user.

Elvin (FITZPATRICK et al, 2002) is a notification framework not originally designed to support collaboration. However, it has been applied to awareness and interaction support. Elvin is a distributor of events regarding observed objects. Event producers send information to Elvin, which is compared to subscription tables containing rules created by consumers. These rules determine whether or not the event will be propagated to the user in question. A user can subscribe to events happening to another user or to artifacts of their shared collaboration. However, should these preferences change, new configuration will be needed.

Among the client applications developed using Elvin, tickertape and CoffeeBiff enable users to let others know of their intentions of going to get a cup of coffee. The

number of people at the coffee break is shown to others, and that lets them know when more than a certain number of people are drinking coffee (which, according to the authors indicates a party is going on). Tickertape displays event information in a small scrolling window and lets users know when events they have subscribed to happen. Information is distributed according to users' explicit event subscription configurations, which indicate the producers they want to focus on and certain rules to filter message contents. A third application was developed to link CVS to Elvin, enabling notifications of file changes. Through user studies, authors determined that it provided support for informal interactions, making digital actions visible and enabling social and content based filtering of messages.

1.4 Multiagent Systems

Multi-Agent Systems (MAS) are a good way to model systems to support distributed systems and work in open-ended environments, such as the one being addressed. A MAS is a loosely-coupled network of problem solvers (agents) that work together to solve a given problem (WANG et al. 1999). Intelligent agents are entities that perceive their environment through sensors and act upon it based on some internal reasoning (RUSSEL, NORVIG, 1995). They are encapsulated computer systems situated in an environment and capable of flexible, autonomous action in that environment in order to meet its design objectives (Woolridge, 1997). Agent-oriented techniques have been applied to a range of telecommunication, commercial, and industrial applications, as developers and designers realize its potential (JENNINGS, 2001).

Agent oriented approaches are well suited for developing complex, distributed systems, composed of interrelated subsystems with variable organizational ties to each other (JENNINGS, 2001) Adopting an agent based view of a system enables a more accurate representation of the decentralized nature of a problem, its multiple loci of control and different perspectives or interests involved. Agents need to interact with each other to achieve individual objective and manage interdependencies, which means they must have social abilities, and leads to their usage in groupware systems (BERGENTI et al., 2002).

A survey of the application of agents in groupware and CSCW can be found in (ELLIS, WAINER, 1999) and (WAINER, ELLIS, 1998). NEEM (ELLIS, et al., 2002),

Personal Assistant (ENEMBRECK, BARTHÈS, 2002) and COLLABORATOR (BERGENTI et al., 2002) are examples of agent approaches used in developing collaborative tools. In a loosely structured scenario, agents may be imbued of the necessary autonomy to control the intensity of information exchange when individuals become closer, or otherwise adjust their behavior according to the groups' needs.

1.5 Architectures for Information Dissemination

Different architectures have been proposed to enable information distribution. For instance, Big Watch (BW) (Kirsch-Pinheiro et al., 2003) is a framework to support the construction of systems that provide past event awareness. The framework distributes event information, and these events are defined and distributed according to designers' determinations. The distribution happens in three phases:

- in the first phase (registering), the designer registers groupware events with the framework,
- in the second phase (monitoring), activities that happen inside the groupware are passed to the framework;
- in the third phase, (notifying), group members are notified of the occurrence of events subscribed to.

The designer needs to identify beforehand which activities are important for awareness effects and register them with the framework. The system enables the definition of filters based on roles, members or a combination of the two, in which certain events can be declared as being of interest to certain roles or members, in order to reduce the amount of information provided.

The COBRA framework relies on constraints to balance both privacy and interruption. In its implementation, a *mediator* is a central controlling instance controls the communication between two elements (called *colleagues*). Colleagues gather or/and distribute information, and the mediator filters this information according to the constraints established. It enables the definition of constraint based filters to represent circumstances and conditions that restrict individual freedom to prevent disruption and protect privacy (Gross, 1999). Several types of constraints may be defined, based on social, physical or legal structures.

Aldunate et al. (2002) present an agent based architecture with two agents: a profile agent and a technical agent. The first constructs, stores, and retrieves information about the user, building the user's mental model, while the second takes care of network infrastructure and communication. These agents provide a location based discovery service, and profile agents attempt to match profiles of users who are in range.

Oh (2004) proposes another agent based awareness system, where each user has an awareness agent that maintains presence information about its user and relays it upon request. This information is picked up via perceptual agents that collect information via cameras or Bluetooth. This information is passed to each user's awareness agent who updates the user's information.

These architectures implement at least two levels, one infrastructural, where communication with the operating system or applications is performed and one for reasoning about preferences or events. They also adopt event based messaging, defining events that are of interest to the user and informing him or her of those when they occur.

1.6 Considerations

Awareness is knowledge about the state of a dynamic environment, which must be maintained over time. Most awareness technologies are implemented as notification servers, where users subscribe to certain events and are informed when those happen (Ramduny et al, 1998). This usually means that some effort must go into configuring the system, telling it what events to observe and send notifications. However, awareness needs change according to the situation, which means notification configurations need to be constantly revised to keep in step with the users' needs. Despite the large number of efforts dealing with awareness, most projects still deal almost exclusively with the question of how to provide mechanisms to integrate and disseminate information, with little or no regard as to the content provided or the selection between different elements. Awareness is maintained through information dissemination and gathering

None of these systems have user profiling capabilities. Additionally, most systems only support pre-existing groups; they do not allow for the discovery of new partners or new collaboration opportunities. In reverse, most recommender systems seek to make introductions and help a user expand his or her contact network, but provide no awareness information nor do they keep track of activities.

IV. Case Studies

In order to better understand loosely structured collaborative work, two case studies were conducted, with two different groups of collaborators. The first was a long term project involving a group of 16 collaborators from different countries in the European Union and the second was a short term project involving a small group of close collaborators.

Case 1: the European case

The goal of this study was to gain a better understanding of loosely structured remote collaboration. The group studied was composed of participants in a research and development project funded by the European Community and involving 16 partners from different countries. A subset of these partners (known as the technical partners) is directly involved in the design and development of prototype systems. The project's lifespan is 4 years, and it is still ongoing.

The project focuses on the development of flexible and interoperable tools to support eGovernment services. Development follows the open source model, and user needs are being elicited through the development of pilot systems in different countries. Later into the project, socio-economic research will address the impact of these technologies on local governments.

This is a fairly large project involving several collaborators who work from different locations and exchange information as necessary. Given the complex nature of the project, many issues are open for research and new technologies must be experimented with. In fact, one of the expected outcomes is a set of guidelines for further research and for the implementation of other systems throughout Europe. Due to the fairly loose definition and open-ended nature of the project, partners had to define more concrete goals and deliverables, setting the expectations for the next few years of work.

The study was conducted through a mix of participatory observation, where the researcher sat in on a two day meeting between the technical partners and interaction analysis, where roughly 18 months worth of forum exchanges and 2 IRC chat sessions were made available for analysis. In addition, informal interviews were conducted to clarify some points. With this study, we were trying to get a better handle on the types

of exchanges the ago on in the group, and how much time is spent on each type of exchange, how awareness is maintained and how coordination and self adjustment are achieved.

Team organization and system support

The project started as a response to a call for proposals by the European Union (EU). Initial participants contacted other partners (using their intensional networks) until a group was formed that fit the requirements put forth by the EU call. In this manner, there was no central controlling figure to put the group together. At the beginning of the project, the definition went through a refinement process, as it had been stated in loose terms and no formal specification was given. The partners refined the project's vision, setting the planning and deliverables for the upcoming years. Once the initial planning and work modularization was decided on, partners spontaneously volunteered for different parts of the project, according to their interests or expertise. Coordination duties for different modules are spread between members.

Several teams compose the project group, and each team has different functions and responsibilities, defined at the beginning of the project. In this study, we focus on the technical research and development teams, since these were more readily available to us. These teams have at their disposal a website, meant to be the main project website, where project related documents are stored and discussions can be undertaken. Partial deliverables are posted to the website with requests for feedback, as well as finalized project specifications and plans. Discussion can be carried out through the use of threaded forums, in a private space open only to partners. Interaction on the forums becomes public to all members. The website provides a public, central repository of data and information on the project, and provides a means for interaction between partners.

Partners tried to organize themselves in such a way as to reduce interdependency, dividing tasks between members to require only loose collaboration. The control structure is variable: despite the existence of a partner in the role of coordinator, the different partners have responsibility for different stages of the project, such that coordination duties rotate between them. However, partners treat themselves as equals, regardless of the appointed coordinator or the overall coordinator. The partner in the role of coordinator carries the burden of writing consolidated progress reports and

reminding other partners of deadlines, but each partner is expected to contribute their own progress reports and to provide input and feedback on the finalized documents before these are sent to the European Commission (EC) for review. Therefore, responsibility is shared between partners, who are all accountable for their parts and for the project as a whole.

Group members have relative freedom when deciding how to go about their business and how to organize and divide their work. This generates a number of problems, as partners organize themselves based only on a tacit understanding that the project must be completed by a given date. They are not hierarchically organized; instead they self-organize and coordinate their own efforts in order to achieve their final goal. Additionally, being from different organizations, partners are often engaged in multiple projects at a time (including projects that fall outside the scope of this one), and must therefore manage constraints and interdependencies arising from the existence of different foci. This configuration corresponds to our focus of research, as it is a loosely structured and self organizing group.

In spite of occasional face-to-face meetings, partners have difficulty coordinating tasks, exchanging ideas and keeping abreast of each other's work. This happens mostly because there is a fair amount of improvisation and adjustment that goes on within each team, and partners end up by stepping within each others' work boundaries and sometimes duplicating work as a result of immediate, individual needs. Additionally, some of the partners collaborate more closely, having individualized exchanges by email or messenger and sharing work, ideas or making decisions about different aspects of the project. Other partners are oblivious to these discussions, and therefore have little opportunity to interact and provide feedback or ideas at a stage where these might more easily be incorporated.

To understand the inner workings of the technical group and how it organizes its tasks and interactions, we conducted an analysis of their online forum logs (which is accessible by all project members), of two online chat sessions conducted early into the project and participated in a two-day face to face meeting, during which, we observed the interaction dynamics and noted meeting topics, annotating problems that came up.

Asynchronous Interactions: Forum Threads

The forum consists of the main form of group interaction, since it is open to all project members. At the time of the analysis, the main forum contained 76 threads (708 messages), the first one dating from 22/06/2004. We analyzed logs up to 18/11/2005, thus a year and 5 months of messages. Average thread length was of 9.32 messages, and average participation rate is of 4.23 distinct participants per thread. Messages in the forum were individually analyzed and coded, and some particular references were highlighted (such as mention to offline work or communication via other, non-public media). Threads represent conversations between members, and so it was interesting to us to verify what the main foci of interaction were. Messages were coded according to their general content, and threads were classified into one of the following categories given the messages that composed them:

- Scheduling – these threads usually contained two distinct lines of discussion: the first, regarding the actual scheduling of a meeting between partners, revolved around agreeing on a location (which might be online via IRC) and finding the best dates for all partners to meet for synchronous interaction. These also included instructions on how to get to the meeting location, where to stay, etc. The second line of discussion regarded an agenda for the meeting. An initial agenda would be suggested by the partner in charge of meeting coordination (a variable role, depending on the stage of the project). Messages would then be exchanged regarding the items that should be in the agenda, and partners would contribute new items when they saw fit. The final agenda would be defined as a result of this negotiation between partners. In addition, partners would sometimes post documents with their ideas or presentations to these threads, so that others could take a look and discuss them at the meeting.
- Announcements – these threads consisted mostly of single messages, announcing an upcoming event (e.g., a conference or the next review date) or a reference thought to be of interest to other partners. These were usually one-shot threads, containing messages to which other partners would not reply.

- Requests for Contributions – since much work was conducted in parallel and without a central coordination scheme, each partner was expected to contribute to the consolidated reports and documents to be presented to the EC. In these threads the coordinator for a given step of the project would request each partner’s contribution to a deliverable, to combine them into one single document to be sent to the EC. The coordinator would usually post this document on the thread for revision by the partners before sending it to the EC. These threads sometimes included clarification messages, where partners would try to elicit what exactly was being asked and what was expected of them.
- Technical Discussions – these threads usually started out with one of the partners asking a question or requesting opinions on a topic, and others then offering their points of view on the subject. A discussion would ensue, with partners exchanging ideas and technical information. The discussion might concern a particular technical question or something that the partners believe may come up during an EC review.
- Status Reports – these threads were similar to announcement threads, except they contained progress reports for one (or more) partners. Partners might post the results of an offline meeting or documentation they recently generated, for instance, so that others have a chance to check it, become aware of what they are doing and provide feedback. Some status reports generated discussion. The main difference between these and announcements, is the fact that announcements regard events extraneous to the group, whereas status reports concern information or artifacts generated by one of the partners.
- Planning – planning threads contained discussion about work subdivision, task assignment, interdependencies, deadlines and project planning adjustments. These were sometimes formalized in documents to be sent to the EC, and were subject to frequent revision. Planning threads also included those in which partners exchanged information to organize themselves, orchestrating a group presentation or other collective activity. In these, it was usually decided who would present what,

partners would share their presentations and others would send in their opinions and ideas.

- Reminders – this type of thread appeared less frequently. It consisted mostly of partners reminding each other to update their information, or that they should publish their papers on the website. There was usually no follow up to these messages.

For the most part, forum threads fell into one of the aforementioned categories. In a few occasions, the theme shifted from one category to another. In these cases, we took the predominant tendency of the thread and classified it accordingly. Forum message breakdown is shown in Table IV-1.

Table IV-1. Forum message distribution

Message Type	Number of Messages	Percent
Scheduling	278	39,27%
Request for Contributions	198	27,97%
Planning	108	15,25%
Technical Discussions	76	10,73%
Status Reports	31	4,38%
Announcements	13	1,84%
Reminders	4	0,56%
<i>Total</i>	<i>708</i>	<i>100</i>

We also had access to two forms of synchronous interactions that happened in the group: face-to-face meetings, to which many partners came and discussed project issues over one or two days, and online chats through IRC, where partners logged in at a prearranged time to discuss pending issues. We looked at two IRC chat logs and attended one face-to-face meeting where we observed the group’s dynamics.

Face-to-face meeting

The six technical partners were present at the face-to-face meeting in which we participated. We noted the topics discussed and the dynamics of the group. We also asked informal questions to clarify partners’ cooperative arrangements. The meeting had been previously agreed upon on the forum, and an agenda had been set according to the group’s perception of points that needed to be discussed. The meeting lasted two days, and featured mostly technical discussion on diverse aspects of the project on the

first day, but was evenly split between technical discussion and planning/organization in the second day.

During the first day, eight themes were treated, of which only two regarded planning (one concerned clarification of what had been previously decided and who was responsible for what and the other concerned what was to be presented at the next review). The remaining six were technical discussions on diverse aspects of the project. These roughly followed the same general format: one of the partners would present some work in progress and solicit feedback from the other partners, who would then provide their thoughts. During one of these discussions, overlaps were found between different members' ongoing work, and the discussion then turned to handling these interdependencies to avoid possible duplication of work. In this case, it was clear that the partners were unaware of each other's work, and some duplication had already happened as a result.

During the second day, nine different topics were discussed. During the first technical discussion, in which almost all partners were involved, some planning also took place, as partners tried to decide on timeframes for deliverables and checkpoints and who would be responsible for what tasks. After assignments were agreed upon, the technical discussion resumed. The following discussion was also technical and the third one regarded planning and establishing interdependencies between partners: one of the partners solicited data from the others, since she could not perform tests of her module without these data. The following discussion (fourth) again regarded planning, as one of the partners asked others to formalize what would be on the report and who would be involved. After that, during an informal pre-lunch discussion, all present checked one of the partners' meeting notes and decided these would be used as the meeting minutes. Partners had been taking individual notes (there was no official, consolidated, minute taking) and it was decided these were the best organized, so would be kept as official minutes. After lunch on the second day, two of the discussions were technical discussions and the remaining two were planning discussions, regarding (1) where to publish documents and shared resources and (2) what to present at the next EC review.

It should be noted that, on the second day especially, quite a few interactions occurred "on the side": partners had informal discussions in groups of two or three and agreed to collaborate on certain aspects of the project, planned to undertake certain tasks together or exchanged information with each other. These informal arrangements were

possible due to the proximity and the sudden ease in determining mutual interests or similar work, and they were sometimes announced to the group (but not always.) These collaborations were not carried out “in the open”, on the forum for all to see, but through email, messenger or smaller face-to-face meetings only between the partners involved. In fact, one of the partners stayed an extra day to work with the host on one of their tasks. Thus, others become oblivious to the smaller group’s activities until these are reported on the forum.

IRC chat meetings

The first IRC chat meeting began with half an hour of technology setup and organization, where partners helped each other get logged in and accustomed to the tools being used, before the meeting actually started.

In the first IRC meeting, four technical partners were present. Of the 10 topics discussed in this meeting, eight were technical discussions, with plenty of back and forth between the partners. One discussion regarded planning and the remaining one regarded scheduling for the next meeting. In general, the dynamics were similar to the face-to-face meeting, and parallel discussions leading to particular arrangements and further meetings between subsets of the partners present also happened here.

All six technical partners were present at the second online meeting. It started with a status report, where partners stated what they had accomplished (or needed) to accomplish certain tasks. The ensuing discussion established interdependencies between partners and what they needed from each other, which covered the second item in the agenda as well. This was coupled with planning for an upcoming review, and partners deciding what was going to be presented. An extensive technical discussion and some planning on the subject of ontologies followed, as partners tried to decide on approaches and assign responsibility for portions of the work. The final part of the meeting concerned scheduling for a later meeting, to be held before the project review.

In the IRC chat meetings, as in the face-to-face meetings, a few parallel conversations happened. These involved a subset of the group planning for further work and establishing individualized cooperative arrangements. In IRC, this becomes harder to manage, as there is no way to establish whom a sentence is addressed to. However, having the side discussion in plain sight made it easier for partners to understand what

arrangements were being made between other partners and to better envision the group as a whole.

Considerations on the Cases

There is a clear distinction between the types of discussion found in the forum (asynchronous) and face-to-face or IRC (synchronous). More than half the messages in the forum regarded scheduling and planning (386 messages) while during the face-to-face and IRC meetings technical discussion prevailed (19 technical discussions in total, versus 13 discussions related to planning and scheduling).

The large number of requests for contributions in the forum betrays the way work is organized in the group: partners are assigned different tasks and are expected to provide their progress reports for the generation of documents to be delivered to the EC. This shows that partners tried to organize themselves so as to minimize interdependency. The progress reports, however, indicate that partners are conscious of the need to keep others abreast of their ongoing work, and that they value the feedback others can provide. These discussions regarding deliverables or ongoing work enable partners to align their work so that it can be more easily integrated at a later stage. They also ensure that partners maintain a shared understanding of the project and of what still needs to be done.

Partners clearly preferred more synchronous means of discussing technical issues, which enabled them to more rapidly reach a decision. While messages in the forum were usually responded within a day, sometimes it would take up to a week for someone to post a response. Quite often, no response would be posted unless partners were explicitly named and requested to provide some input, which brings to light another interesting aspect of the group's work: regulation is exercised socially. While there is a partner in charge of coordinating tasks and producing the final reports, each partner is aware of his or her own responsibility with the group and does not want to be the one that "lets the ball drop". Social pressure leads partners to try to meet deadlines with the others and maintain a certain level of commitment. When it was not possible to provide what was needed by the deadlines, partners would provide an appropriate excuse and an expected date for the required information.

Not evident in the numbers above are the parallel work and discussions. Many threads contained references to work done off the forums (in private discussions via

email, chat or face-to-face) or requests for side communication (communication off the forum, via email, phone or chat). There was at least one of these messages per thread that had more than 10 messages. Once a thread got too long or the discussion got too involved, partners preferred to switch to more synchronous media, where discussion could be carried on more easily and without the time lag existent in the forum.

While not the best medium for discussion, the website provided a shared repository that was valuable for the partners. When there were doubts regarding task assignments, schedules or previous decisions, partners would refer to documents archived on the website (especially after these had been delivered and approved by the EC). Partners would also frequently ask for documents to be posted on the forum in addition to being sent by email, so that they would be permanently archived and so that others could have access to it. The website is, in fact, a rich source of information (all project documentation is uploaded, as are deliverables and reports sent to the EC).

In one interesting case, a discussion between two partners was undertaken on the forum, which allowed all partners to follow it. A few days into it, a third partner saw fit to join in the discussion, providing his ideas on the topic. Soon after, yet another partner joined in, and the thread turned into a lively discussion of technical aspects, in which all technical partners participated. Open discussions enable partners to become aware of issues and to join in when it relates to them or when they are interested. This is akin to a situation found in shared physical workspaces: had the two partners been discussing in the hallway or during a coffee break (or possibly even in one of their offices), others would have been able to join in more easily.

Conversations undertaken in a public space enable users to follow developments and join as needed. Had all partners been in a shared office location, it is likely that they would have collaborated much more frequently, as most of the “parallel arrangements” would have been undertaken in the open, and others would have had the opportunity to contribute when appropriate.

It should also be noted that a great deal of negotiation went on. Most planning and scheduling conversations involved negotiating dates and assignments, but some of the technical solutions also involved deciding on the best solution and how it should be implemented. When a solution was presented to the group or when feedback was solicited, the group would discuss and determine whether other approaches were more desirable. Even simple requests for contributions sometimes involved negotiation: for

instance, when a partner solicited others contributions and someone was unable to comply within the deadline proposed, a negotiation process would take place, with partners proposing alternatives until a consensus was reached and a new arrangement was set.

The main difficulties pointed out by users and observed during the meetings were keeping track of others' activities (what was doing what), to avoid duplication or to discuss the approach at an earlier stage. New members joining: one of the user's expressed a desire to know when partners changes their configurations, bringing someone in. This happened quite frequently, as partner organizations also had other projects and members had to attend to other projects. Additionally, side conversations and groups formed to deal with particular activities, and these should also be supported.

Case 2: the Journal case

In this case, we analyze the activities and communication between members of a small, cohesive group of individuals writing a journal paper. The goal of this study was to better understand communication breakdowns within the group, and the focus was on one incident that caused considerable attrition within the group: duplication of an activity (namely, the translation of the paper to English). The group was composed of 4 members of a research group who discussed in person and agreed to write a paper at an initial moment, motivated by a hierarchically superior figure (their supervisor).

At the initial meeting, they agreed on the paper's topic, picked a target journal and determined a deadline. All had extensive experience with email, had written papers before and had knowledge they were to bring to the paper. Thus, they were familiar with the task at hand and the available communication means, had experience working via email and had input to bring to the paper. The group worked together on this paper during a 3 week-period. Despite 6 face-to-face meetings, most work was undertaken individually, and communication was undertaken via email, with individuals coordinating tasks and negotiating different aspects of the text electronically.

The selected strategy was to conduct a social network analysis of the communication between group members, mapping collaborators, communication and artifact exchange. With the actors' help, we elicited the group's communication flow and activities and constructed social networks in the following manner:

1. each node in the network represents one actor;
2. edges represent communication between actors;
 - a. communication, may include file transference (represented by a file icon);
 - b. dashed edges represent background communication (not undertaken via email).
3. each graph is a snapshot that represents one step in the group's process.

Thus, a sequence of graphs represents the full project time span and maps the communication and file exchange among group members (this could easily be transformed in to an animation.) The full graph is drawn as time passes: with each message exchanged, an edge is strengthened or new members are brought into the fold. This allows us to see the patterns emerging from the data, and to analyze the structure of the resulting graph.

The project started a little before March 31st (there was a face to face meeting before messages started flowing), and ran through April 11th. A daily message and event log follows, along with an activity map. Core group members are J, V, W and Z, who worked most actively on the project. Peripheral members are T, M, A, X and B, who were consulted for diverse reasons during the writing of the paper (in particular, T is the translator, who checked the paper before it was sent). For simplicity's sake, we refer to all actors as he, regardless of their actual gender. There were men and women involved in this collaboration. Except for the translator, none of these peripheral players sent messages to the rest of the group: all were contacted face to face and did not get closely involved with the work in progress. The sociograms were constructed from the records of email exchanges obtained from J, V, W and Z (a partial log is shown in Table IV-2). We used the full set of emails exchanged during the project. Pre-processing this data involved removing redundancies, as emails were provided by both senders and receivers.

Table IV-2. Partial message log for the project

Date/ Time	From	To	Content & Attachments
March 31			
12:16	V	W, J	Proposed structure; WebServices2006.doc
April 2			
23:15	V	J, W	Status update, next steps; WebServices2006-V.zip, diagramas.doc, gcd.zip
April 5			
00:08	V	J, W	Content update; WebServices2006Artigo1.doc
1:59	J	V, W	Content update, comments; WebServices2006Artigo1.2.doc
21:23	W	J	Replies to comments, discussion
21:55	J	W	Comments on content
April 6			
01:42	V	J, W	Content update; Webservices2006Artigo1.3.doc, grid-database.rmx, grid-database.rmd
12:09	W	J, V	Affiliation info correction Webservices2006Artigo1.4.doc
19:47	W	J	Additional author info
19:30	W	Z	Last version of paper Webservices2006Artigo1.5.doc
April 8			
16:23	J	W, V	Final version, info to send to translator Webservices2006Artigo1.5.doc
21:22	W	T, J, V	Request for translation Webservices2006Artigo1.6.doc
April 10			
17:42	J	M, A	Report, coordinate, solicit feedback Webservices2006Artigo1.7.doc
19:11	W	J	Status report, coordination
22:00	J	Z, A, W	Reply from M
April 11			
00:18	Z	J, A, V, W	Submission confirmation DAPD-S-06-0039.fdf

In addition to the email messages, there were offline meetings happened frequently throughout the project. At the last few, the group discovered where they had gone wrong. There was a kickoff meeting between J and W on the 28th, then another between V and W on April 2nd. Then, there were meetings on the 3rd, 5th and 6th as the deadline approached, and a last one on the 10th. In these, members coordinated their

tasks and discussed the topics and contents of the paper, eliciting feedback from the peripheral participants. At the meeting on the 10th, actors became upset when they discovered that they had been working over each other, as sending of the paper to the translator had not been announced and, as a consequence, one of them had also gone through the trouble of translating it unnecessarily. Some of the actors weren't even aware others had joined until the very last day, when they found out at the final meeting.

The log shows a great amount of work being undertaken, initially between a smaller group of collaborators. There are discussions and different versions with pieces of the paper in writing are exchanged. When a reasonably stable version is reached, a new member is brought into the group, who starts to work on the paper and question some of the concepts. The previous members continue working, and the paper is sent to a translator for revision. At the final meetings, actors realized they had doubled up work and had not properly coordinated. The final issues are resolved face to face, and one of the actors is left with the responsibility for sending the final version of the paper after the last group member send his comments.

Group activity is represented in the sociograms in Figure IV-1: in the initial moment, nodes J, W and V work in parallel and interact frequently, discussing, exchanging information and versions of the paper. In a second moment, Z is brought into the group, with the file being sent to him and he starts working on the project. J and V are unaware of the new member of the group, and continue their pattern of collaboration. The third graph shows J and W coordinating to send the paper to T, the translator. Again, Z is not notified, and ends up duplicating the translator's work. In a fourth frame J request feedback from a peripheral member of the group, M. The dashed lines represent knowledge that that was being done. J, V, W and Z had agreed in person that J would send the paper to M, so it there was no explicit communication, but it was common knowledge that that would be done.

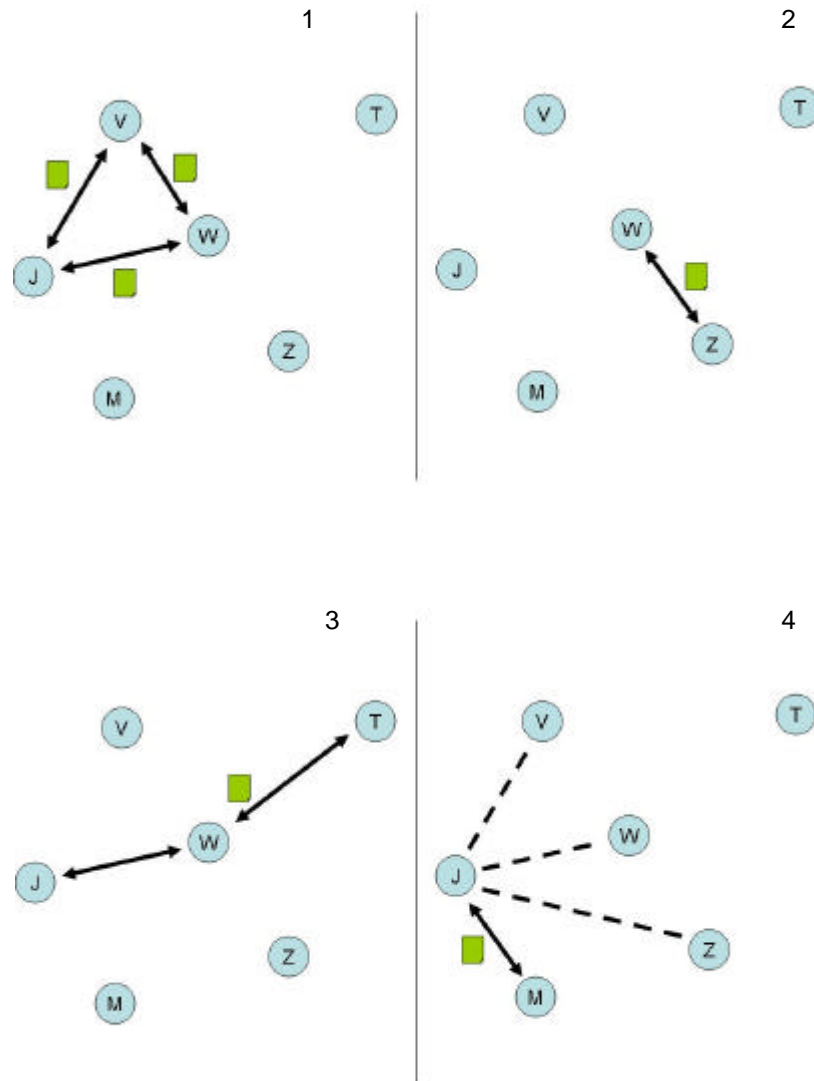


Figure IV-1. A series of sociograms corresponding to different moments in the group’s activities

We measure the outcomes of this project in two ways: whether the task was accomplished and whether the experience was satisfactory. The paper was written and submitted as planned (task accomplished). However, in interviews with three of our four subjects, it became clear that the collaborative experience was far from satisfying. All participants complained of the confusion in communication and of the extra work being undertaken that didn’t need to be. It seems that the group lacked synergy and was fairly fragmented from the start. Comments such as “I didn’t know they were sending the paper to a translator”, “I wasted my time translating this” and “I had no idea he could translate it” were frequently heard.

For instance, one of our subjects complained he wasn’t told the paper was being sent to the translator: he spent the night revising the paper when there was no need, that

work was wasted. Yet another complained that someone requested several changes a few days before the deadline (he hadn't seen the paper before). Worse, one of the collaborators didn't even know another was involved until they all met the day before the deadline to make final adjustments. At the end of the process, they all thought they had worked too hard, were stressed and irritated. In general, the process involved lots of discussion and extra effort that might have been avoided had collaborators been better aware of the social world they were operating in.

Looking at the graphs, it seems clear that the group's boundaries were changing (which is to be expected), and that individuals were not aware of those changes. At a given point, node W became more central, as indicated by its degree of centrality. With the increased centrality, comes additional responsibility: to avoid doubled work and misunderstandings, actors need to be notified of actions that impact on the group's work (such as sending the file to a third party or adding someone to the loop). That was what prevented additional problems when actor M was notified and his comments were requested. Since he already had a stronger tie to actor J, it was left to J to deal with M, sending the request for comments and transmitting the reply to all other actors, but it was common knowledge that that would be done. M is not a local actor, so he was not personally involved and could not participate in the face to face meetings.

In more general terms, it seems that, in a situation of peer work, where individuals have equal say in the project, the ideal configuration would to have all members receive information on all events, which would yield a fully connected graph. When an actor moves to a different position (as is the case with actor W, who became more central as he connected to other actors) or the graph structure changes, something needs to be done. Either this is a legitimate change and the actor must now keep track of interactions and message flow or the change was involuntary (due to a mistake or forgetfulness) and the more "democratic" structure is desired. In the latter case, an underlying system could detect this structural change and "fix" this problem by letting others know of new additions to the group or events that are going on.

The main causes of the breakdowns were:

- Current members did not know of new member's capabilities/skills
- New member did not know of previous arrangements
- Members did not communicate to others execution of activity (duplication)
- Members were unaware of new member joining

This information could have been disseminated in order to obtain better results.

Given our observations in these two case studies, we noticed that information regarding artifact manipulation, group composition and related activities was necessary for the group to function well. Based on these, we developed an architecture, a model and a method to address information distribution. In the following sections, we present a solution to the problem of disseminating information based on these observations.

V. An Agent-Based Architecture for Information Capture and Organization

Given that in loosely structured environments people often select and work with different systems and platforms, we propose the adoption of an agent based layer as an integrative solution. Intelligent assistant agents could help users manage their many ties to peers, linking to the different applications and data sources. We propose an agent-based architecture to support automatic information collection, organization and dissemination. In addition, given the relative independence of each individual, we adopt a peer-to-peer architecture, to create a looser structure and enable more flexible group participation.

A Multiagent Architecture for Knowledge Management

The OMAS (Open Multi-Agent System) architecture stems from a series of intertwined projects, and provides an infrastructure for the development of agent based systems. In OMAS, “coteries” are groups of cognitive agents that work closely together. There is no internal control structure in a coterie, and coteries can participate in higher-level interaction through broker agents. Agents exchange messages asynchronously via broadcast. This information exchange enables them to update their internal models and pursue their goals. The architecture is shown in Figure V-1. There are two types of agents in the OMAS platform:

- Service agents (SAs): provide a particular type of service corresponding to specific skills.
- Personal assistants (PAs): interface with the user. Their skills are devoted to understanding and presenting information intelligently and in a timely manner.

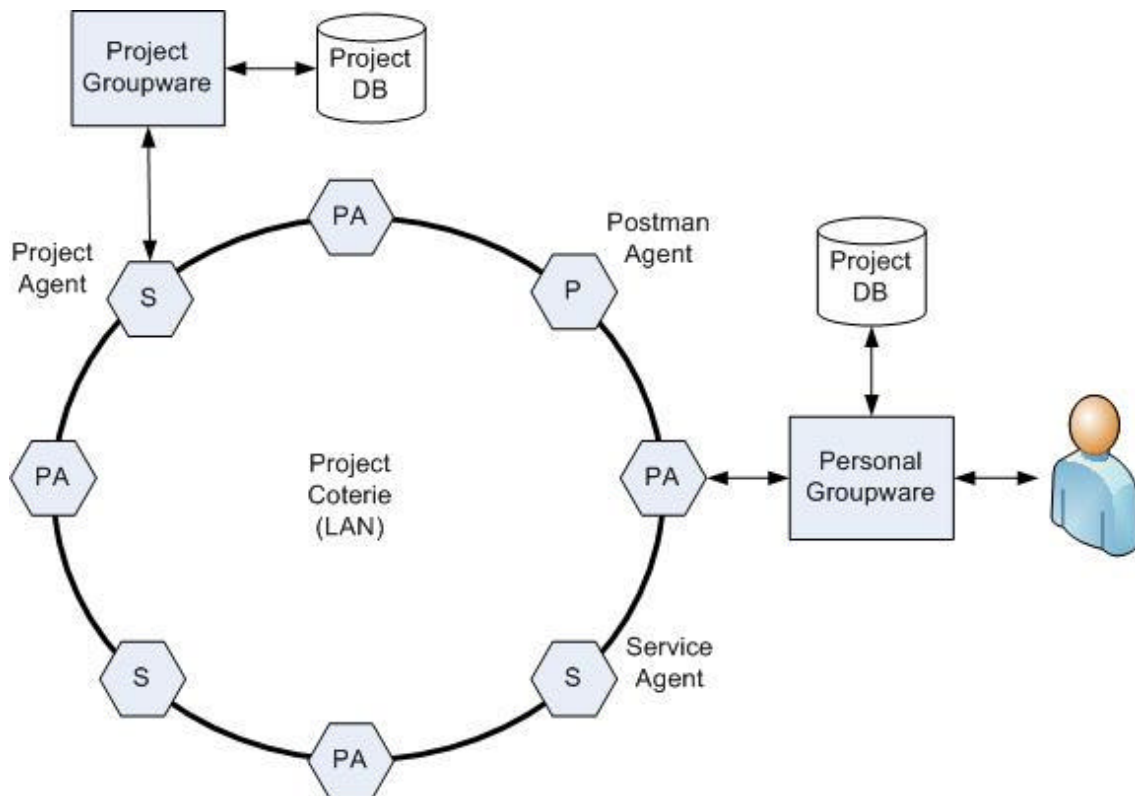


Figure V-1: OMAS multiagent architecture (adapted from TACLA, BARTHES, 2003)

Agents interact with groupware systems, extracting and managing information. For instance, a *project* agent is a service agent specifically designed to interface with project management software. Another service agent, the *postman* agent controls messages sent and received.

In OMAS, each agent is an independent cognitive subsystem, with the internal structure shown in Figure V-2. Knowledge is modeled through several different ontologies, used for reasoning. Agents share a common basic structure, but Personal Assistant agents have a user interface and a user model (shown in gray) that service agents do not. The following components make up each agent:

- Net interface: implements communication protocols.
- Control module: is responsible for the agent behavior.
- Skills: the set of services an agent can provide. A skill can be expressed as procedures or rules.
- World: a representation of other agents (and their skills) and of the environment.
- Tasks: a representation of the agent's current task.

- Ontology: a module to handle the agent and domain ontologies. Ontologies are used to represent skills, tasks and other agent's skills.
- Self: representation of the agent's skills, data stored in its memory and goals.
- User model: user preferences and a model of the dialog with the user.
- User interface: enables communication with the user.

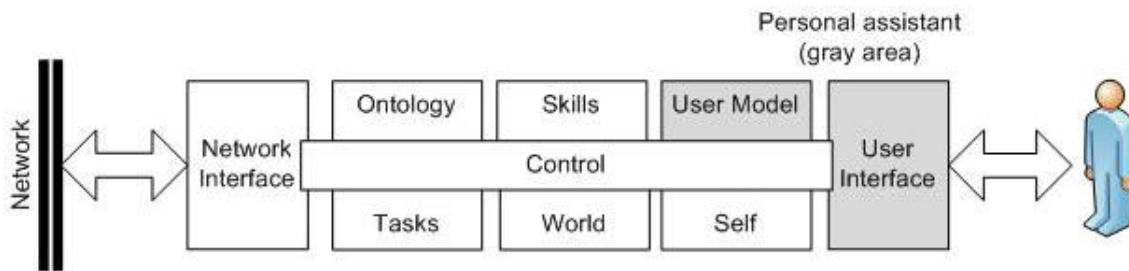


Figure V-2: OMAS agent internal architecture (from TACLA, BARTHÈS, 2002)

CUMBIA, our multi-agent framework, echoes the architecture proposed by OMAS (Tacla, Barthès, 2003): each user is assisted by a group of interacting agents that perform tasks individually to assist the user. Agents collect and process data, providing information when necessary. In addition, we adopt a peer-to-peer architecture to the system, where different coteries representing each user exchange knowledge.

CUMBIA: A Framework to Support Opportunistic Collaboration

CUMBIA is an agent-supported, peer-to-peer architecture to support the management of multiple working spheres, particularly the formation of impromptu groups and establishment of collaboration. It is based on the concept of “unplanned interactions”, or interactions that occur without prior planning by either interactant. These contribute to increase the awareness of what is happening in the workspace and of what others are working on. If the group does not share the same workspace, these informal interactions stop happening and the chances of new, spontaneous collaboration happening also drop.

A Framework for Collaboration Support

As the focus of the CUMBIA architecture is unplanned collaboration, the two main requirements are:

- Facilitating encounters between users with similar working contexts

- Enabling interaction between users;

Agents are designed to collect information about users' activities and look for potential collaborators. As in OMAS, CUMBIA provides each user with a cluster of service agents (a "coterie") to control applications and assist with knowledge management and collaboration tasks. Agents are in charge of identifying potential cooperative situations and providing relevant information in a timely manner. The CUMBIA framework assumes that the group of agents has a shared goal (supporting user collaboration), so no negotiation between agents is necessary.

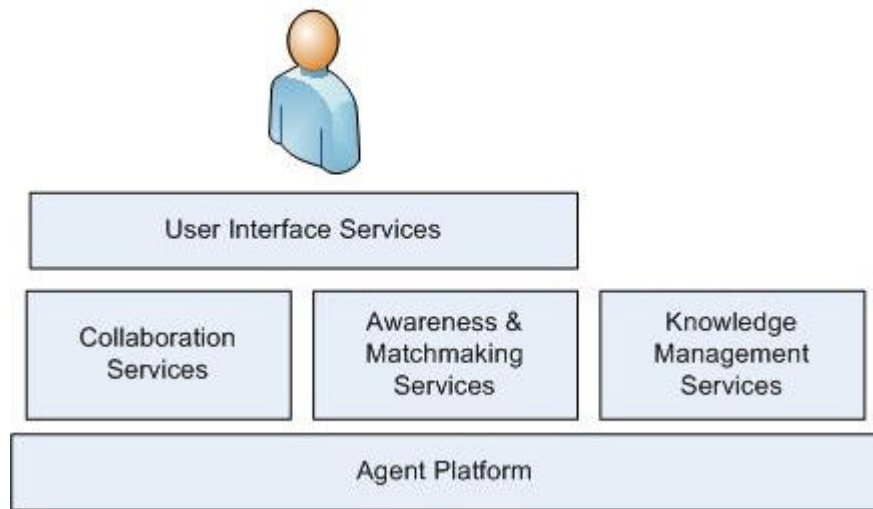


Figure V-3. CUMBIA Service Architecture

In the CUMBIA framework, there are four types of service agents that interact to fulfill these requirements, providing the different services shown in Figure V-3. Four types of services need to be provided to achieve the aforementioned goal. Agents provide these services to support encounters between distributed users:

- **User Interface Services:** display information and allow the user to interact with the system, specifying parameters and information to the agents. This is performed by the PA, as it interacts with the user.
- **Collaboration Services:** enable quick and easy establishment of contact when the possibility for collaboration arises by providing tools for cooperation (forums, messaging, etc.). These are performed by SAs connected to groupware tools.
- **Awareness and Matchmaking Services:** search for other users with whom it might be interesting to establish contact, contact other agents for their users' profiles and work contexts, compare user profiles to current

context and work environment. Performed by intermediary SAs, which are not necessarily connected to groupware tools but receive and process information about the user.

- Knowledge Management Services: manage user's personal data, build user profiles based upon this data and keep track of ongoing activities, including document usage, searches, collaborations and research. Performed by SAs that interface to individual productivity tools (eg., text editors) and to the operating system.

The coterie's goal is to determine a user's multiple working spheres. A working sphere is a locale grouping relevant documents, interactions, contacts or links related to an ongoing collaboration. At any given moment, a user will be working in one of his or her working spheres and may switch between different working spheres. In practical terms, a working sphere determines what resources belong together: documents and links may be saved anywhere on the user's hard disk and be associated with one or more working spheres as shown in Figure V-4. When a user switches to a different working sphere, these documents and contacts become live, as they form part of the contextual environment for this particular activity. Only one working sphere will be on the forefront at any given time, although multiple working spheres may be active simultaneously, indicating that there is ongoing work within different subnets.

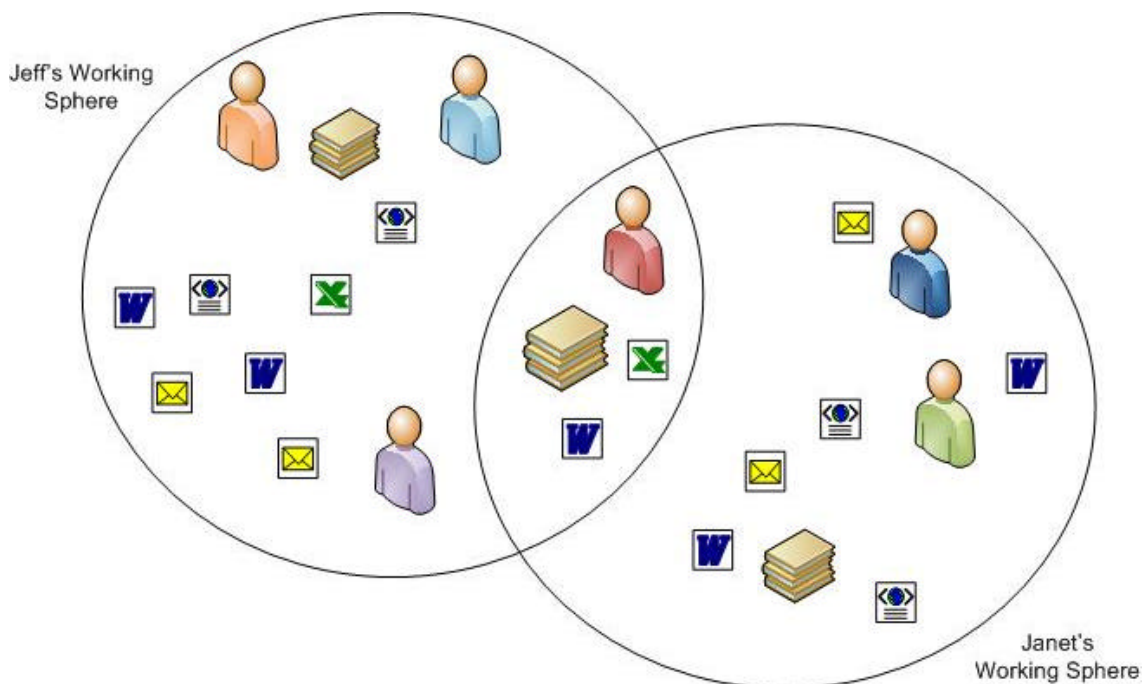


Figure V-4: Working Spheres

Each CUMBIA agent has an individual goal, and communicates with others so that the coterie may achieve its global objective. In this fashion, agents may adjust according to the environment. Each of the services provided is described below, as are the agents involved, their goals and reasoning.

Knowledge Management Services

Knowledge management services involve collecting and organizing information and providing profiling functionality to the system. Agents are in charge of tasks relating to the collection and organization of user data and resources, and the construction and maintenance of user profiles that can then be exchanged with other agents or used to organize resources. They also provide an interface to other groupware and individual productivity systems and the operating system. Agents automatically infer profiles from user activity, but users should always have the last say on their profiles, being able to correct the information and determine which information can be made public and which is to remain private. In our environment, we consider that a person is always working within a working sphere. Working spheres contain projects, documents, contacts, etc. relating to the project at hand.

Working spheres are related to documents, people, collaborations and research, but are inherent to each user, as different roles and hierarchical positions may generate a different view upon the same group context. This is in accordance with the fact that individuals have personal views of reality and organize their work accordingly. The following information may be tracked by Knowledge Management agents:

- **People:** a user's contact list, classified into different categories, such as personal or work contacts, previous, current or potential collaborators, researchers, etc. Some information may be inferred from email or messages sent and received by the user.
- **Contact Information:** information necessary for another person, to contact the user: Name, Title, Email, Phone, etc. Possible data sources for this information would be email, messenger and contact management software, which store information about users contacts.
- **Areas of Interest:** general areas in which the user has some interest. These may be automatically or manually setup, and may be ranked by

interest and activity level. This information could be gathered from files manipulated or websites visited, using information extraction techniques.

- History: agents may track pages the user accesses when navigating the Internet. Work activity should be logged to create work and collaboration histories that may inform future interactions.

Rating mechanisms can be used so that users can rate their peers or resources (sites or documents) according to the importance to the work or project in progress or to an area of interest, providing more information the system can use.

Awareness and Matchmaking Services

Matchmaking services involve allocating users and resources to their respective working spheres, so that related work and resources may be grouped together in order to better maintain their collaborative contexts. Opportunities for collaboration are identified by matching a user's active working sphere and activities with other users' activities or interests. When two working spheres are similar or related (or when a working sphere and another users' resource are related), an opportunity for collaboration might exist. Given the information used by each user, we look for documents that match those in other users' environments.

Matching can be done in real time or "in the background": real time matching occurs when the user is working, inserted in a context and the system searches for potential collaboration opportunities based on that work. Background matching looks for users whose working spheres or resources may be related to working spheres or areas of interest the user is involved in. This is meant to speed up searches: agents independently pre-search the space to build and store simplified user models for potential collaborators, which are then used to make initial matches and search deeper for matches.

Collaboration Services

When an opportunity for collaboration comes up, a user should be immediately notified. Opportunities are time sensitive, and the user should be given information on the other user on the potential for reciprocity (how likely is it that) or that might motivate him or her to enter collaboration. This information may include past partnerships and cooperative behavior or common links between the two individuals.

After the identification of collaboration opportunities, an individual may become an incidental or an active collaborator in another user's working sphere. Incidental collaborators provide occasional suggestions or attend meetings every now and then. Active collaborators are inserted in the project and have to deal with schedules and deadlines. In this case, it's important to know each participant's status and whether any tasks are dependent on him or her.

The initiation of collaboration should involve as little overhead to the user as possible, which means the interface should enable easy activation of a communication channel to establish initial contact. The users may choose to engage in longer interaction, in a chat or message exchange. Another possibility is to have the agent notify the user whenever it locates some information or document that might be useful to a peer and send that information automatically (after asking for permission from the owner). In this fashion, a user doesn't have to worry about locating adequate documents, histories or appropriate information to be sent to others.

Standard collaboration support tools (discussion lists, messaging systems, shared whiteboards, file sharing mechanisms and email) can be provided by agents, which can become active according to the needs. This would help with one of the issues of remembering the preferred communication channels for each of the ties.

User Interface Services

User interface agents provide information to the users and request information from them. UI agents also mediate requests between other agents and users. The basic interface should display little information, so as not to disrupt the user. However, most information should be easily accessible with a mouse click or rollover. An interface agent can select different ways to display information depending on the situation. For instance, it can reorganize a contact list when priorities change or flash icons when a contact becomes active.

Pilot Study: Navigator – a System to Promote Encounters Between Potential Collaborators

To analyze the potential of the CUMBIA framework and gain a better handle on loosely coupled collaboration support, a proof of concept system was implemented. Navigator is a multi-agent, peer-to-peer application to support opportunistic

collaboration that adheres to the CUMBIA framework. In this system, agents gather users' activity information from the computer and construct a profile which is exchanged with peers in the network. Navigator defines time based activity profiling and matching methods necessary for recommendation of potential partners. Users are matched based on their current activities, interests and expertise.

In large communities, finding experts or partners can sometimes be hard. To find new partners, one often asks for recommendations from known acquaintances. This is a form of social navigation, whereby one searches for others in the network, who might fit required profiles. Time is an important factor in this type of situation, as needs, interests and expertise levels change rapidly. Navigator is implemented as a peer-to-peer system to assist social network navigation and search, based on the CUMBIA architecture. Opportunistic collaboration happens when participants use their knowledge about others to identify unforeseen opportunities to collaborate. This type of collaboration can be viewed from two different standpoints: interactions leading to joint work and interactions with the goal of providing assistance. Informal communication is vital for collaboration, as it enables individuals to keep abreast of what's going on in an organization and of others' skills, availability and capabilities. Navigator is a system to distribute information about other's activities, in an effort to promote opportunistic collaboration. The following agents were created:

- Knowledge Management: agents to interface with word and PDF (*Miner Agents*), to access the operating system's task lists (*Tasker Agent*) and to build user profiles (*Profiler Agent*).
- Collaboration: in the implementation, chat functionality was built into by an *Interface Agent*. A better solution would be to link the agents to an external messenger system through an additional agent.
- Awareness and Matchmaking: agents communicate with other coteriees and exchange information about their users, comparing profiles and finding matches (*Matcher Agent*).
- User Interface: an *Interface Agent* displays information about other users, ordered according to potential for collaboration. An alternative radar interface is also available, to convey proximity between peers.

User profiling

Navigator implements 4 agents in the KM layer: Word and PDF Miners, a Tasker and a Profiler. When first setting up the environment, agents build initial profiles by scanning all available information provided by the user. The first step is to process textual documents and contact networks. The following processes are undertaken:

- Processing textual documents: *Miner* agents process documents using an established information retrieval technique. Keywords are extracted from each document and relations to the documents are established. Navigator implements TFIDF.
- Processing task lists: the *Tasker* agent collects task list information from the operating system to determine:
 - how much time the user spent on a given resource or task;
 - how long ago this resource or task was last accessed.
- Building profiles: the *Profiler* agent collects the information extracted from the environment and external applications and consolidates them into user profiles.

When the user is at work, his or her behaviors are observed to furnish some extra information. As the user works, agents observe his or her behavior and process documents in the same manner as when they are batch processed. The following additional information is extracted from the users' behavior:

- Time spent on documents: total time spent on a document is measured
- Access frequency: the system keeps track of how many times the document has been accessed, and when the last access was.

Users' interests change over time, so what is useful or necessary in one moment may not be in the next. To take this aspect in to account, each user profile has two dimensions, a dynamic, short term dimension and a stable, long term one. These are constructed based on an analysis of users' documents and activities, as shown in Figure V-5. Both long and short term dimensions are calculated as a function of time spent on resources.

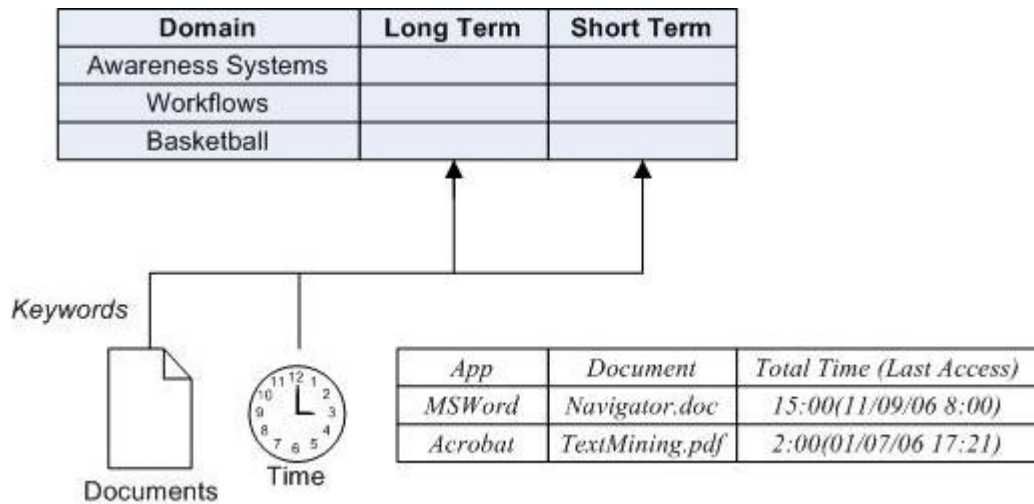


Figure V-5. Navigator User Profile

A profile contains a user's ID in the system, a name or alias and a list of his or her areas of expertise. These areas of expertise are elicited based on document contents. Keywords are extracted from a user's work resources (word documents, pdf files, web pages, email, etc.), and their frequency counts are verified, using a standard information retrieval technique.

Documents are clustered according to keyword co-occurrence, to determine general "activity domains". In most systems, this keyword data is all that is used to build a user's profile and match users. In our system, we also track time spent on each document, which is used to calculate a user's expertise and level of interest on a subject (respectively, the long and shot term dimensions of the profile). These are obtained by a specialized agent that keeps a record of the user's time allocation while at the computer. The activity list is a list of applications and documents used by the users, the accumulated usage time and the last time they were accessed by the users.

The long term dimension of the profile should reflect a user's expertise on a certain domain. Accumulated time spent studying or working on documents should lead to a rough estimation of how knowledgeable he or she is on the domains that document pertains to. Thus, the long term dimension is cumulative, and represents the user's accumulated experience in a certain subject. Profiles are sent to other users over the network.

We assume that a user's resource access patterns should reflect his or her current interests or ongoing work. However, interests and priorities change frequently (for instance, when projects start or end, or individuals move). Information that is valuable or interesting at one moment may not be on the next. Thus, the short term dimension is

transitory, and the values assigned to each of the domains decay with time, to represent a user's changing interests and focus of attention. The more a document is the object of the user's attention, the "hotter" it becomes. Repeated access to the same resource indicates renewed or continuing interest on the document or subject.

All data relating to the domains is tagged according to how recently it has been used and to the total time spent on it. Every time a user accesses a different domain of his profile, this domain is reinforced in its long and short term dimensions. In this fashion, both expertise and interest are calculated as a function of time.

In the current implementation, each resource is weighed according to its total usage time. This weights carries on to terms in the document and to domains to which they pertain. In this fashion, terms from a document that was more heavily used are considered better expertise identifiers than those from other documents.

To calculate the interest profile, documents are again weighted by their usage time, but this weight decays according to the time spent since last access. Thus, these values will decrease with the passing of time, representing fading interest. However, if the user turns back to the resources after a while, the interest is again reinforced. We believe a temporal activity based analysis should be more useful than a simple keyword based one, as it should elicit changes in interests or work needs. For our initial estimate, we have assumed that 1 hour of work equals 24 hours of interest, that is, for every hour spent on a document, the associated interest will last for 24 hours. Thus, if one spends 3 hours on a document, their interest would be reduced to zero after 3 days. This was an initial guess used in navigator, and no further research was undertaken to verify its applicability. It was not used in our later developments.

Awareness and Matchmaking

The identification of an opportunity for collaboration is a process performed by the *Matcher* agent. This process begins by finding individuals in similar contexts (it is always better not to have the user change contexts) and then finding resources these users might share. Once a link has been established, users can be directed to collaborate synchronously and exchange thoughts and ideas regarding each other's work.

Navigator tries to identify people who might become collaborators. To accomplish this, the system processes the profiles received from peers and calculates the similarities between them and its user's profile. The system then shows other users in a

list with their respective similarity ratios calculated. Similarity is calculated using information retrieval techniques.

This calculation enables the extraction of the terms that best define an interaction (composed by a set of textual messages). This technique can also be used to extract keywords from documents (resources) being manipulated by the user.

In this fashion, it becomes possible to tie resources together, linking individuals through artifact manipulation and inserting these into their respective working spheres. Other indexing and retrieval techniques could be substituted through the implementation of other agents.

When two users are working (reading, writing) on similar subjects, the system should indicate proximity, since their current foci (activities) is a match. Long term profiles indicate who experts are on certain subject. Through the long term profile match, the system can determine whether there are other users who know about the subject on which a user is working. Although a specialist might not currently be working on the subject, the system can still use historic information to match it to others.

Having a bidimensional profile enables the system to perform different types of matches, which correspond to different opportunities for collaboration. When considering two individuals, a user is the base point of comparison and each peer is a potential collaborator. We envision the following opportunities for collaboration (summarized in Table V-1):

- A user's short term profile, when matched with a peer's short term profile: this represents a possibility of collaboration and work sharing, as both individual will have immediate needs and be working on the same topics. This may lead to cooperation or labor division;
- A user's short term profile, when matched with a peer's long term profile: the user is actively working on the subject, and the peer is someone who has worked with it before and has experience that may be valuable to the work in progress. The user could ask the peer for assistance with the ongoing work;
- A user's long term profile, when matched with a peer's short term profile: the user is in a position to assist someone who is currently

working on a subject he or she has tackled before. This is an opportunity for the user to offer his or her services as a consultant;

- A user's long term profile, when matched with a peer's long term profile: both peers have worked with the subject before (possibly not together). They may share experiences and lessons learned, discuss alternatives and choices made, reflecting on them and possibly arriving at new conclusions when the problem next poses itself. Their discussions could be a valuable resource to workers who are currently tackling these problems.

Table V-1. Different types of opportunities generated by matching distinct profile dimensions

	Peer	
User	Short Term	Long Term
Short Term	Cooperation	Requesting advice
Long Term	Offering advice	Discussion

The proximity between users' profiles represents proximity between interests and expertise. It is calculated using the co-sine measure of distance between a user's short term profile and the long and short term profiles of each of his or her peers. This calculation yields a value between 0 and 1, where the closer to 1, the more similar the users' profiles are (in reverse, the closer to 0, the more dissimilar the users' profiles). The biggest advantage of this model is that it uses the time spent on each document, which we can use to privilege the terms which are more often or more recently read (depending on what we are looking for).

Collaboration and Interface

Navigator implements one agent to display the information gathered and system calculations, as well as control direct communication with other peers through a messenger interface.

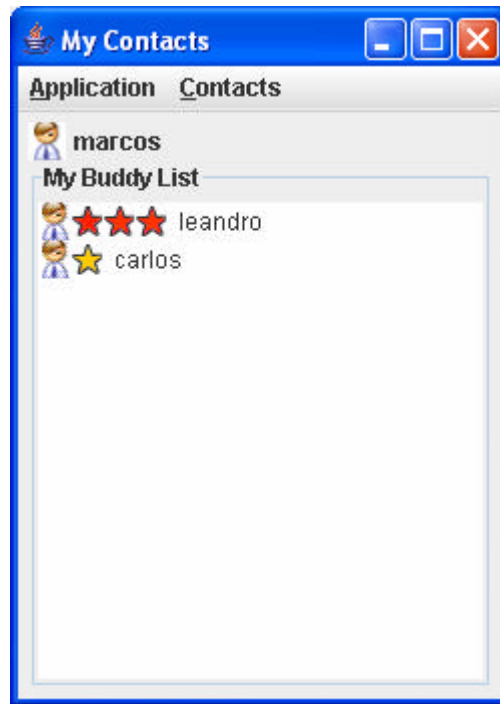


Figure V-6: Contact list view

The main interface is similar to a messenger contact list (Figure V-6) and shows users for which levels of proximity were calculated, indicating similarity ratings. Recent similarity (ongoing work) is represented by the number of stars attached to the user's icon, where the greater the number of stars, the higher the match (see Table V-2). Expertise is represented by star color: yellow, orange or red, indicate growing levels of expertise, as shown in Table V-3. Users may contact each other using a chat functionality built into the system.

Table V-2: Similarity of recent work

Number of Stars	Similarity
☆	0% - 25%
☆☆	25% - 50%
☆☆☆	50% - 75%
☆☆☆☆	75% - 100%

Table V-3: Expertise similarity

Star Color	Expertise level
★	0% - 33%
★	33% - 66%
★	66% - 100%

The system also has a radar view, shown in Figure V-7, which displays the user at the center and other around him or her, with proximity indicating closeness in match

and shirt color indicating expertise (as shown above). This enables not only the visualization of proximity to the user, but also to other users and it allows users to disambiguate between two contacts that may have scored in the same range.

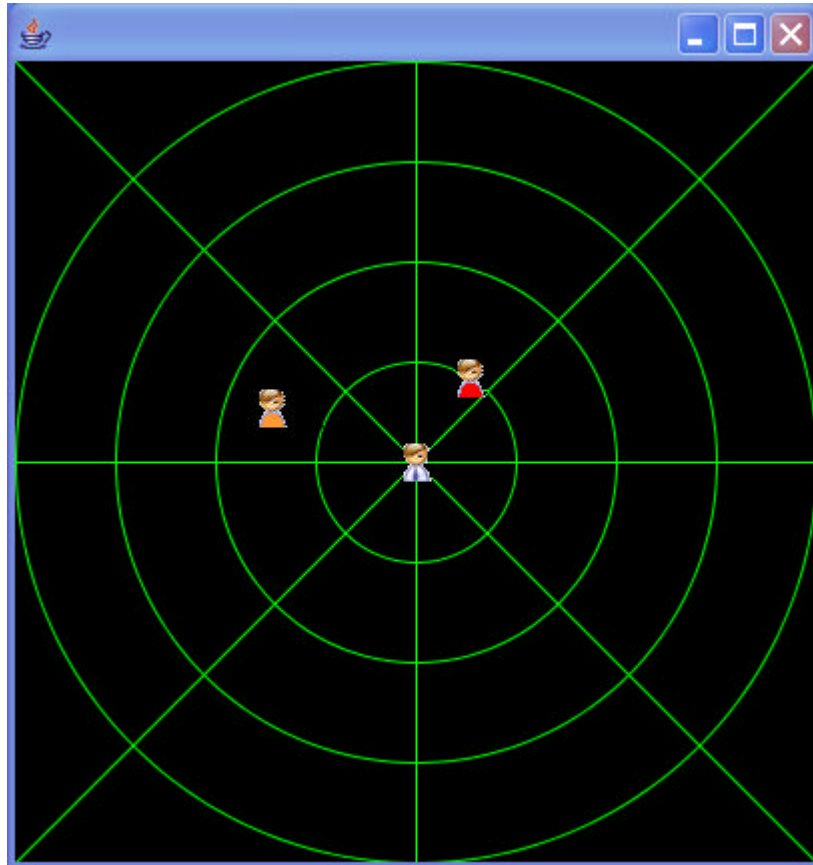


Figure V-7: Radar view

Figure V-8 shows an overview of Navigator's interface, showing the contact list, where, for each contact, a document list is also shown. These are documents currently being manipulated by the peers at that moment. This contributes to increase awareness of other users' activities, showing what the others are working on that is related to the user. The chat interface is also shown, which is the communication channel provided so two users can engage in collaboration.

Technical Details

Navigator was implemented using the COPPEER framework (Miranda et al, 2006). It provides an infrastructure to facilitate the development of agent based, peer-to-peer applications. COPPEER agents can act in parallel and react to the actions of other agents. The collective behavior is the result of the individual actions of each agent,

which can be mobile. The application was written in Java, and uses Lucene (LUCENE, 2007) for textual indexing and searching, PDFBox (PDFBOX, 2007) to access PDF™ files and Jacob (JACOB, 2007) to read Microsoft Office™ files.

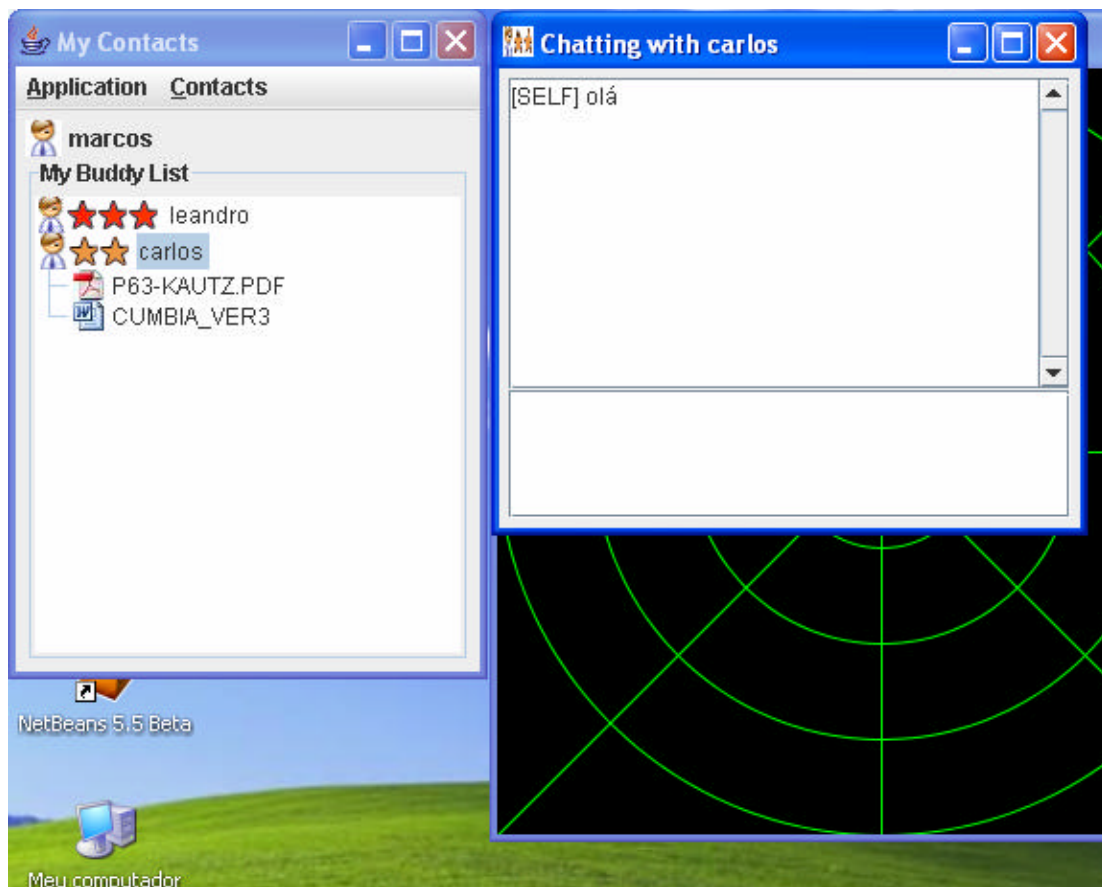


Figure V-8. Navigator interface overview: peer list window, chat window and radar display (partially shown)

Navigator allows a constant fluctuation of users in the network. Each peer manages its own connection with other peers, without a central controlling entity. Navigator implements five types of agents, which have already been mentioned in previous sections and are listed here: text miners (Miner), task gatherers (Tasker), profile builders (Profiler), matchmakers (Matcher) and interface (Interfacer) agents.

- *Miners* analyze users' documents and extract the data to build the profile. These agents are specialized according to documents types: there are currently agents to read Microsoft Word™ (DocMiner), PDF™ (PdfMiner), and plain text files (TxtMiner). Miners index files and combine them into one full index. This consolidated index will be used by the profile builder agent (Profiler) to create the users' profiles.

- The *Tasker* agent monitors running processes by querying the operating system directly, capturing information about the execution of applications and resource usage. This is the activity information used to calculate users' profiles.
- The *Profiler* agent aggregates the activity and keyword frequency information generated by Miner and Tasker agents, into a profile, calculating both the long (expertise) and the short (interest) term dimensions. These are put together using an aggregation function, as described before.
- The *Matcher* is responsible for comparing its user's profile with other users' profiles. It uses data generated by Profiler and incoming profiles from the network. Two different matches are calculated for each peer: a user's short term profile against other user's short and long term profiles. In each case, this yields two measures of proximity, and each peer will be categorized accordingly.
- The *Interfacer* agent coordinates the graphic interface and all communication with the user. It uses the calculations made by the Matcher agent to provide a visualization of opportunities for collaboration. Figure 1 displays the user interface: there is a messenger like contact list, where each peer is assigned a set of colored stars. The number of stars indicates the short term profile match (possibility of work division), and colors indicate expertise match (possibility of consultation with an expert). The user can start a chat session with other peers to engage in active collaboration, should both parties so desire. Double clicking on a peer displays his or her current activity list. An alternative visualization is provided: a radar display where the user is the center surrounded by peers. Proximity indicates closeness in recent activities, colors indicate expertise.

Due to the need to keep the user's profile up to date, the Miner and Tasker agents execute continuously, recording information at regular intervals. The interval is customizable by the user (the default is 30 seconds). Data is sent from one agent to another, and the profile is updated accordingly. Figure V-9 shows information exchange

among Navigator agents. The user interacts only through the interface, where other users' names and degrees of proximity are shown. Through the interface, users gather knowledge about others in their domain and how they might help with studies, problems or tasks. Agents are proactive in their recommendations and in profile construction. When there is a possibility of collaboration, it is shown without user intervention.

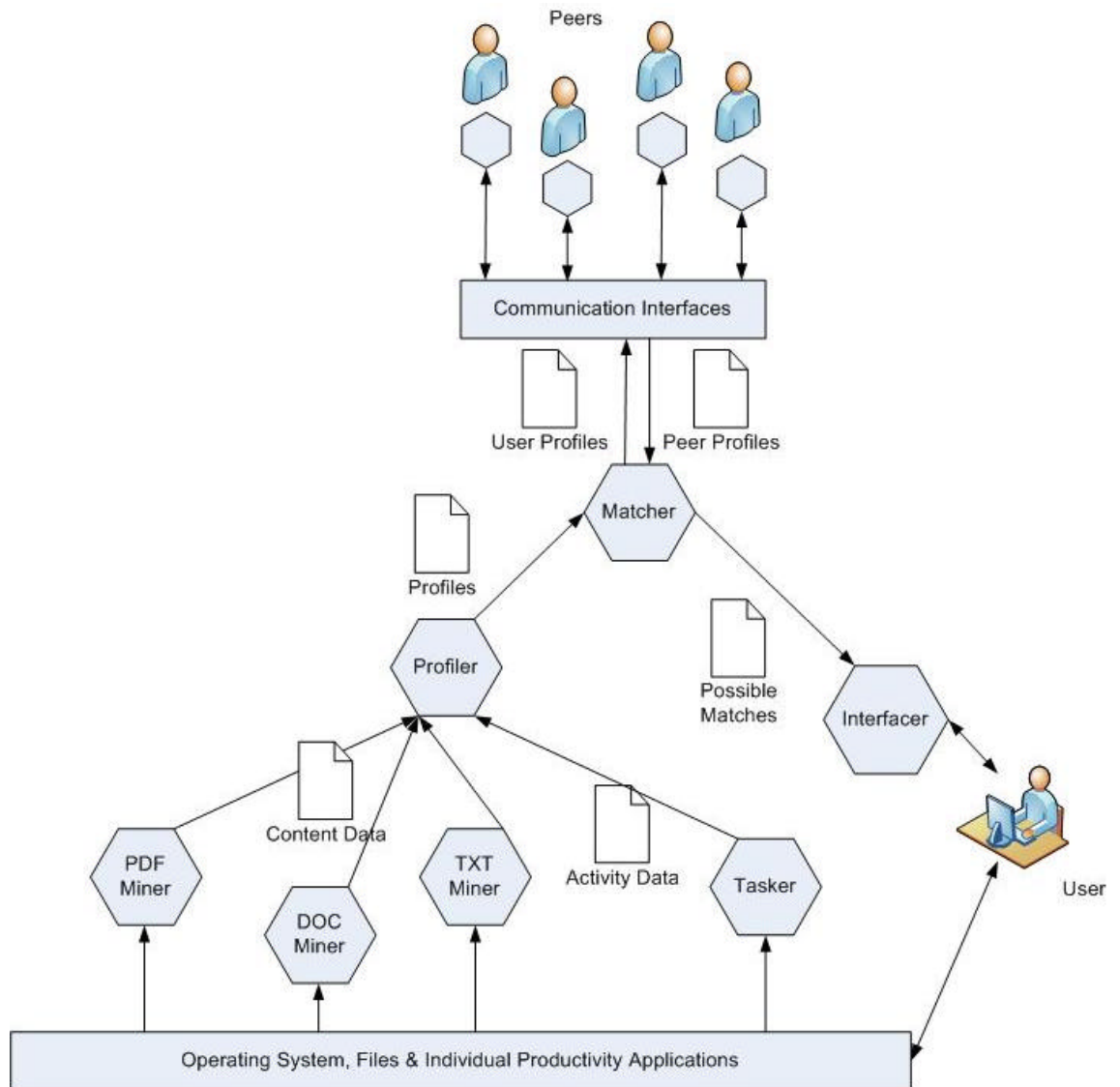


Figure V-9: Navigator Agent Architecture and Communication

Lessons Learned from CUMBIA and Navigator

One of the advantages of Navigator is the way in which the information is processed for user profile creation. Navigator agents search within the user's documents for key terms and their frequencies. However, Navigator extends that approach by taking into account the time spent on these resources and how recently they were used.

This alters the calculation of the relevance of each of the document's, to reflect it being the focus of the user's attention.

Navigator's goal was to support loosely structured, opportunistic collaboration in remote settings, by facilitating unplanned interactions. Agents located potential collaborators based on an analysis of activities performed by their users. Even though no formal experiments were conducted, a few problems and observations were gleaned that helped direct our research. One of the problems found when running the system a few times was the large number of messages sent and received between agents. Even with as little as 5 users, the system quickly caused memory overflows and halted. Early tryouts quickly flooded the system with messages, given the large number of data being transmitted. This is a technical difficulty that can be circumvented through configuration of message sending behavior (how frequently messages are sent, etc.), but a concern nonetheless.

Upon reflection, our experience with CUMBIA and Navigator systems indicated that they were too open in nature. The fact that Navigator searched for anyone in a potentially large network would most likely lead to a large number of matches, generating a flood of information that a user might find hard to handle. A large amount of information would not be very useful to the user, as it would leave many activities and events to be manually sifted through.

Navigator goes from artifacts to people (look for manipulation of similar resources and find out who is acting upon them), which is a frequently used strategy (e.g., see FONER, 1997). We believe this strategy would not function as well for ongoing collaboration, as it doesn't analyze the involvement of contacts with the resources being analyzed, focusing only on activities and interests. While it may be good for finding new contacts (as it searches for people regardless of their prior relationships with users), this strategy is not necessarily the best for ongoing collaboration. In this situation, it seems it would be more appropriate to search initially for collaborators and then for the artifacts used in their collaboration. With this in mind, we set out to design a method that would place greater emphasis on ongoing collaboration, selecting contacts and information that would be of interest to the user, given ongoing collaborations.

VI. A Conceptual Approach for the Discovery of Collaboration Spheres

To be able to assist a user in a dynamic, networked environment, an intelligent system needs to have an up-to-date model of its user's contact network. The system also needs to be able to determine what portions of this network are active (what the state of each tie is), and what each link concerns. This model should be able to indicate who a user's current collaborators are and what their shared collaborative context is.

Inspired by Gonzalés and Mark's (2004) concept of *working sphere*, we coined the term *collaboration sphere* to represent a construct that ties together active portions of the user's intensional network and the themes and resources related to it. It is similar to the notion of *working sphere*, which brings together events involving a group of people with a common motive and resources, under the viewpoint of a single user. A working sphere represents an individual's view of a collaborative project, preferences, hierarchical view, etc. A collaboration sphere bridges several working spheres, providing a unified view of a collaborative project. Working spheres belong to different individuals but relate to the same joint effort, such that when activity happens in one working sphere it should be correlated to someone else's working sphere through the collaboration sphere that links them. In this fashion, events involving only one group member may be tied to other group members and their working spheres. A schematic view of collaboration spheres is shown in Figure VI-1. Note that a collaboration sphere will attempt to link elements from different working spheres, there may be elements that may belong to a working sphere but remain private to an individual. In the same manner, individuals may form subgroups within one collaboration sphere and constitute collaboration sub-spheres for those interactions.

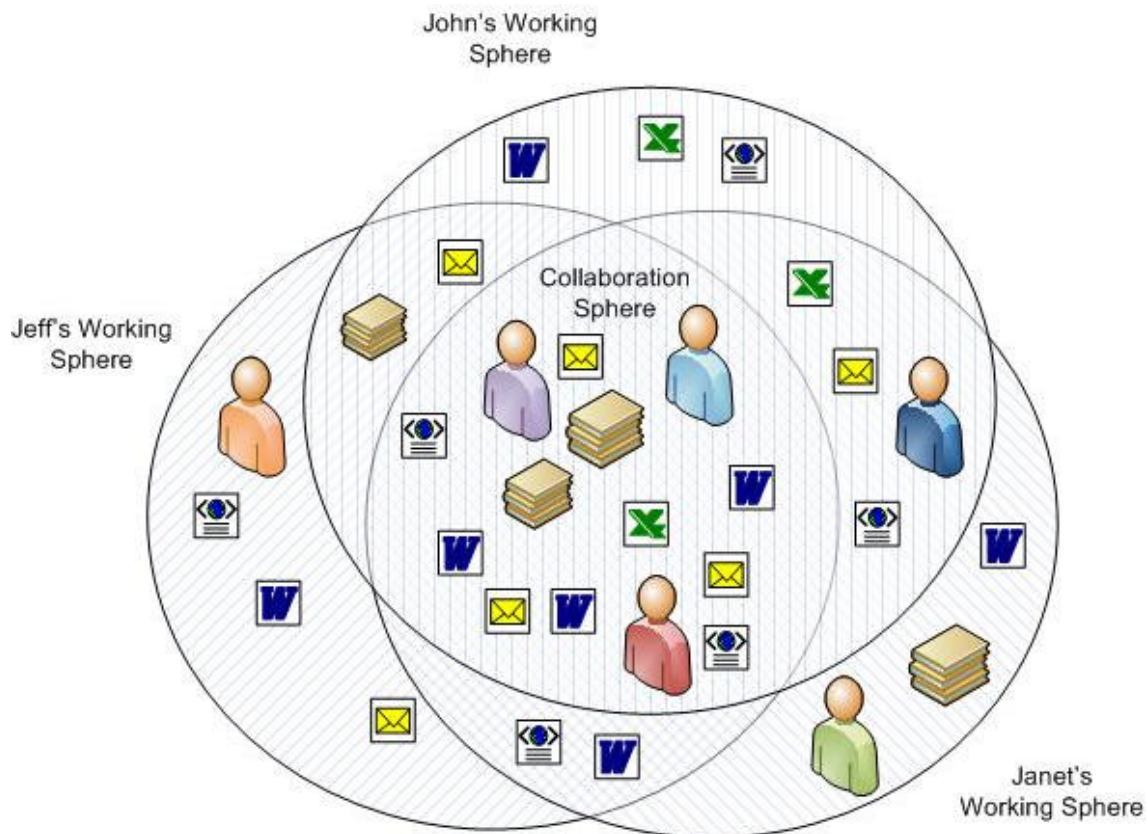


Figure VI-1: Working Spheres and Collaboration Sphere

A collaboration sphere ties together portions of the user's intensional network, the themes and resources related to it and portions of other users' networks. It describes a relationship between two or more parties, and contains information about personal, organizational and project aspects of a project undertaken within these relationships. Additionally, it contains information about prior activities, which establishes common ground and generates expectations of others' behavior based on prior history. This helps users decide who they will be working with and adjust their behavior accordingly.

A collaboration sphere contains a record of past and ongoing interactions. It must reflect the state of each tie, including ongoing projects, with a special emphasis on variables that influence a user's actions. Characteristics of a collaboration sphere are:

- State: whether the sphere is active or dormant;
- Nature: whether it is personal or business;
- Formality: whether it is formal or informal;
- Participants: peers involved in the project;

- Project Structures: describes organizational structures adopted for this project. These may be formally defined (assigned by the organization) or informally distributed (volunteers pick up the various tasks that need to be performed).
 - Task subdivision: how tasks are divided between participants;
 - Task interdependency: where are the interdependencies between tasks;
 - Reporting structure: who should an actor report to when done? Who are results to be sent to?
 - Role assignment: what are the roles in this project and who is in which position?;
 - Ongoing work: how is work progressing within the group project?
- Constraints: what constraints exist for this context (ex. secrecy, NDAs, deadlines, available tools, etc.)
- Themes: what are the subjects this context deals with?
- Goals: what are the goals of this collaboration?;
- Locales: describe the sites and means through which collaboration is undertaken.
 - Sites (shared spaces or tools and communication media adopted);
 - Means (shared artifacts and resources).

These variables were taken from contextual elements described by Borges and colleagues (Borges et al., 2004) and elements described as influences to individual decision making in collaborative environments described by Bernoux (1995), and our observations in the cases studies. The elements and their interconnection are shown in Figure VI-2.

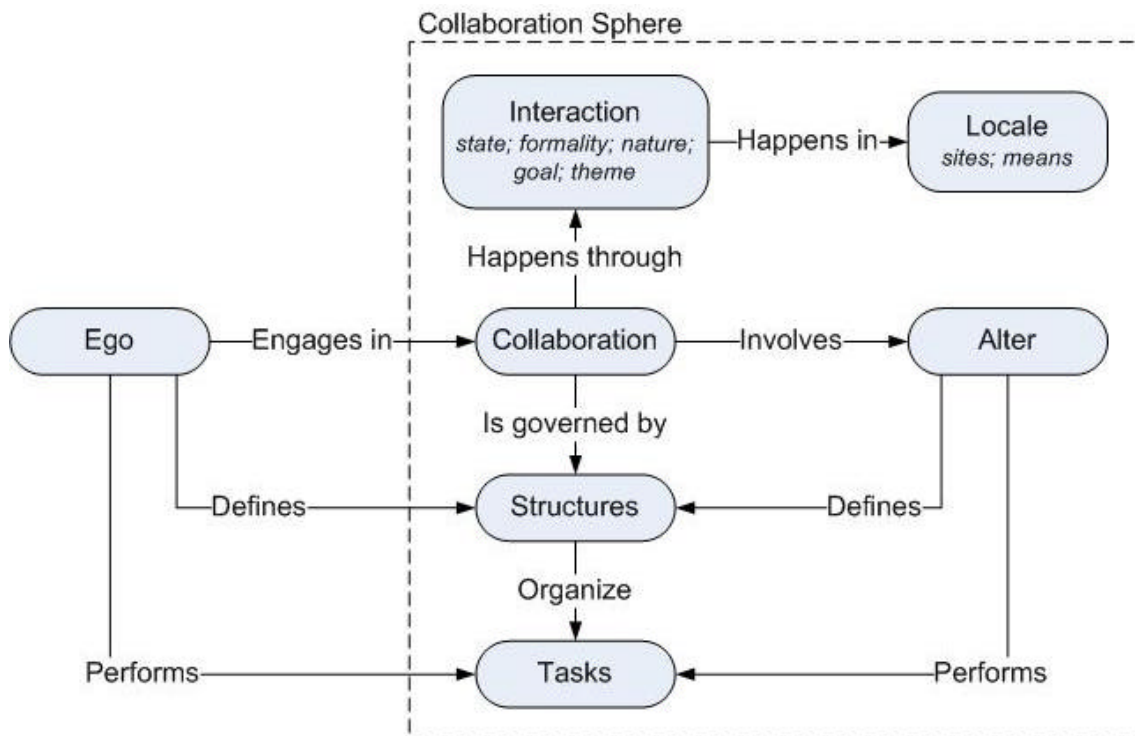


Figure VI-2: Ontology of collaboration sphere concepts

It should be noted that a tie is a rich, individualized concept that reflects the relationship between two individuals. It exists in and of itself, regardless of working or collaboration spheres and is constructed over time, which means it includes interaction histories and personal preferences or affinities. It may have several working spheres associated with it, as a user may have collaborated with one person more than once, and each of these would describe one joint project. Each person also has certain individual characteristics that reflect their skill levels and abilities, as viewed by the user. This means each user should maintain a model of each peer in order to better be able to navigate his or her intensional network.

Characteristics such as nature, formality and communication media may also be viewed as characteristics of the tie as opposed to a particular context. Many relationships have both personal and business facets to them, and more research needs to be done to determine how people manage these aspects and to what extent a collaboration sphere is influenced by them.

In networked settings, individuals need to take stock of the different collaborations they are involved in, their roles in each one and the organizational constraints they are subject to. Managing multiple collaborations involves understanding one's social environments, being able to tie people and resources to each of these, and identifying opportunities for interaction, and two challenges to individuals

working in these settings are remembering and communicating. For instance, remembering that an activity or a document is relevant to a particular group and communicating with others about it helps strengthen individual ties and maintain the flow of joint work. To facilitate multitasking between several collaborations, social contexts and activities need to be tied together, which would enable information to flow and easier context switching by each participant. This is accomplished through collaboration spheres.

In this scenario, certain situations may constitute an opportunity for strengthening a tie, or there may be information pertinent to ongoing collaborations that could be relayed to other collaborators. Regardless of the scenario, detecting and supporting relationships involves determining the state of a tie (whether active or dormant) and what artifacts and information are involved in this collaboration.

Supporting work in dynamic, loosely structured conditions involves keeping the user abreast of ongoing work, providing him or her with information to enable decision making and adjustment to unexpected situations, helping the user remember his or her multiple working spheres and collaborators and communicate with others as necessary. In this section, we present our approach to the identification of collaboration spheres through interaction analysis. A collaboration sphere is a rich concept that involves several elements that describe collaborative arrangements. It is beyond the scope of this thesis to try to elicit all of them. We attempt to determine the two fundamental elements, people (who) and artifacts (what) involved in a collaborative project. The information that supports adjustment and self organization concerns events relating to the elements contained in each collaboration sphere, especially people and artifacts that are involved in it.

Constituting Collaboration Spheres

Our goal at this stage is to develop a method for determination of collaboration spheres. We have devised a two-tiered method to reach that goal, discovering collaborators and tying these to resources or activities (that involve information manipulation). With this in hand, it become possible to analyze users' activities and allocate them to collaboration spheres as appropriate, distributing event information as needed. The method presented is based on our experiences with the CUMBIA (Vivacqua et al., 2005) and Navigator (Vivacqua et al., 2007) projects described in the

previous section. Initial tests with CUMBIA and Navigator quickly led to the realization that filtering mechanisms were necessary to reduce the amount of information gathered and distributed. With this method, we move towards the goal of helping users manage their ties to others and support loosely structured collaboration, by identifying working spheres and binding them to collaboration spheres, focusing on current collaborators and providing information related to joint tasks.

In the remainder of this chapter, we assume a distributed, peer to peer system (as described in the previous chapter) to link users to their peers, through the provision of awareness information relating to active collaboration spheres. In this scenario, each user is assisted by a coterie that reasons about users' interactions and activities and exchanges information with other peers, creating an information flow to keep group members informed of ongoing activities. Thus, each coterie's goal is to *maintain awareness between peers by distributing information about the activities of its user's acquaintances*. To reach this goal, the agents:

1. collect information generated by the user while working on his or her computer;
2. exchange information with other users' collectives; and
3. provide information to the user about his or her alters' activities.

Given the large amount of information available, filtering mechanisms must be put in place to select the information to be distributed and displayed. The agents focus on active collaboration spheres. The model presented here fits into the user model in the agent architecture presented by Tacla and Barthès (2003), described in the previous chapter.

Since communication is at the basis of any collaboration and users in loosely structured collaborative situations coordinate heavily through linguistic exchanges, we elected the analysis of interactions as the cornerstone of our approach. We expect this to work especially well in cases where users rely heavily on email, messenger and other communication media to establish contracts and get work done, as is the case in distributed groups.

Interactions are analyzed structurally and for content. Given that attention is a valuable resource and that users will actively attend to the information or messages most important to them, we equate attention dispensed with importance and activity. Attention foci are determined through an analysis of the user's ongoing interactions, and

consider that colleagues that receive more attention are the ones with whom the user is actively collaborating. Supporting loosely structured ongoing collaborations involves discovering active ties in the network and the work they refer to.

Communication doesn't only activate different parts of one's network: patterns that reflect closer collaboration emerge over time. These observations lead us to an approach based on the analysis of electronic records of interaction (e.g., email, fora or messenger logs). An inherent limitation of this choice is that this approach will only work as long as the majority of interactions are conducted through electronic media, leaving a trace that can be analyzed computationally.

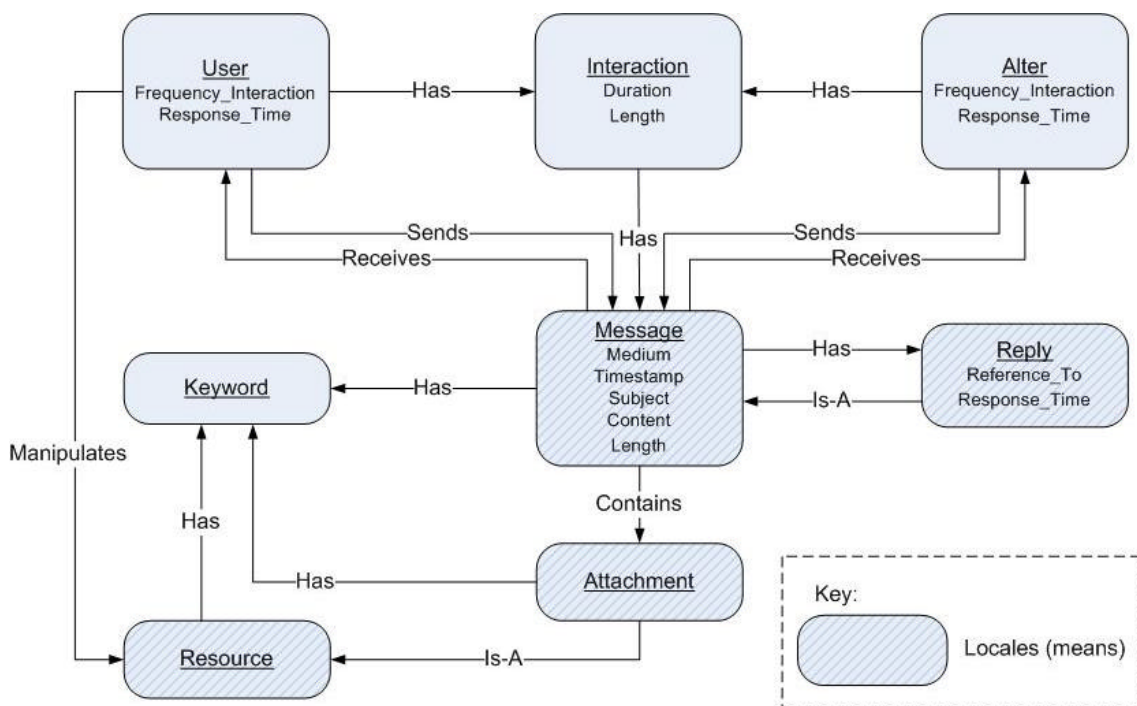


Figure VI-3: Ontology of concepts involved in the discovery of collaboration spheres

An ontology of the concepts involved is shown in Figure VI-3. This information is organized to represent ongoing relationships and interest foci. The *user* is an individual who performs his or her own work in a computational environment and interacts with others (*alters*) through email or other computer-based media, and is the center of analysis. A tie is a relationship between two users, built over time, through a series of *interactions*, which may be work related or personal, creating a multifaceted relationship. An interaction involves a series of *message* exchanges (messages and their *replies*) between two users. These could potentially happen over email, messenger, discussion fora, in person or on the telephone. Messages may contain *attachments*, which are also *resources* manipulated by users and included in the working spheres.

Messages and resources are described in terms of *keywords* that provide an overall theme and help establish proximity between different working spheres.

An egocentric network is built by linking the user to his or her many contacts. This network can be viewed as a tree with ego (the user) at the root and his or her alters (the people who have sent or received messages from ego) at the first level. Resources and information related to the interactions is then added as a second level (themes, artifacts, etc.) Determining ongoing collaborations is a question of determining which ties are currently active and need to be monitored and pruning the answer space accordingly. To that end, branches at the first level are weighed to represent intensity of activity. To be able to provide activity information (to maintain awareness), the system must allocate activities and resources to the appropriate collaboration spheres, which makes it easier to keep track of events that may be related to ongoing work.

That which is relevant to the user we call the *focus of interest*, and its determination is a two-step process, shown in Figure VI-4:

1. discovering which peers the user might be interested in (selecting nodes at the first level); and
2. figuring out which activities relate to this collaboration sphere.

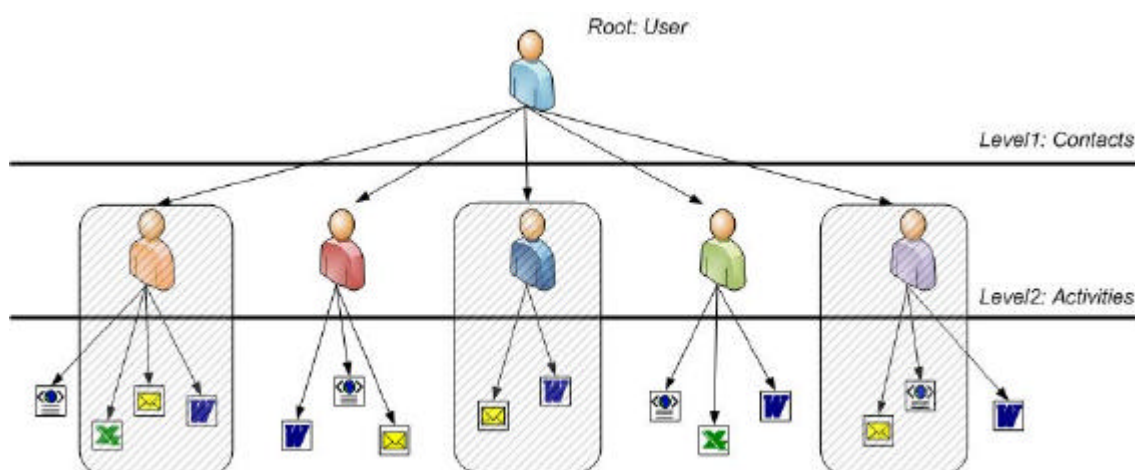


Figure VI-4: Two level search to distribute information

Email Based Interactions

Given that email has become one of the most frequently used communication media, an instantiation of these concepts was created for email-based interactions. In

this scenario, information about ongoing interactions is gathered from email logs and used to reason about a user's needs. We assume that each message contains a sender and a recipient, date, subject, content and attachments. When a message is a reply to another message, the response time is also saved, in order to verify urgency. In the user's network:

- *Ego* is the user (normally determined by looking at the *From* field of outgoing emails, or manually informed) and
- The *Alters* are the other email senders and recipients (taken from the *From*, *To*, *CC* and *BCC* fields), which form the user's list of acquaintances.
- Each *Message* may contain one or more *Attachments*
- *Interactions* are composed of email messages and their replies (determined via *Subject*, *Message-ID* and *Reference-To* tags present in message headers) grouped together, with each reply is qualified by the time it took the user to respond (extracted from the *Date* field). An interaction, in turn, is qualified by its length (number of emails) and duration (time from first to last message).
- A *tie* is characterized by the frequency of interaction between alters, i.e. how often they exchange mail. This helps determine "normal" behavior towards each alter, which can be used to detect increase in communication that indicates collaboration.

Figure VI-5 shows a UML representation of the data as it is extracted from data sources for processing. As working spheres and intensional networks are individually controlled, we adopt an egocentric standpoint, analyzing each person's interaction records individually and moving from these to the groups they relate to.

For ego, average frequency of interaction (how often he or she sends/receives email) and average response time (how quickly does ego respond to any email) are also calculated as a whole and per alter (how often does ego send email to alter A and how quickly does ego reply to messages from alter B). Co-worker groups are determined through co-occurrence in message headers, and provide an initial formation of a collaboration sphere.

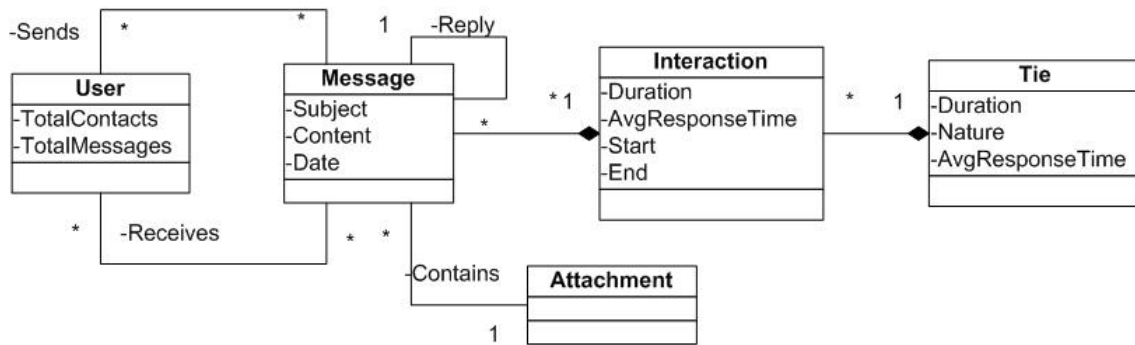


Figure VI-5: UML model of user interactions, applied to email messages

Discovery of Active Collaborators

The full set of user acquaintances is extracted from the interaction logs (all contacts who have sent or received messages to the user). The first step in discovering collaborators is to obtain a list of all potential collaborators (*alters*), that is, all individuals with whom the user (*ego*) has communicated at some point. This forms the user's full personal social network, from which collaborators will be drawn. These can be obtained by extracting senders and recipients (*to*, *from*, *cc*, *bcc*) from messages sent and/or received by ego. Each alter is related to the user by a tie quantified by the number of messages sent from ego to alter and received from alter by ego.

We thus define the following sets (in terms of email interactions):

Senders = all contacts found in the FROM headers

Recipients = all contacts found in the TO, CC or BCC headers

Recipients(Ego) = all contacts found in the TO, CC or BCC headers of messages sent from Ego.

And the following variables:

Contact = any individual who sends or receives email

Sender(Msg_i) = contact found in the FROM header of a specific message

Recipients(Msg_i) = contacts found in the TO, CC or BCC headers of a specific message

This full set is often quite large and contains several alters that can be discarded. We interpret each message received as an attempt at gaining the attention of the user, and each message sent as a conscious expenditure of attention (Davenport and Beck, 2001). Tyler and Tang (2003) suggest that contents of the Outbox are more important

than the contents of the Inbox in this type of analysis, as they reflect interactions the user has actually decided to engage in. Thus, from this initial set, we remove all alters to whom ego has not sent any messages (that is, individuals who have sought ego's attention but have received none, so they are most likely involved in a lower priority collaboration sphere if at all). The *Discard* operation is defined:

Discard(*contact*):
 if (*contact* ∈ *Senders*) ∧ (*contact* ∉ *Recipients*(*Ego*))
Discard = *True*
 else
Discard = *False*

The remaining contacts are the *alters* in ego's network, the ones ego is effectively devoting time to. After removing alters that received no attention from the network, ego's "normal" behavior is inferred from previous interactions. The average interaction rhythm is calculated in terms of: number of messages sent per day/week/month and time between the receipt of a message and its reply. These indicators are calculated for each alter, and this information is added to each ego-alter tie. This information is also calculated globally, for those situations when new alters appear in the network (either as senders or as recipients of messages) and there is no prior data with which to assess ego's behavior towards them:

MessageSent(*ego*) = Message where Ego is the sender

MessageReceived(*ego*) = Message where Ego is the recipient

Message(*ego*, *contact*) = message from ego to contact

TotalTime = total time since the first message in the log was sent

TotalTime(*ego*, *contact*) = accumulated interaction time between ego and contact (time from first to last message from ego to contact contained in the log)

Co-occurrence in messages (i.e., multiple recipients) indicates the social groups a user is part of. Collaboration spheres often involve more than ego-alter pairs, they frequently involve multiple people (ego and several alters). Discovery of multiple participants simultaneously engaged in a collaborative endeavor happens through analysis of co-occurrence in message threads: groups of alters participating in the same discussion have this thread as a common point of interest (thus, recipients of the same messages would be bundled together as a social world). The fact that each recipient is

consciously selected by the sender (whether it be ego or an alter) as a recipient in a message is taken into account in this step.

$Group(contact_i, contact_j)$:
if ($alter_i \in Recipients(Msg_x) \wedge (alter_j \in Recipients(Msg_x))$)
 $Group = True$
else
 $Group = False$

To determine which groups are capturing the user's attention, the system looks for discrepancies between the user's current behavior and the user's normal behavior. Variables that characterize email exchanges are the number of messages exchanged and response time. For each alter, we compare the current behavior to the normal behavior (the previously calculated average).

$AvgActivity(ego) = Count(MessageSent(ego)) / TotalTime$
 $AvgActivity(ego, contact) = Count(Message(ego, contact)) / Time(ego / contact)$

Having established "normal behavior", a system can verify when ongoing behavior falls outside of these parameters. When communication levels (in terms of number of messages exchanged) rise above average, the system considers this to be an active subnet, where ego is engaged in collaboration. And the activity function is defined:

$TimeBetweenMsgs(ego, contact) = Time(Msg_i(ego, contact) - Msg_{i-1}(ego, contact))$
 $Activd(contact) = \text{if } (TimeBetweenMsgs(ego, contact) < AvgActivity(ego, contact))$
 $Activd(contact) = True$
else
 $Activd(contact) = False$

Note that, when there is no prior activity data for a contact, the method substitutes $AvgActivity(ego, contact)$ for $AvgActivity(ego)$.

Alters or groups receiving more attention than usual are the ones a user is actively interested in. A series of replies in a period of time shorter than the average, or an intensification of exchanges (i.e., more messages are being exchanged than usual) indicate ongoing collaboration. Groups in which the user is very active are of more interest, with activity providing an indication of the focus of attention.

The first level prune tries to answer the following question: given the universe of user acquaintances, which ones are the user collaborating with? The emphasis is on finding active collaboration spheres, in order to seek information on activities that relate to joint work.

There are certain rhythms to work, and activity within a group changes according to the need. Thus, a group may be very active for a period of time and slow down after a certain point (e.g., project completion or reaching a milestone). This means the system must constantly check for the formation of new groups and changes in activity patterns. This activity is reflected in the levels of interaction among group members, which means that a system must actively capture and process data to maintain an up-to-date model of a user's collaborations.

It should be noted that a user's collaboration spheres are not defined only by a group of individuals, but also by the shared work that brings them together. This means that content analysis is needed to disambiguate interactions, defining the social worlds as a set of individuals with a shared theme, goal or project. It can thus be used to determine what activities relate to each social world. Resources accompanying interactions are automatically allocated to the collaboration sphere, as they include conscious activity and selection on the part of the user. The final thus is:

Determination of Resources and Themes

Resources are attached to collaboration spheres with strength proportional to their level of correlation to the sphere, as measured directly by a TFIDF vector comparison. After determining which parts of a user's network are active, we turn to the determination of which topics they concern and which resources and activities are related to them. This is accomplished through an analysis of the contents of these interactions, and subsequent matching of these to resources being manipulated by group participants.

The determination of which activities are related to a collaboration sphere is done through content analysis and matching of interactions to ongoing activities. Our proposal is to apply standard keyword extraction and indexing techniques to elicit message and activity themes. Ongoing activities with contents that match the contents of ongoing interactions (determined in the previous step) are allocated to their respective collaboration spheres.

For textual analysis and matching, keyword vectors are built to represent interactions and tasks. These are constructed using the TFIDF algorithm, described in chapter 2, which generates weighed keyword vectors given textual documents, and compared using the vector space model, using the cosine measure of proximity. Given that most of the activities under consideration are information processing tasks that involve a large amount of textual information (word processing, website surfing and searching, chat, etc.), this should be a feasible approach, which should elicit activities that are related to previous conversations. Being established methods for information retrieval and matching, TFIDF and cosine measures have been extensively applied and tested, with good results.

Information retrieval methods yield a score that is used to establish how well matched to a collaboration sphere a resources is. It should be noted that Attachments and links included in messages are treated differently: as they involve an active selection and distribution of content, they contain “user intelligence” and are immediately allocated to the respective collaboration sphere, which means they will also be taken into account when indexing and retrieving information in the next step.

Textual files (pdfs, word documents, web pages) are processed for keywords in the same manner as messages, and compared to ongoing interactions, yielding a relation between individual tasks and the social worlds a user is inserted in. The higher the similarity score, the more related the resource is to the collaboration sphere. Hence, allocation to collaboration spheres includes a “relatedness factor”, which is assigned according to how well matched the resource is to the collaboration sphere. In the case of attachments and links, these are attached with a strength of 1. Other resources are assigned factors according to their similarity scores. Thus, links are established according to similarity rating results. An active system must keep track of its user activities, periodically recalculating activity and message matches.

In this thesis, we are only looking into resource allocation to the appropriate collaboration spheres. Lower level events (e.g., editing a paragraph of a file) are left for further work. More fine grained control, analysis and display would enable the user to “drill down” to obtain more information. However, at this stage, we are concerned only with the constitution and allocation of people and resources to collaboration spheres, as these must first be discovered in order to enable further elicitation and analysis of events.

VII. Analysis

In this chapter we present a few analyses undertaken to assess the viability of the described methods. A full validation would require a long term study involving several users and activities, and would be hard to undertake in the available timeframe. Therefore, we designed a few analyses to see how well the method described would perform.

Analysis of Interactions to Determine Active Collaboration Spheres

Prior to analysis, we collected all the threads, saving individual messages in a database, extracting sender, thread titles, dates and attachments. Most of the attachments were word or excel documents, so they were manually converted to pdfs for ease of processing. The data set consists of 1023 messages, subdivided into 112 threads, sent over a period of roughly 24 months. The thread type classification seen in Chapter 3 was not used in this analysis, other than noting technical discussion threads. Cross referencing results with classification is left for another analysis. This dataset correspond to the forum messages of the European project described in Chapter 4. These are confidential data, so all names and affiliations have been omitted.

Analysis I: Interaction Patterns in European Project Dataset

The first step of our method analyzes interaction patterns to elicit active collaboration spheres. To verify this step, we took the data from the forum and checked on changing interaction patterns, to elicit different work groups within the larger groups. A few adjustments were made because the data came from a group forum as opposed to an individual email log and the set of acquaintances was already limited. As participants in the forum are all somehow related to the project at hand, it was difficult to determine if they could be left out through an analysis of whether messages were sent to them or not, as messages posted are open to all. However, only a few responded to each posting, so we count these as the actual recipients of the message, as their attention was drawn to the subject. Posting to a thread can be interpreted as interest (as it represents a commitment of attention and time), and subgroups were identified in this manner. Posters to the same thread formed subgroups dealing with particular issues. This also means that threads to which nobody responded either failed to capture the interest of the group or did not require a reply (e.g., announcement or FYI messages).

Messages were counted in the following manner: each message sent to a thread was counted as one message sent to all others that engaged on that thread, which means each member sent at least one message to all other members, given that some of the organizational threads elicited a high response ratio. At this stage, our hypothesis was:

H1: if message activity is representative of collaboration, then technical partners who collaborate more closely with B will have higher interaction levels (number of messages sent and received).

It is possible to observe sub group formation, as a few participants band together at certain points in time. It is also possible to observe the different interaction patterns, which reflect the different discussion participants engage in. Figure VII-1 shows a graph of interactions for the whole group throughout the period in question (almost two years). As one can notice that there are periods of high activity interspersed with others of little or no activity (presumably, holidays or vacation time, as evidenced by the august periods). We point this out because the individual interaction graphs in Figure VII-4 were created based on periods of greater activity, in an effort to avoid long periods of inactivity that render the graphs difficult to read.

In the interaction graphs presented later, the egocentric viewpoint is used, and ties to others are constructed according to the number of messages exchanged. We selected participant B as ego, as he had been a fairly active participant, with 102 messages sent to peers over the period analyzed, participating in 44 threads. He was also a technical peer, which we felt would yield more interesting results.

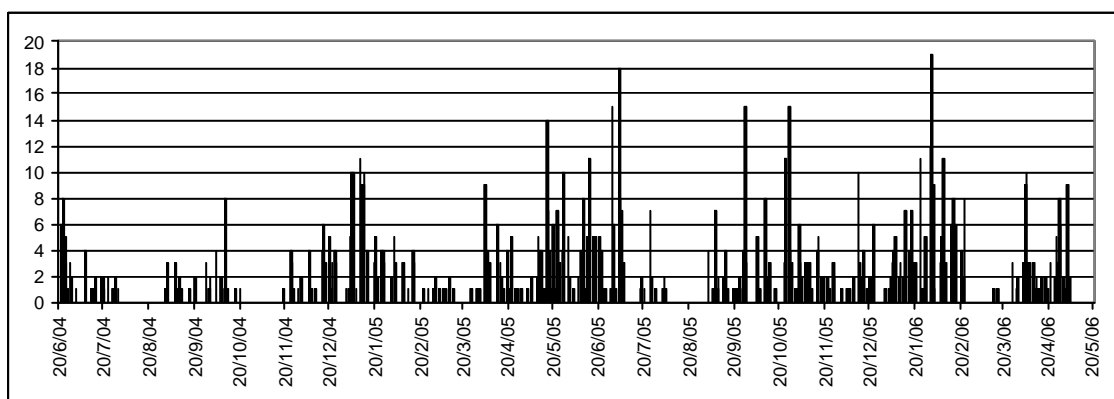


Figure VII-1: Total postings per day

Looking at B's activity pattern within the group (Figure VII-2), we can see there are moments of intense activity interspersed with others of lower activity. Interaction

patterns across time (messages sent to others), show the different rhythms in which participant B engages others in discussion.

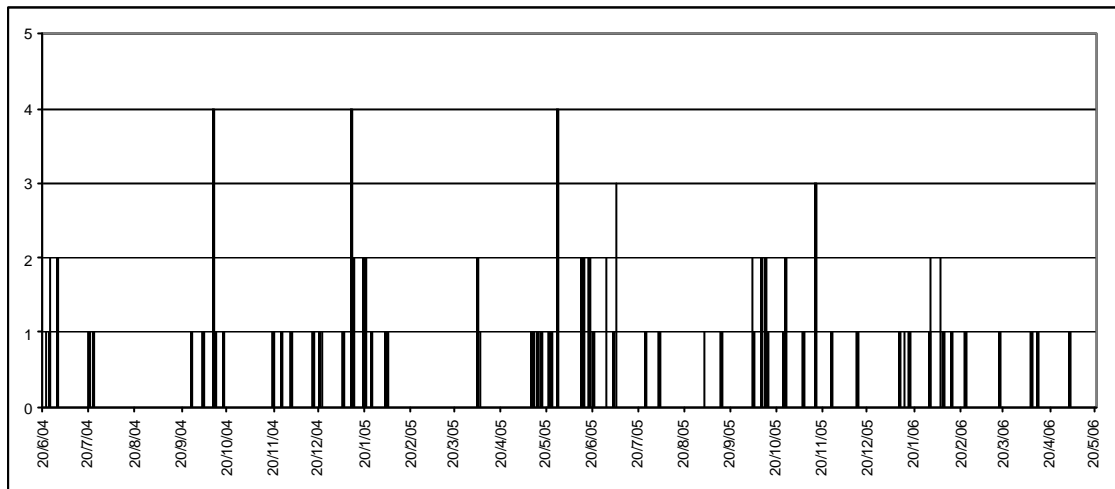


Figure VII-2: Messages sent by B per day

An analysis of messages sent and received from participant B to other participants shows that message frequency varies greatly among the different participants. Not surprisingly, the peers with whom B most frequently interact are other technical partners. By using message exchange as an indicator, we expect to arrive at the group of technical collaborators (A, C, D, F, J, P).

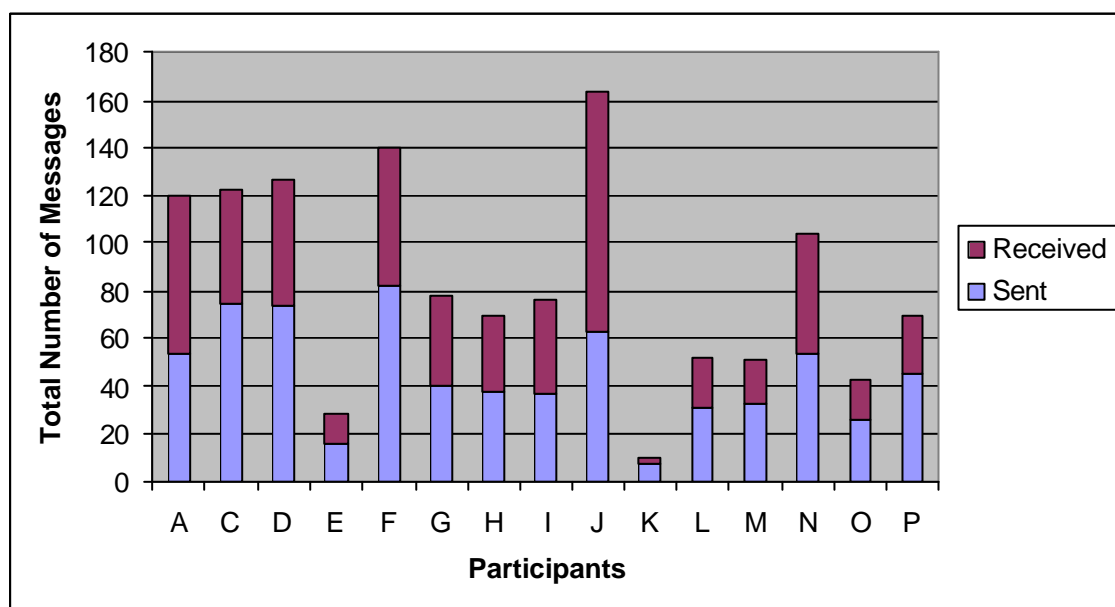
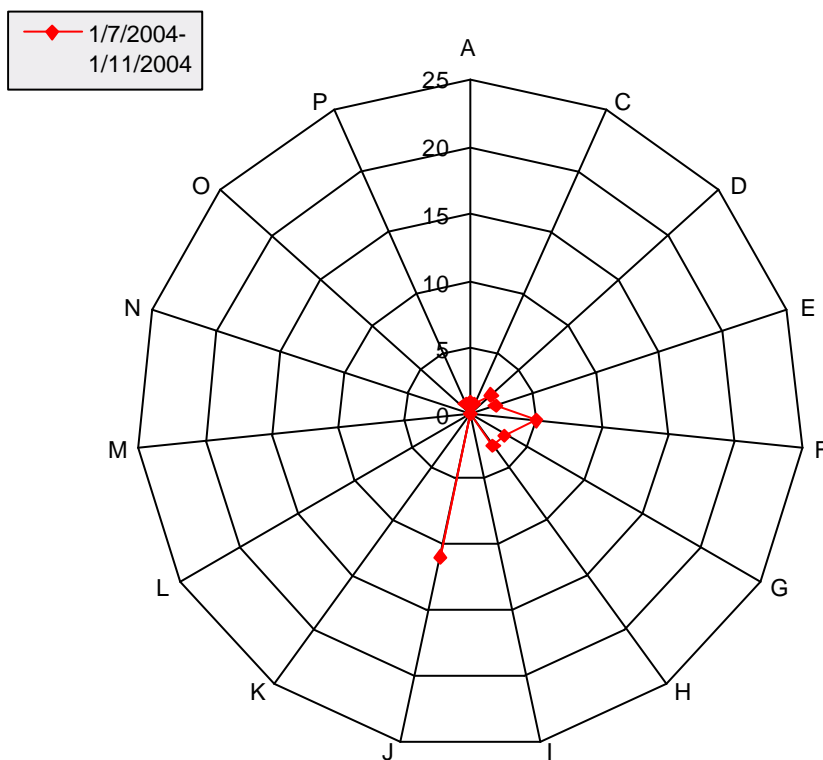


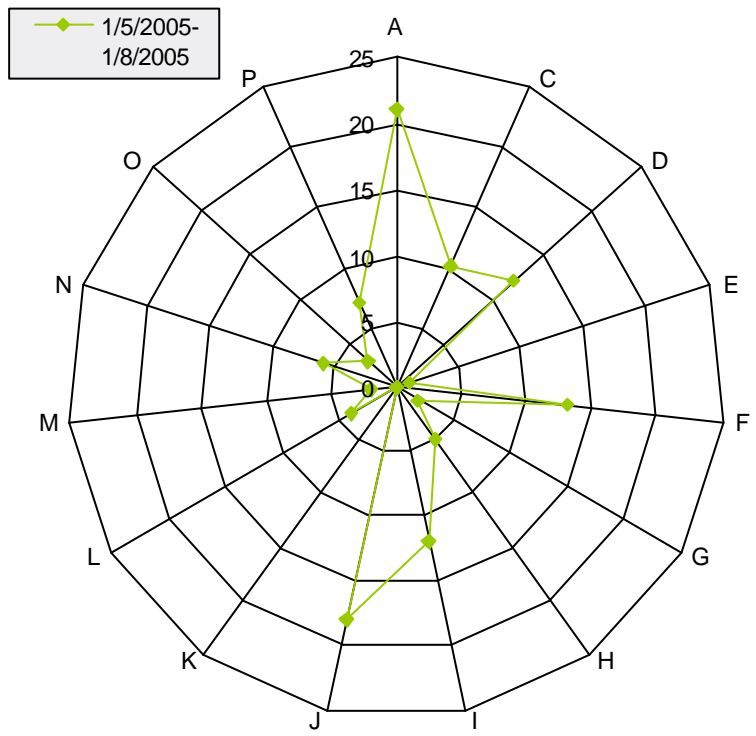
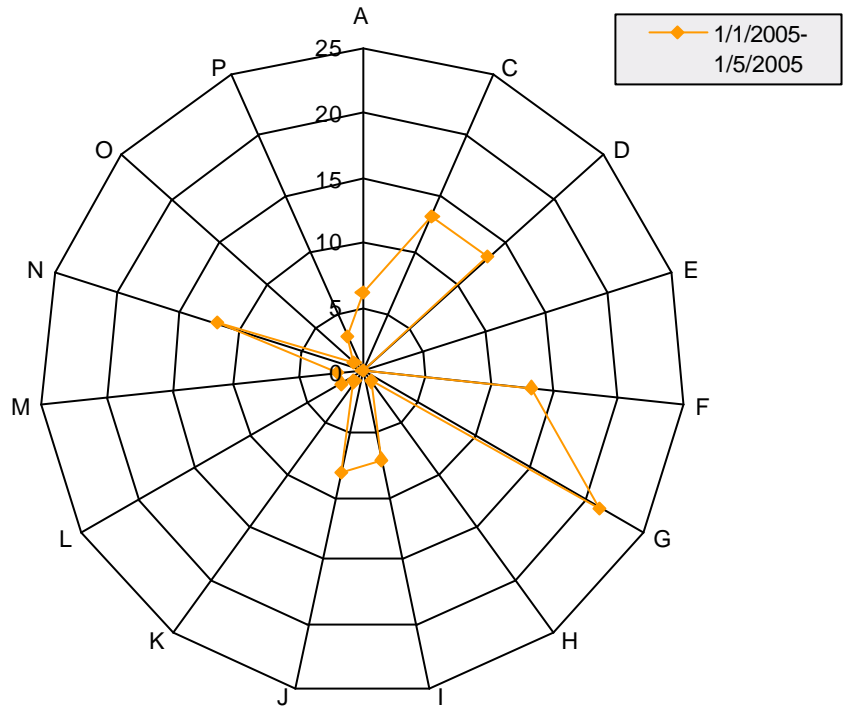
Figure VII-3: Messages exchanges between member B and other members of the group

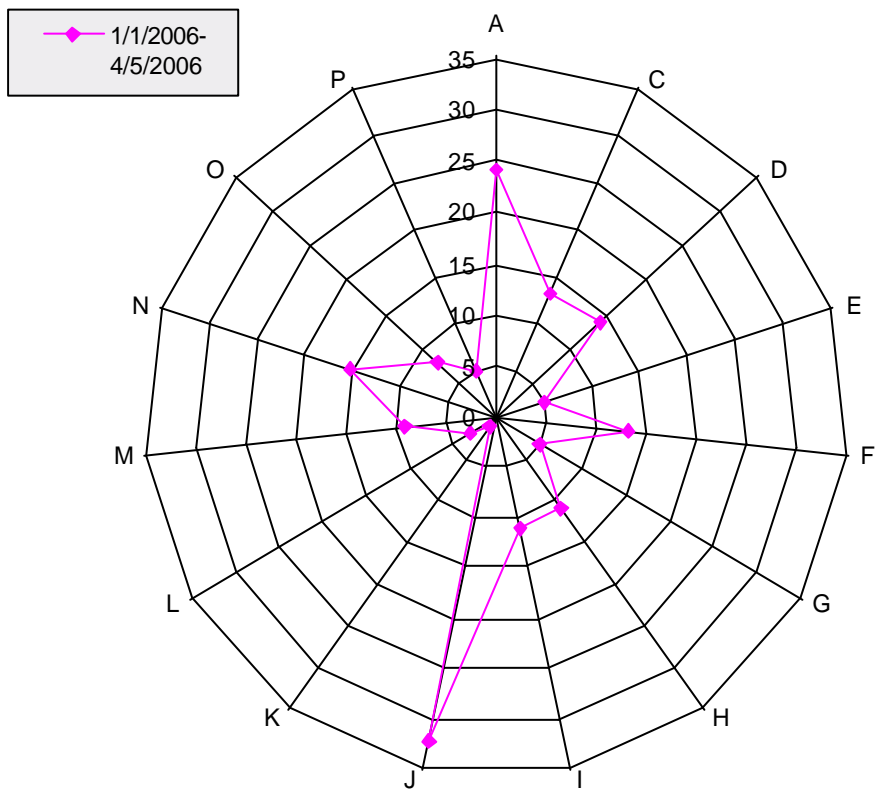
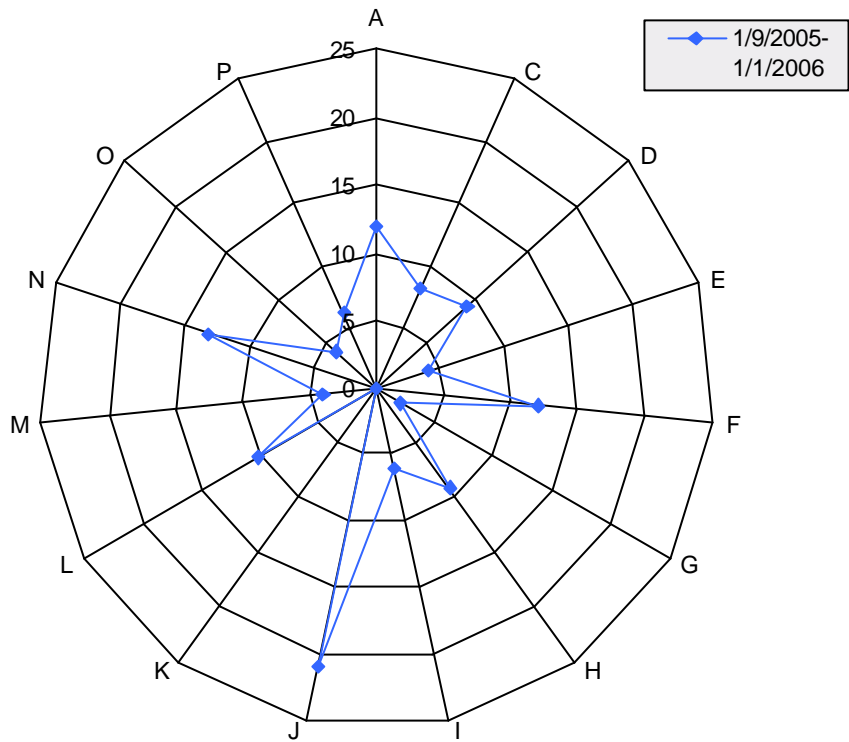
Upon verification, technical partners did score higher in the message exchange, which confirms our hypothesis. The unexpected low score was member P: he had a poor record of participation in the group as a whole (participation in only 22 threads, most

for scheduling and confirmation of attendance). An unexpected high score was member N, who was, in fact, one of the “clients” for the application being constructed by B and collaborators. Closer collaboration with N was necessary as they provided the case under study, furnishing data and requirements for the modules under development. This provides evidence that collaboration equates with interaction.

Figure VII-4 shows interaction patterns for user B in different periods, with the different alters, in terms of messages sent per day (how much attention B is dedicating to his alters), to illustrate the activation of different ties. The last graph is an overlaid view of the previous five. Observing these graphs, it is possible to perceive relationships becoming active through communication, which represent shared interest or work. Different partners become active at different moments, showing that, even in the small group under analysis, members organize into subgroups as necessary.







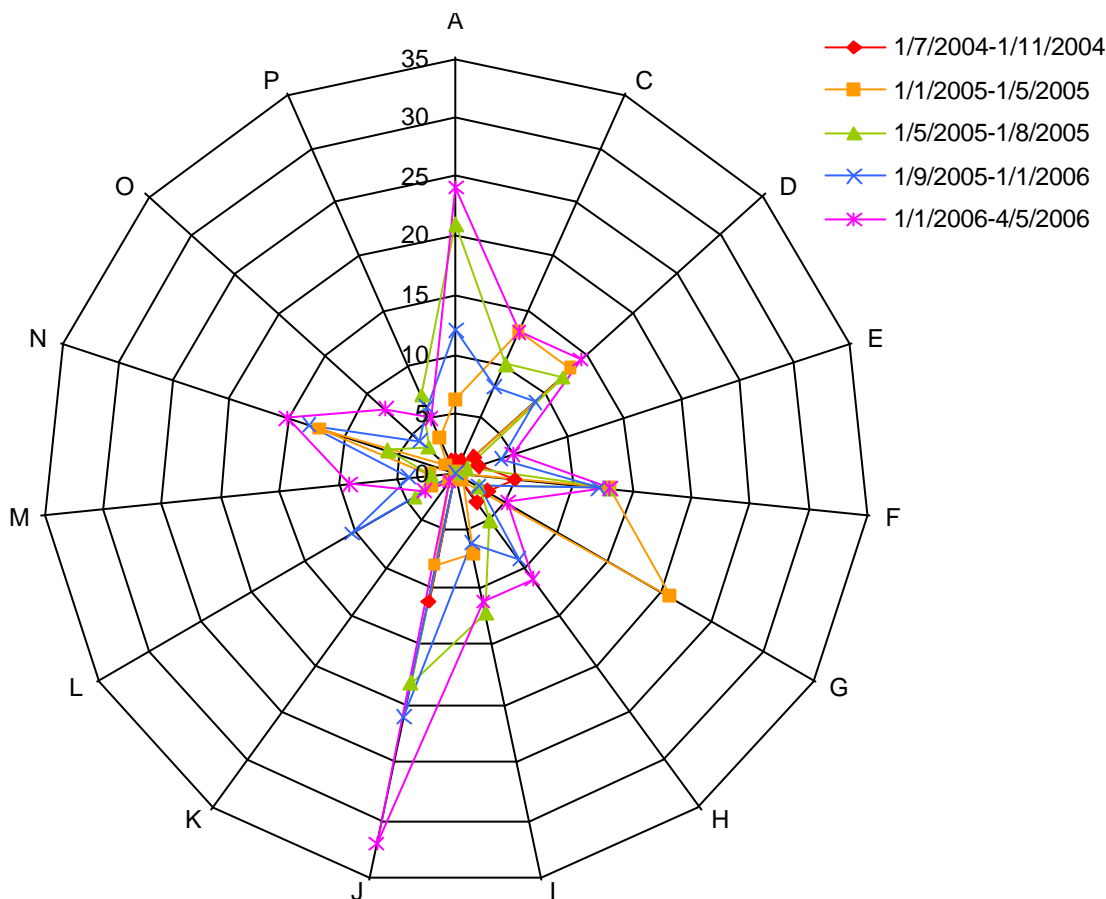


Figure VII-4: Number of messages sent per day from B to other members in different periods

Analysis II: Interaction Patterns in Email Dataset

For a second analysis, a visualization tool was built to enable further verification of some of the assumptions involved and to obtain user feedback. This tool performs structural email parsing, extracting senders and recipients and builds sociograms, according to the methods described in the previous section. Ties between users are qualified in terms of message count (messages sent and received), but no content analysis is performed.

An interface displays the sociograms, with the possibility of exploration by slicing the data into different temporal intervals and sources, as seen in Figure VII-5. It implements a spring-embedded graph layout, using the Fruchterman-Rheingold force model (Fruchterman and Reingold, 1991). This algorithm treats a graph as a set of nodes that repel each other, but are connected by springs which attract the nodes. It generates a mapping that reflects node proximity while attempting to minimize line

crossings. The visualization was built using the Java language and the JUNG library for graph construction and display. It should be noted that this is an analysis interface, and not intended as an interface for a final system. This was used to interview users regarding their social worlds and how they relate to ongoing work and awareness needs, to obtain a preliminary analysis and get user feedback before proceeding with system implementation. We interviewed 4 users regarding the email patterns that showed on the visualization.

To perform a verification of the first step (determination of collaborators), we worked with three assumptions:

- (1) that collaboration spheres are reflected in email;
- (2) that participation in a collaboration sphere indicates interest;
- (3) that activity levels for a collaboration sphere change over time.

Users' email logs were processed, and these users were interviewed with their respective sociograms at hand for inspection. Our users are all heavy email users with several thousand email messages in their mailboxes. One is a university professor, and the three others are students at the same university. They often communicate with students and colleagues via email, frequently exchanging files with their collaborators and recommendations in their messages. The majority of their activities consists of writing documents (papers, reports, grant proposals), studying (reading documents of webpages) or coding. We explored their sociograms with them, slicing the data in different ways. We asked users if:

1. the cliques they identified in their sociograms were related to projects or other collaborations going on at the time;
2. different cliques became active when the temporal range was changed;
3. patterns of message exchange reflected projects;
4. the social worlds in which the user had not participated (other than as an "observer") were of interest as far as peer task awareness.

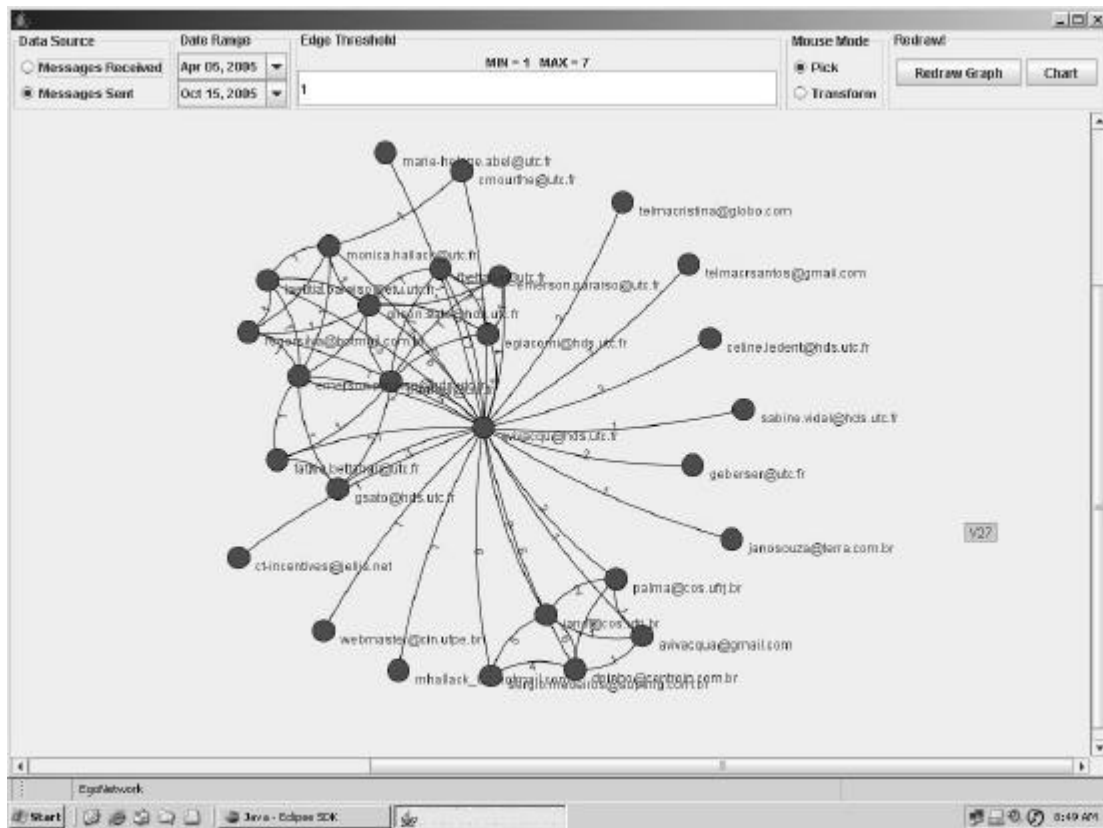


Figure VII-5: Visualization screenshot, several cliques are visible.

When asked, all our users were capable of relating existing cliques that showed up in their sociograms with ongoing collaboration, even if historic. However, not all of these social worlds were related to ongoing work. In certain situations, cliques represented groups with non work-related shared context (e.g., students in the same department), who were not in direct collaboration. This observation confirms that social worlds do map to cliques, but it doesn't follow that all these represent joint work.

While further investigation is necessary to determine how to differentiate between work and non-work interactions, it should be noted that it is hard to dissociate the personal from the work aspects when dealing with interpersonal relationships. Many relationships have a dual nature, being partially work related and partially personal, and these two facets have an effect on each other.

Any changes in time slots brought different groups to the forefront, displaying changes in activity levels. Social worlds did become more or less active according to group needs, and the user's participation in them also varied in intensity, indicating interest. Slicing the data into shorter time periods considerably reduced the number of messages and the corresponding visualization, making it easier to identify subgroups. This confirmed that activity levels are reflected on email. Inspecting the temporal graph

(seen in Figure VII-6), where time is sliced into daily email exchanges, changes in the interaction pattern could be seen. A dormant relationship suddenly sprang to life, with emails being exchanged daily (sometimes several messages a day, depending on the urgency), and then died out when deadlines are reached. This confirms the assumption that temporal patterns are detectable through email. Changes were often quite abrupt, going from no interaction to four messages a day overnight. While this was expected, we had also hoped to see softer patterns, where interactions would gradually increase with time.



Figure VII-6: Time based interaction graph

All users confirmed that social worlds in which they did not actively participate were not of enough interest to warrant the creation of a collaboration sphere. An analysis of available data revealed that most of these were announcements or mailing lists that were not related to any projects. Project-based mailing lists usually had users' interest, and they engaged in active participation. Users wanted to be aware of their closer, more immediate collaborators and activities related to pressing projects, where there was a lot of coordination to be done, and had no desire to be aware of everybody's work, although in some cases they would like to remain superficially aware of what was going on in a group. This indicates that, for task awareness purposes, incoming threads in which the user has not participated can be left out. In computational terms, this significantly reduces graph size, and, consequently, memory needed and processing time. As an illustration, in one case this meant well over 70% of the total contacts found in the mailbox. Large amounts of spam and announcement mailing lists account for most of the emails left out in this particular case. However, some false positives were also observed, where alters who had multiple emails had some of them removed because ego used one address only. This difficulty can be overcome through the creation of aliases that bundle multiple emails together to represent one alter, so as not to lose relevant messages.

Within the emails, there were several instances of project-related social worlds, usually qualified by intense interaction in a shorter period of time (weeks or a few months). This suggests a way of more effectively picking activity-related groups. Further data inspection showed that structure alone was sometimes not sufficient to tease activity apart, especially when there were overlapping social worlds. These needed to be qualified according to the interaction themes, so that they could be effectively differentiated. There were a few overlapping social worlds, including temporal overlaps, where a group worked together on more than one project at the same time. Within our data sets, there were also a few social worlds embedded in other social worlds.

Number of messages sent by ego proved to be a reasonable qualifier for ties, unlike the number of messages sent by an alter. Some alters sent over 200 messages over a 6-month period and were neither collaborators nor of any interest to the user. A user's outgoing messages reflected user's participation and interest in social worlds more accurately. Engaging in conversations involves an investment of time and effort that indicates a certain level of commitment to the group. Accordingly, the algorithms process a user's outgoing messages first, determining the relevant alters and then fits incoming messages in with these.

Analysis of Content to Elicit Resources

Analysis III: Content Matches in EU Project Dataset

To further verify the applicability of our method, we analyzed the use of content analysis to tie interactions and resources through the application of information retrieval techniques to measure the correlation between documents. These similarity scores are used to associate resources to message exchanges, determining how well inserted into a collaboration sphere a resource is. The primary measurement is the level of correlation between two distinct documents: the higher the similarity yielded by a TFIDF calculation (described in chapter 4), the more correlated they will be. To that end, we devised a series of similarity matches, to verify how well different variations would work. The file set contained files sent as attachments, files posted to the forum and external files, inserted to see whether false matches would be found. Indexes were created for messages, threads and files, and similarity was calculated between them.

Content matches were performed by taking messages and indexing them, and then using these to retrieve documents manipulated by the user. We indexed full threads, as these were the focus of activity. After removing stopwords and stems, as described previously, we extracted keywords for each of the threads. Keywords for 2 sample threads are shown in Table VII-1, to illustrate the diversity of themes involved.

Table VII-1: Sample keywords for threads

Thread	Most Frequent Keywords (with frequencies)
8	activebpel: 5; code 15; contribution 7; documentation 11; libraries 12, module 11
11	amount 5; bank 5; commission 3; contract 3; distribution 4; file 5
83	conference 3; discussion 3; engineering 3; government 5; impact 3; partners 6; presentation 14

We matched messages and thread keywords to pdf documents, to see whether related documents would be retrieved. The top 10 matches for each thread were taken for analysis. Table VII-2 shows a sample thread’s theme and a description of the documents obtained with the match score, which would be used as the indicator of pertinence of an artifact to a collaboration sphere. This score is calculated by taking the keywords found in the threads and searching for them in the documents in question.

Table VII-2: Matches for one thread (top 10 shown)

Thread theme: “open source issues and next review meeting”	Score
1. details of one partner F’s pilot applications	0.495258
2. developer’s second chat log	0.477502
3. whitepaper describing pilot applications	0.472957
4. pilot ontology and reference - partner A	0.435104
5. developer’s second chat log	0.414138
6. definition of metrics and measurement baselines	0.401056
7. storage and data access details	0.369296
8. storage and data access details, revised	0.332887
9. pilot applications - partner P	0.326288
10. draft version of project deliverable	0.305564

One of the issues is finding appropriate cutoff points, where documents are no longer relevant to the collaboration sphere and can be discarded. Technical discussion threads yielded the best results, and had high proximity scores with technical documents delivered and with the logs of two online discussions.

The first analysis we conducted was between individual messages and files. An analysis of the similarity values between them yielded poor results. Scores were low overall (see Figure VII-7). Almost all of the higher scores were false positives, where a

message is linked to resources unrelated to its content. Most of these occurred because messages were short and contained proper names that also figured in the documents analyzed. These became the basis for the match, as there were few words in the messages. A few notable exceptions occurred: high strength correlations were found between a message that contained agenda items for a meeting, and documents with agenda items (for that meeting and other meetings). Another strong match was between a long, highly technical message about ontologies involved in the solution and documents which discussed the ontology construction lifecycle.

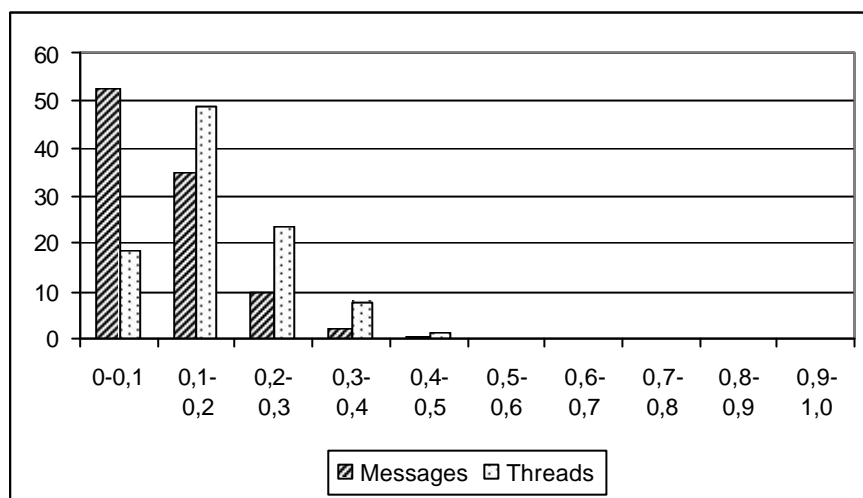


Figure VII-7: Similarity rating distribution for 10 ranges of values, for messages and threads, as a percentage of total ratings.

Threads also scored higher than isolated messages in the resource match. This is understandable, as the series of messages in the thread provides extra content and establishes a discussion. As we noticed that the shorter messages yielded unreliable replies, we selected the 10 longest messages to analyze regarding precision, in our case an analysis of the correlation of the documents recalled to the message in question. The longest messages contained technical details of the project: some were specifically technical discussions, and others were documentation reviews questioning technical aspects. In all of these cases, the top three matches were relevant to the contents of the message, despite their apparently low similarity ratings (all around 0,3). A similar procedure was performed for threads, with similar results: among the top threads, similarity ratings were about 0,4, going over 0,5 in some cases. On a related note, very few files unrelated to the project were brought up by these matches. For instance, there were only 25 outside files within the 1120 results for resources related to threads (about 2%). It should be noted that some of the external files actually did bear some relation to

ongoing discussions, as they related to ontologies, agents and semantics. Completely unrelated resources (e.g., Common LISP Reference Manual) did not figure in the results.

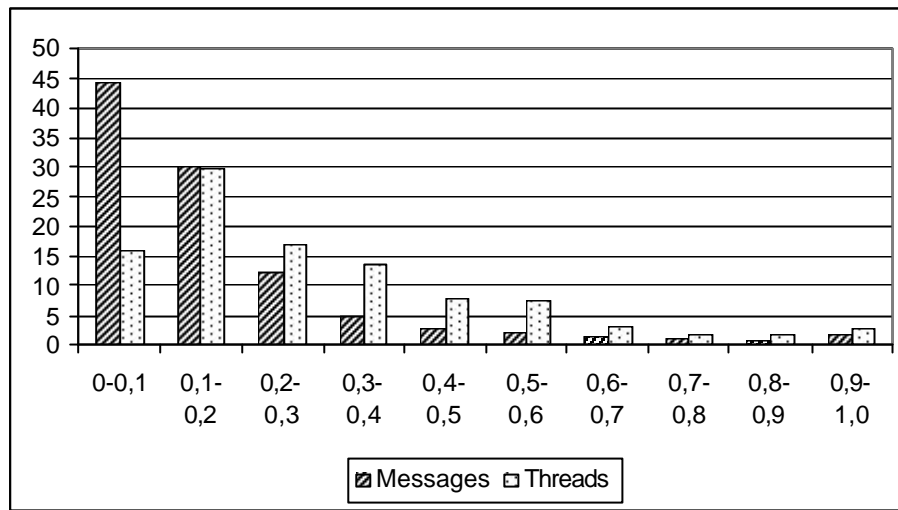


Figure VII-8: Similarity rating distribution for 10 ranges of values, for messages and threads (with attachments included), as a percentage of total ratings.

From our prior study of the journal paper group, we came to believe that attachments and links would provide additional content, which might be more related to the project than messages themselves, which is why we assign them to collaboration spheres automatically. To check on this assumption, we created indexes that incorporated the contents of attachments to the body of messages. We expect these will provide better matches than messages by themselves. We would like to test the following assumption: *if attachments contain more relevant content than messages, then attachments will have higher similarity scores when compared to files than messages by themselves.*

This can be verified through a comparison of scores between messages where the attachment content was considered and messages without the attachments. Figure VII-8 shows a graph of these distributions, for messages and threads. It is easy to see that ratings are greater when attachments are considered than not, as attachments provide added content with which to compare documents. This corroborates our hypothesis that attachments are valuable source of project related data. In a fashion, by using links and attachments sent with messages, we are taking into account users' recommendations, as someone has already gone over these resources and judged them

useful to the group. Thus, linking an attachment to a collaboration sphere is in fact a correct operation.

Our last verification worked the other way around: we selected 30 files (10 attachments, 10 deliverables and 10 outside files) and attempted to match them to threads, to see whether it would be possible to link an artifact with interactions that surround it. Again, we compared these files to threads and found fairly low ratings (see Figure VII-9). However, there was a distinction in ratings between outside files and project-related files (whether attachments or not): outside files never rated above 0,3, while other files could go up to 0,75. With the inclusion of attachments, ratings went up considerably, further separating project files from unrelated files. This is an important point, as it can help establish a difference between external and internal resources.

We consider results to be good on this account. This analysis provides evidence that it is possible to match resources to social contexts, and the top 3 matches were all relevant matches, in that they were actually related to the document in question (high precision). It should be noted that the best scores came from technical discussions, which were closely related to documents describing aspects of the projects, even without the attachments for added context.

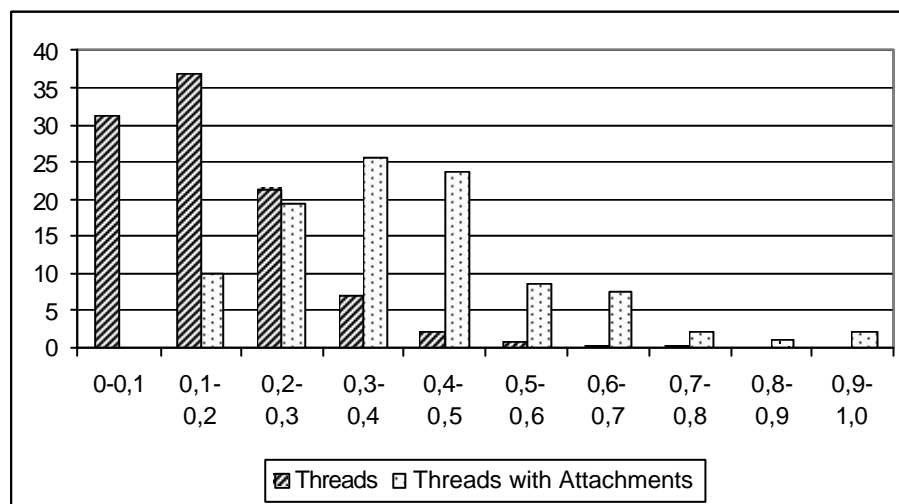


Figure VII-9: Similarity rating distribution for 10 ranges of values, for matches between resources (files) and threads (with and without attachments), as a percentage of total ratings.

We were surprised to find that ratings were much lower than we expected. One of the main problems is message length, as there were many very short messages, mainly in scheduling or meeting preparation threads (e.g., “so-and-so will be present to represent company X” or “here is company Y’s contribution to the report” or “please

send in the quarterly management report”). The large number of planning and scheduling messages may account for the poor results, as meeting scheduling and contribution requests usually generate only short messages that have little to do with the project themes. However, some of the contributions were sent in as attachments, in which cases they contribute to the definition of themes in the collaboration sphere.

To measure how well the matches worked, we manually analyzed the top 3 files retrieved for each match, to see whether they were effectively related to the message or thread in question. Matches were rated 0-3, with 0 being unrelated, 1 for related, 2 for strongly related and 3 for those cases when attachments were correctly assigned to their respective threads or messages. While this would probably have had more accurate results had it been scored by the owners of the data, as they were technical documents, the researcher was able to score them without any problem.

We randomly selected 10 messages and verified how well they connected to files: 4 out of 30 files scored related and only 1 was strongly related (an IRC chat log was matched to a technical discussion). One message actually led to its attachment. Messages by themselves were short and therefore did not provide good matches. The same procedure was adopted for threads, and scores were better. Out of 10 randomly threads (30 matches), one linked to one of its attached files (the thread contained the message mentioned before), and there were 5 files strongly related to the threads and 1 related one. The remaining files were unrelated. This seems to indicate that, barring specific situations (e.g., technical messages) message content does not correlate well to documents. This had already been gleaned from the analysis of the messages in the journal case study, which provided initial indication that attachments were better content providers than messages themselves.

Another surprise was the number of matches due to proper names: quite a few of the top rated matches were short messages that contained proper names (e.g., “participants A and B will be present representing company X”) that were matched to for instance, activity reports that contained these people’s names. While not a particularly useful match, it does bring up a possibility, of giving special treatment to proper names. These could be extracted from sender fields or signatures, and be explicitly searched for in messages or documents.

We somewhat misjudged the group’s interaction style: we thought many project discussions would happen online, but these were the minority (although when they did

happen, they provided excellent content for an information based match). At the face to face meeting, participants took the time to discuss technical details and make decisions, and tried to achieve a division of labor to minimize interaction in between meetings. This is consistent with observation by Olson and Teasley (1996) that distributed groups display a tendency to change the organizational structure and task distribution in order to reduce coupling of non-located members, thus reducing coordination needs.

Informal interviews with group members revealed that they formed smaller subgroups that interacted more frequently, but outside the forum (via email mostly, but also IRC and on the phone when necessary), and this data was lost for the purposes of this study. They also reported visiting each other in between group meetings when necessary, and in fact stayed on after the meeting to work together. The range of media adopted could also pose a problem, as it would be hard to cover all different interaction channels. However, this also brings up new opportunities, as each channel has different affordances and will be chosen depending on specific needs and desires. Given the user's choice, one could infer whether the message was urgent, for instance.

Putting together the two parts of the approach, the active groups, as indicated in the first stage of the analysis would have their content transformed into keyword vectors and these could then be matched to documents belonging to the user or to resources in use (webpages). Once the collaboration sphere was constructed, information could then be distributed as necessary. For instance, when the thread pictured above became active, partner F would be informed of the existing relationship to others when he edited his pilot details document. Alternatively, participating partners would be notified when one of them edited the whitepaper describing the pilot applications.

VIII. Discussion and Future Work

This thesis presented an approach for information dissemination, through the discovery of collaboration spheres, and an analysis of this method. The approach presented provides a way to explore information needs to enable self organization in loosely structured groups.

This is a step towards the discovery and description of collaboration spheres within which a user operates. This information could be useful in assisting the user automatically configure groups, remembering who should receive a certain of information, who is impacted by work on a given file, who else has worked on or accessed a particular resource, what others are doing that the user should know about or what the user is doing that should be relayed to others.

We can foresee a number of scenarios where collaboration spheres can be useful: for instance, as a user interacts, resources are presented that relate to the collaboration sphere in question. Another possibility is to present the user with information on peers that relate to the ongoing activity, so the user will know how this impacts on others (a variation on this is to have systems automatically disseminate information to provide awareness, which introduces a privacy problem). A resource could also be tagged with the social context it pertains to, so that it goes with it when it is sent or received to others. This would benefit, for instance, a newcomer to the project, who would better understand the multiple roles and individuals that were involved in constructing this resource, not to mention arrangements, deadlines and related information. The important part is tying individually produced artifacts with their social and collaborative contexts.

This technique could be used to construct systems to support networked, distributed, loosely structured work, by assisting individuals multitask between different groups and maintain their networks alive through communication. Understanding and adjusting to unforeseen work events is an everyday task, which could be better supported through automatic inference systems. This involves not only maintaining awareness of the group, but also understanding how people interact during the construction of artifacts (maintaining both a present and a historic perspective), and, conversely, what artifacts are involved in each relationship.

Nardi and colleagues (2002) state that two processes are fundamental in networked work: remembering (people, their interests, ongoing collaborations, etc.) and communicating (with the peers, whether involved in collaboration or not). Collaboration Spheres could help users remember their ongoing projects by relating individual and group work. Gonzalés and Mark (2005) also state that a common problem when multitasking among several collaborations is remembering arrangements made and not letting anything “fall through the cracks”. Provision of information relating the individual to the ongoing collaborations helps users keep in step with the different projects, while also keeping track of others’ work and discovering opportunities for interaction and discussion. Additionally, by comparing interactions and elements from outside the collaboration spheres, it may be possible to find new resources that would be of interest.

With this thesis, we contribute to the Computer Supported Cooperative Work domain, with the introduction of a concept (Collaboration Sphere) to bind together individual workspaces. This concept underlies the creation of shared spaces, bridging individuals who work alone and exchange messages to collaborate. We also present a method to constitute these working spheres, finding people and resources that fit together within a collaborative context. This method is an extension of Rodden’s (1996) model of awareness, as it determines a user’s focus of attention to distribute information. It could be used to determine collaborations within a user’s focus of attention, to locate related documents or ongoing work related to these and spread activity information about these, enabling users to become more ware of each other’s activities. Results from this research and related work conducted during the development of this thesis have been published in a number of refereed publications, listed in Appendix A.

Limitations

Our findings are applicable to any situation where it would be beneficial to group together individuals with shared activities and their resources, based on individual activity, eliciting informal or ad-hoc groups. One example of a system that might benefit from using an approach such as this is the ActivityExplorer (GEYER et al, 2006), which requires that individuals manually allocate elements (interactions, artifacts, etc.) to their ongoing activities.

We expect this method will not provide good matches for activities such as coding, despite the fact that it involves document manipulation. This is because coding involves a different language, which is usually not present in discussion (save for situations when specific coding problems are being discussed, and there were none present in our sample message set). We look at this, however, as an opportunity for building the algorithms into software development platforms, in such a way as to integrate systems such as Eclipse and CVS with communication between developers.

The approach encounters a serious limitation in groups that are even partially collocated: we discovered early on that it did not work well in settings where participants have the opportunity to meet in person, as they usually prefer to conduct work based discussions in person. An examination of the message logs from the journal case showed that most messages contained little or no content about the project itself, consisting mostly of document handoffs and meeting scheduling (about 70%, 22% were technical discussions and the rest were notifications of meetings that had happened). This indicates that this approach will work better for groups where the majority of discussions happens through electronic means.

Our approach works for groups that interact, coordinate and exchange ideas using text based media through the computer (e.g., email, messenger or discussion fora). Arguably, it does not work with video, audio or other media, where other methods of analysis would be necessary. It can only provide matches for activities such as document composition, reading or browsing the web, where textual documents are manipulated and can be correlated to discussion.

One of the difficulties encountered when attempting to perform content analysis was language: in our initial message sets, all of the discussion was undertaken in Portuguese, whereas documents were written in English. This language difference may prove to be a great barrier for any system adopting this strategy. The approach also suffers from a bootstrap problem: to establish interaction patterns, it needs prior records of interaction, otherwise it might take time to learn new the user's work patterns.

Comparison with other methods

In this section we present a brief comparison with a few related systems. These are systems to distribute awareness-related information or help organize work into activities. Although not exactly the same as our proposal, they share similarities which

is why we list them here. A comparative table is shown in Table VIII-1. The following systems were selected (all have been described in Chapter 3):

- PIÑAS (MORAN et al, 2002) a framework to support potential and actual collaboration;
- Community Bar (MCEWAN, GREENBERG, 2005), a system that provides information through different media items;
- Activity Explorer (MULLER et al, 2004; GEYER et al, 2006) a system to support lightweight, informally structured opportunistic activities;
- ContactMap (NARDI et al, 2002), a system that enables users to visualize and organize their contact networks; and
- Elvin (FITZPATRICK et al, 2002), a notification framework that has been applied to awareness and interaction support.

Table VIII-1: Comparative analysis between collaboration spheres and other approaches

	PIÑAS	Community Bar	Activity Explorer	Contact Map	Elvin	Coll. Spheres
<i>Determination of focus</i>	Manual	Manual	Manual	Automatic (one time)	Manual (rules)	Automatic
<i>Consider interactions</i>	No	No	No	Yes	No	Yes
<i>Creation of collab spheres</i>	Manual	Manual	Manual	Manual	Manual	Automatic
<i>Joining collab. Spheres</i>	Manual	Manual	Manual	N/A	Manual	Automatic
<i>Event selection</i>	Automatic	Automatic	Automatic	N/A	Automatic (rules)	Automatic
<i>Adaptation to changes</i>	No	No	No	No	No	Yes
<i>Need central system/server</i>	No	No	Yes	Yes	Yes	No
<i>Privacy</i>	No	Yes	No	No	No	No

The table shows a comparison along 8 axis: determination of focus concerns the capacity of the software to actually determine where a user's attention is being directed; consider interactions asks whether the system takes interactions into account in its calculations; creation of collaboration spheres asks how collaboration spheres are

created, whether automatically or manually (by the user). Joining collaboration spheres concerns the way in which a user may join a collaboration sphere; event selection asks how events are selected for distribution; adaptation to changes concerns the systems' capacity to adapt to changes (new groups forming, old group dissolution, etc.). Need central server determines whether the system is centralized or not and privacy shows whether each system deals with privacy, if at all.

Future Work

Further work needs to go into completing the information in the collaboration sphere, regarding roles and hierarchies, task assignments, arrangements, deadlines, etc. Additional analyses of the different structures found in the networks might yield information on hierarchy and positioning within the group. Other research has indicated that there are interaction patterns that denote role attribution or the organization of a team, and email-based analysis can be used to elicit them (Fisher and Dourish, 2004). We expect different roles and organizational patterns will have different information needs (e.g., core vs. periphery members differ in terms of nature, quantity and depth of the information desired), but further research is needed to clarify this point, establishing the types and level of granularity if information distributed to each group member.

While a user will probably not be interested in keeping close track of the activities of members of larger groups with whom he or she has little in common, he or she may want to receive periodic summaries or reports on how work has been progressing within the group. Thus, awareness can be seen as a continuum, with needs tied to users' levels of participation in groups. The user might desire to have more or less information (regarding both depth and frequency) about others, depending on the level of involvement. Further research is needed into this topic, to better determine depth and level of detail of the information desired.

Further work can also go into better analysis techniques that take into account the semantics of the interactions. One of the possibilities is the creation of an ontology, relating it to terms to elicit details about the contracts established between individuals (for instance, promises to deliver part of the work, or the expectation that something will be ready by a certain date, existing reporting structures, etc.) Another possibility is to use speech act theory, as suggested in (COHEN et al, 2004), to classify messages. Another approach that could be incorporated is presented by Raposo and colleagues

(2004), who use conversation clichés to elicit commitments between partners. Clichés are state machines that control the dialog sequence. In well known domains, it may be possible to apply this technique to determine what contracts are being established between partners. Using an ontology or dictionary to categorize words or infer additional meaning from messages could help infer roles and contracts made between parties. Another possible application of collaboration spheres is to measure collaboration levels between individuals, through an analysis of their interactions and attachment exchange. The methods described could potentially be applied to monitor the health of the collaborative effort, and to suggest remedies when necessary.

As expected, the analysis of attachments indicated that these are good indicators of relevant content, as they are directly related to user decision making (selecting the resource and the recipients). Finding other instances of this type of situation (e.g. forwarding messages) would also help create a better definition of distribution methods. In fact, the low scores found when comparing messages to resources lead to possibilities of analyzing attachment text to elicit content and collaboration themes, leaving messages out.

By keeping track of the exchange of attachments, it would be possible to elicit a pattern of contributions to the artifact under construction, follow changes made by different group members (with a more in depth analysis of the document itself), and possibly even design a method to calculate authorship of the document by analyzing the amount of work each partner put in (a naturally controversial issue, which deserves more thought).

Cross-referencing statistical data with a classification of thread types (e.g., scheduling, technical discussion, meeting preparation) might also elicit different communication patterns for different types of threads. These could later be used to generate better expectations of response times and levels.

Another difficulty lies in separating personal from work-related interactions. However, this may turn out not to be a problem, as many relationships have a dual aspect to them, being simultaneously work related and personal. This means that a tie can be strengthened with non-work related information and this will likely benefit the individual at a later date. In fact, Nardi and colleagues (2002) point out that network maintenance activities often involve remembering birthdays, special occasions or personal preferences and sending personal cards, invitations or gifts to strengthen these

ties. While personal communication may lead to false positives, if a user is directing so much attention and time to collaborations of a personal nature, these should also be supported.

To develop a more complete mapping between interaction levels and awareness needs, other variables need to be taken into account, such as response time and message length. These may indicate priorities or roles (e.g., writing longer emails to the boss), for instance. Additional treatment could be included to take into account message sizes. Longer messages would carry more weight in defining the corpus, while short ones would have less weight. One interesting approach would be to correlate message length with thread types (organizational, technical, etc.), which would provide an indication of what types of messages one is dealing with and enable appropriate indexing.

Privacy Issues

Whenever information is automatically collected or distributed, privacy becomes an issue. The automatic management of a user's nimbus is an issue left for further research. The choice of what to make available can be left to the user, by adopting a simple three-tiered privacy scheme, for instance, whereby the user can define whether the information in his or her collaboration sphere is public (anyone can see it), protected (selected peers can see it) or private (nobody can see). The user should be able to select alters or resources that fall within each of the tiers, and who has access to what at the protected level. When an event is found that should be propagated to other peers, it would be checked against the specified restrictions to see if it falls within a specific privacy tier and whether it can be sent to the requesting agent. The same reasoning could be applied to different levels of information granularity.

Activities usually fall into one of the following categories: manipulation of shared objects, manipulation of non-shared objects and interaction between members. We assume that all shared objects and interactions within a collaboration sphere can be made public to members of that collaboration sphere. Examples would be editing or forwarding a file that has already been sent around as an attachment, or having a conversation related to the project with other members of the collaboration sphere.

Modern organizations are composed of networks of interacting actors, where relations between them are subject to constant renegotiation (Bernoux, 1995). More often than not, knowledge is exchanged and work is undertaken through these informal relations between workers, in networks that cut across departmental, functional and organizational boundaries. Thus, modern organizations require coordination and integration of activities across these boundaries, and information systems should provide support for distributed coordination and decision-making, while helping the individual user manage the multiple collaborations he or she is inserted in. Networked configurations give rise to loosely structured groups, where members have the freedom to reorganize as necessary to get work done.

Our approach to support loosely structured groups is based on egocentric email-based interaction analysis. This is a promising line of research that holds many possibilities for further work. Many studies have applied social network analysis to uncover relations between people and patterns of interaction, but few have used these patterns as a basis for a system to actively assist the user. With this work, we build on the Focus and Nimbus theory, adding methods to automatically determine a user's focus. The interplay between interaction and information needs to be further explored, and we believe this is a relevant issue in the construction of collaborative work support, especially as work becomes more decentralized and loosely structured, with actors gaining autonomy and freedom to revise their work arrangements.