Redes Complexas
Aula 13

Aula passada

- Configuration Model
- Propriedades
- Stochastic Block Model
- Propriedades

Aula de hoje

- Busca em redes
- Explorando estrutura
- Navegação em redes
- Algoritmo eficiente e estrutura

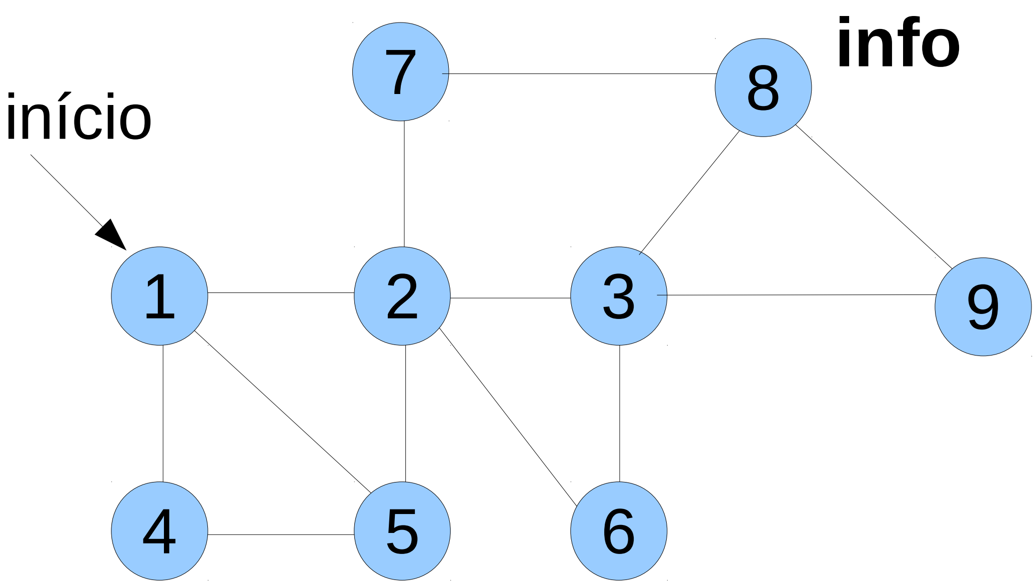
Busca em Redes

- Vértices possuem informação
  - conteúdo das páginas web
  - arquivos em uma rede P2P
  - conhecimento em uma rede social
- Informação distribuída
  - não existe repositório centralizado de quem conhece o que
- **Problema**: localizar informação
  - de maneira eficiente!

Algoritmos de busca!

Busca em Redes

- Vértices possuem determinada informação
- Determinado vértice deseja localizar esta informação
- Algoritmo?



BFS (Breadth First Search)

- Busca em largura
- Iniciado pelo vértice fazendo a busca
  - busca tem identificador único
- Se ao ser perguntado pela primeira vez, vértice não possui informação, então pergunte a **todos** vizinhos
- Se possui informação, responda ao vértice inicial
- **Vantagem**
  - Rápido! Caminho mais curto até informação
- **Desvantagem**
  - Pesado! Todos os vértices serão atingidos
  - Não há como interromper o processo (eficientemente)

Flooding Limitado

- *Flooding*: busca em largura em redes

Como controlar o *overhead*?

- TTL: *Time-to-live*
  - Maior número de saltos que busca pode fazer
  - informação contida na mensagem, decrementada a cada salto
- Usada na prática: Gnutella, etc.
- Problemas?

Random Walk

- Iniciado pelo vértice fazendo a busca
  - busca tem identificador único
- Se ao ser perguntado pela primeira vez, vértice não possui informação, então pergunte a **um** vizinho
  - Qual? Escolher aleatoriamente
- Caso contrário, responda ao vértice inicial
- **Vantagem**
  - Leve! Busca se propaga de vértice em vértice
- **Desvantagem**
  - Lento! Demora para encontrar informação

Random Walk 2

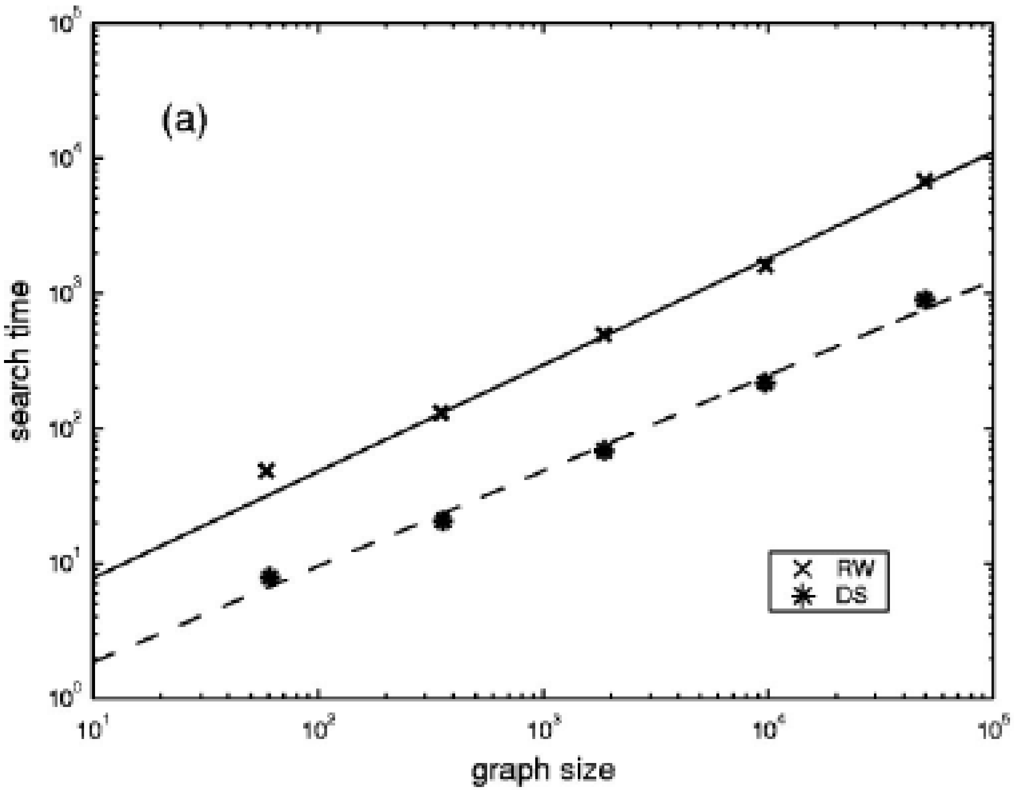
- Como melhorar eficiência da RW?
- **Problema**: passeio pode estar ao *lado* da informação e não encontrá-la!
  - informação em um vizinho, mas outro é escolhido (ao acaso)
- **Solução**: perguntar a todos os vizinhos antes de andar com a busca
  - se algum vizinho tem informação, então pára
- Diminui o tempo de busca, mas não fundamentalmente

Explorando a Estrutura

- Como tirar proveito da estrutura topológica?
  - ex. rede com graus seguindo uma lei de potência
- **Idéias**
  - Propagar a busca para vizinho de maior grau (e não aleatoriamente)
- Vértice de maior grau tem acesso a mais informação
- Direcionar busca para quem sabe mais
- Diminui ainda mais tempo de busca
- High-Degree Seeking Walks
  - proposto e implementado

Avaliação (1/2)

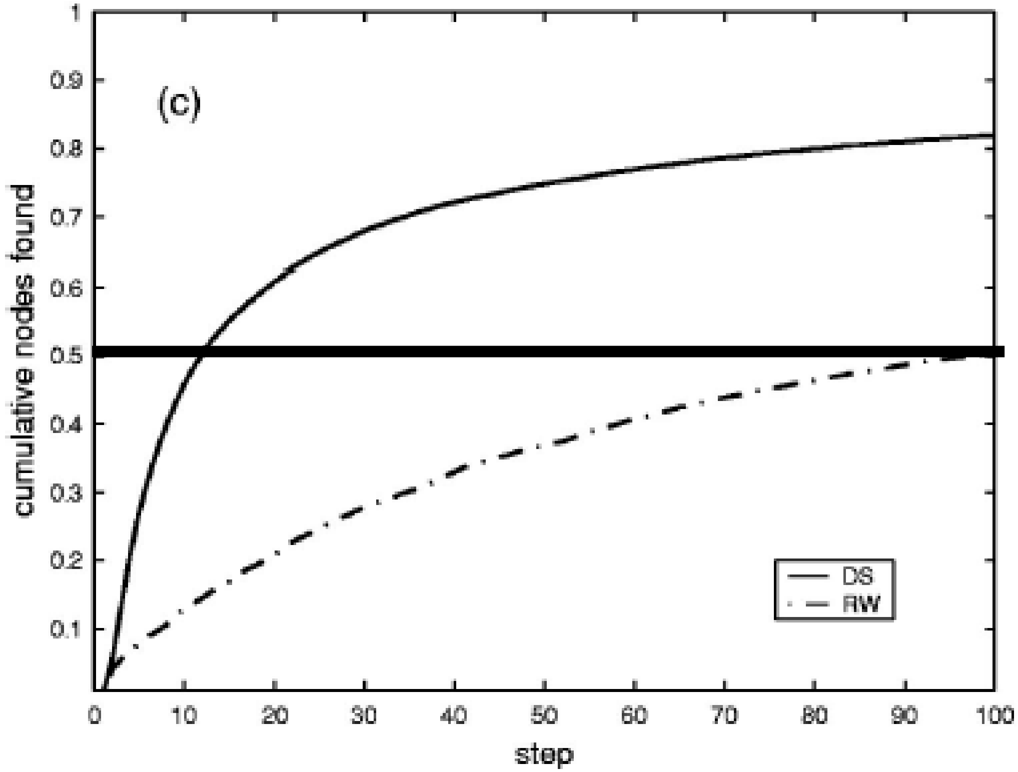
- Rede com lei de potência no grau (α=2.1)
- Vértice de origem/destino da busca escolhido uniformemente
- Busca termina quando chega a um vizinho do destino



- Tempo médio para encontrar destino
- Medido em número de passos
- 10 vezes mais rápido!

Avaliação (2/2)

- “Alcance” da busca
- Fração de vértices pesquisados em função do número de passos



- N = 10000
- DS: metade da rede pesquisada em 10 passos
- RW: 100 passos!

- Problema desta abordagem?

Busca com Informação

- Busca até então era cega
  - não há informação local sobre onde encontrar informação desejada
- Em muitas redes, há informação local sobre informação desejada
  - Ex. ser apresentado ao Barabási através da rede social
- Conhecimento local de informação pode guiar processo de busca

Como explorar informação local?

Experimento de Milgram

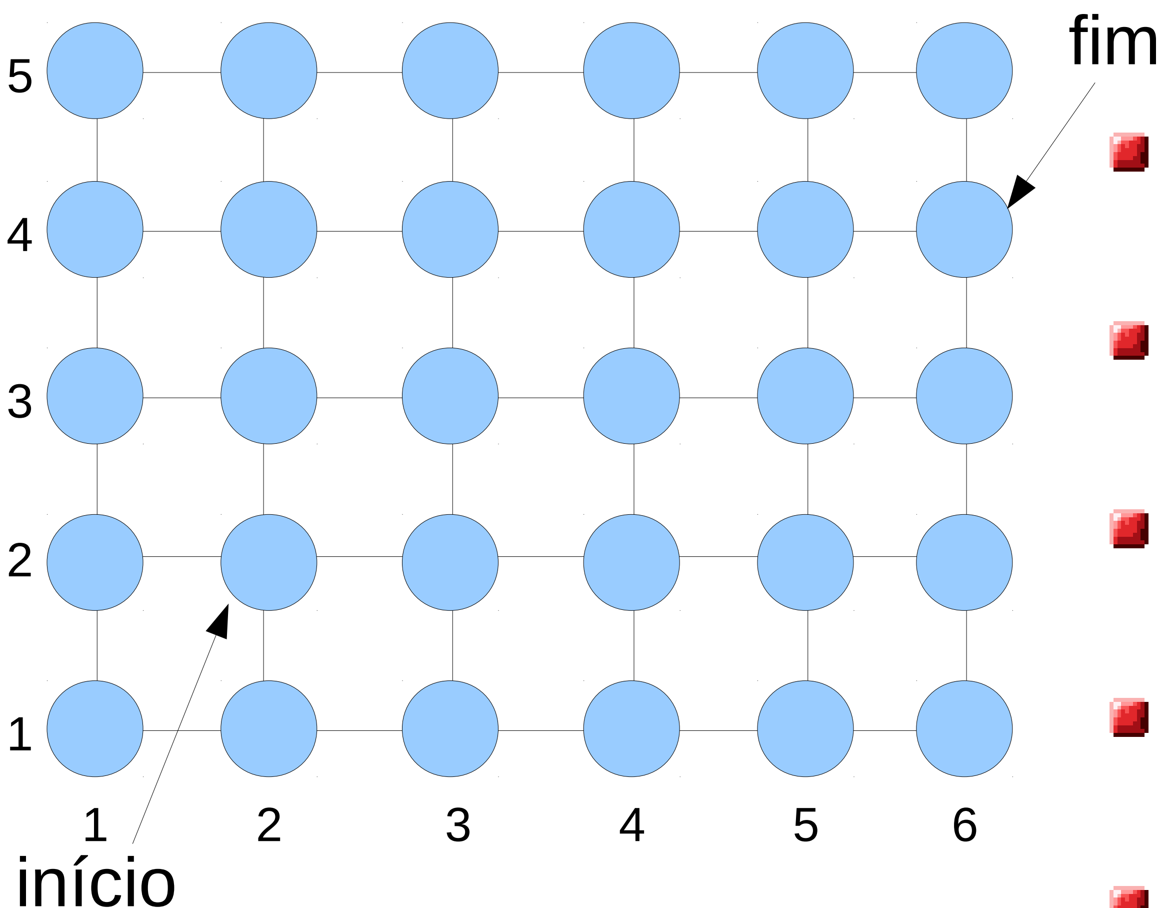
- Enviar carta a um desconhecido através apenas da rede social
- **Surpreendente**: Caminhos curtos existem
  - ~ 6 passos na rede social
- **Igualmente surpreendente**: Pessoas comuns encontraram caminhos curtos
  - rede social é vasta, busca aleatória levaria a caminhos muito mais longos
  - pessoas não têm conhecimento da rede social
- Pessoas navegam bem na rede social
- Mas como?

Algoritmo de Busca Guloso

- Algoritmo guloso
  - a cada passo, busca transferida para vizinho ***mais próximo*** do destino
- Algoritmo é míope
  - considera apenas próximo passo
  - tenta se aproximar o máximo do destino a cada passo
- Métrica para proximidade depende do contexto
  - distância física, distância social, etc.
- Provavelmente usado por pessoas para navegar na rede social

Exemplo

- Rede: grid em 2D
- Distância: distância no grid (em saltos)



- Vértice inicial conhece coordenadas do fim
- Vértice inicial conhece coordenadas dos vizinhos
- Escolhe vizinho que está mais próximo do fim
- Passa busca para vértice vizinho (que repete)
- **Complexidade?**

Desempenho do Guloso

- Desempenho do algoritmo depende da topologia da rede
- Guloso pode **não encontrar** caminho mais curto na rede
  - apesar do caminho existir!
- Estrutura da rede é fundamental para desempenho do algoritmo
  - caminhos exponencialmente menores
- Observação e demonstração feita por Kleinberg em artigo influente

Navigation in a small world
Jon Kleinberg, Nature, 2000 – 1489 (+3900 citações)

Modelo de Kleinberg

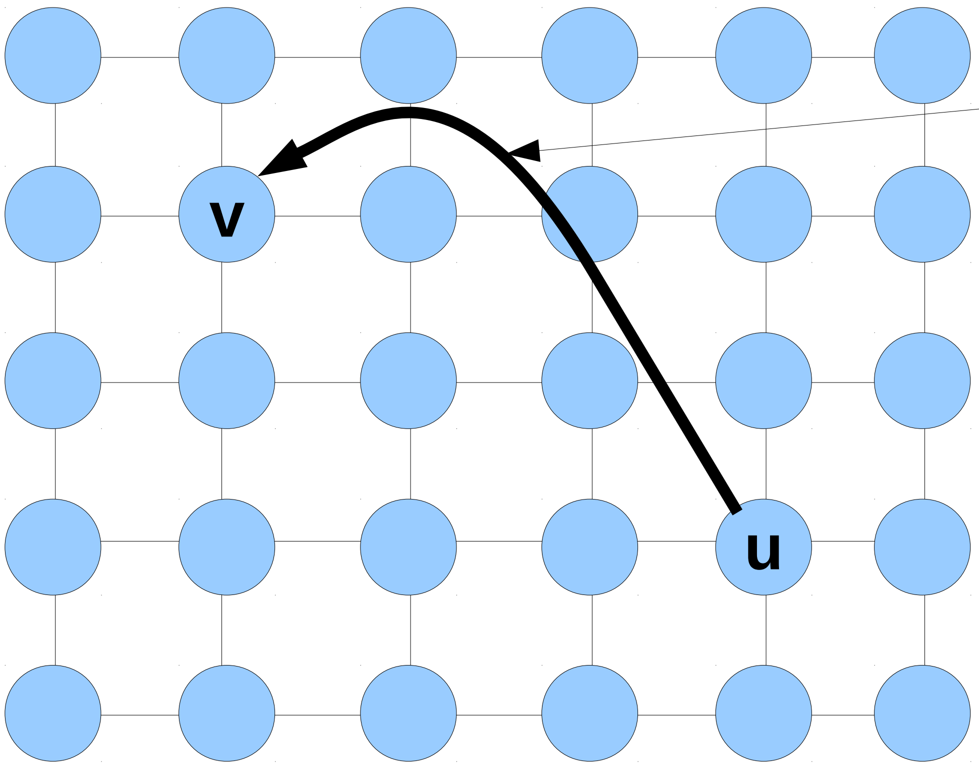
- Rede: grid em 2D
- Distância: distância no grid (em saltos)
- Adicionar "atalhos" no grid
  - criação de caminhos curtos
  - cada vértice adiciona q atalhos
- Atalhos aleatórios
  - probabilidade inversamente proporcional a distância entre os vértices no grid

$$p_{uv} \propto \frac{1}{d(u,v)^{\alpha}}$$

, onde *d(u,v)* é a distância entre os vértices *u* e *v* no grid, α é constante > 0

Modelo de Kleinberg

- Rede: grid em 2D + atalhos probabilísticos



Atalho ocorre com probabilidade ~ $5^{-\alpha}$

- Atalhos é função de α
- α grande: atalhos longos ocorrem mais raramente
- α pequeno: atalhos muito longos podem ocorrer

- **Problema**: Calcular desempenho do algortimo guloso em função de α

Modelo de Kleinberg

- Assumir rede é um látice, n x n vértices
- Cada vértice tem exatamente um atalho
- Probabilidade de atalho de u levar ao vértice v

$$p_{uv} = \frac{d(u,v)^{-\alpha}}{\sum_{u \neq v} d(u,v)^{-\alpha}}$$

, onde *d(u,v)* é a distância entre os vértices *u* e *v* no grid

- Parâmetro α constante > 0
- Desempenho: número médio de saltos para encaminhar mensagem do origem ao destino (T)
  - escolhidos aleatoriamente no grid

Avaliação Teórica

- 0 <= α < 2 ⟶ $T \geq c\, n^{(2-\alpha)/3}$
- α > 2 ⟶ $T \geq c\, n^{(\alpha-2)/(\alpha-1)}$

Tempo médio é polinomial!

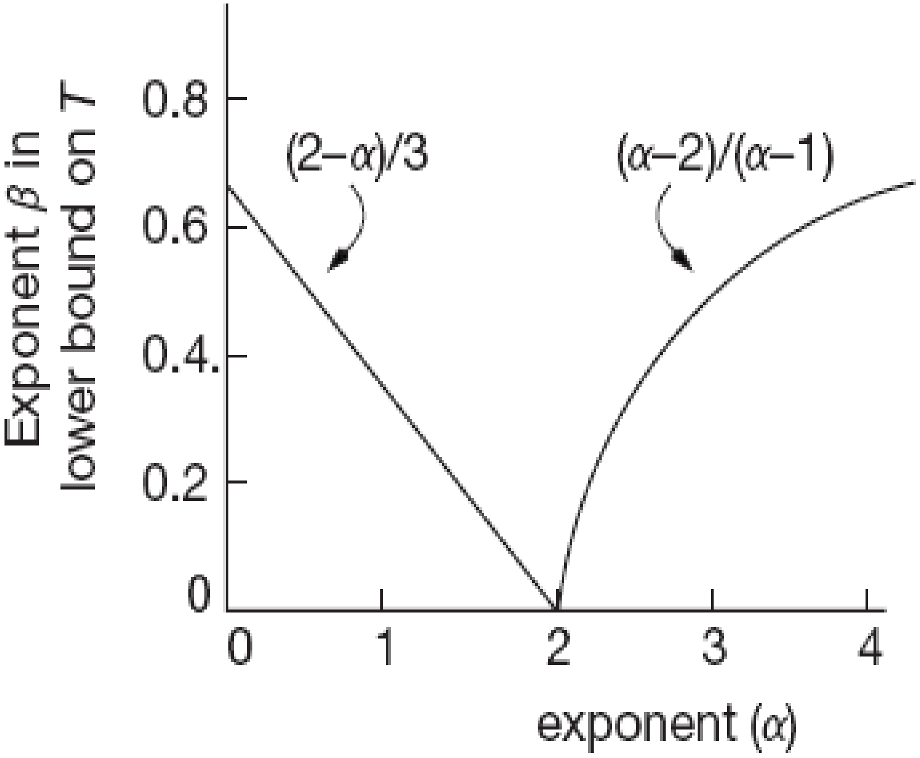
- α = 2 ⟶ $T \leq c \log^2 n$

Exponencialmente menor!

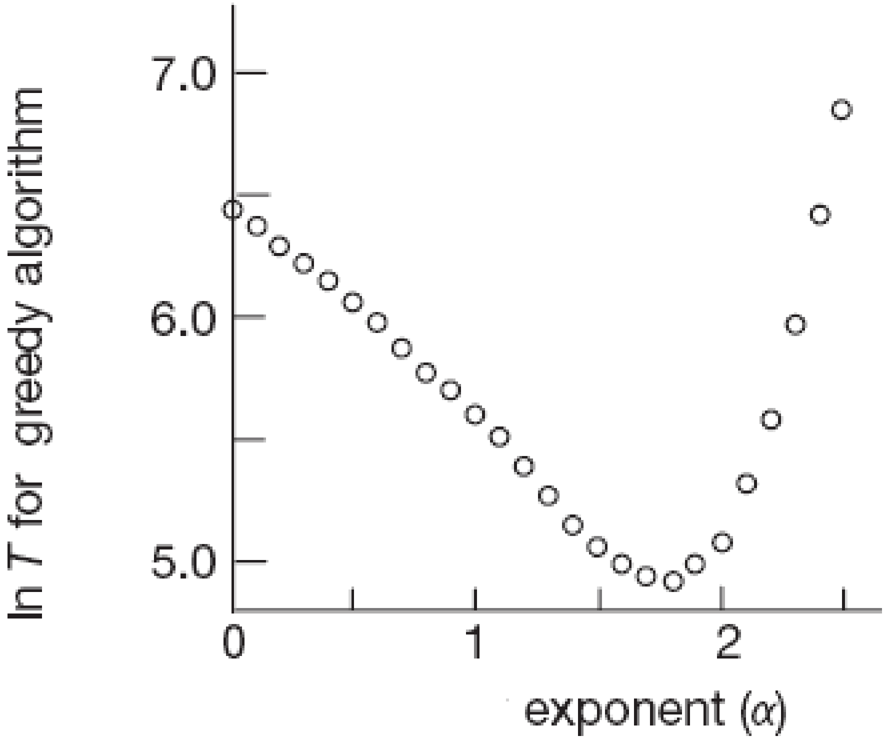
Avaliação

- Gráfico do expoente β

$$T \geq c\, n^{\beta}$$



- Avaliação empírica (simulação), n = 20K, 1K rodadas, log T



Observações

- $\alpha = 2$: vértices tem em média o mesmo número de vizinhos em todas as distâncias
  - efeitos se cancelam
- $\alpha < 2$: distribuição de atalhos muito uniforme
  - algoritmo não consegue explorar os atalhos
- $\alpha > 2$: não há atalhos longo suficientes
  - não existem caminhos curtos o suficiente

Idéia da Prova

- Para o caso $\alpha = 2$
- Argumento geométrico
- Algoritmo procede em fases
  - Fase j se distância até destino é entre 2^j e 2^{j+1}
- Quantas fases, no máximo?
  - log n
- Número médio de passos até mudar de fase?
  - encontrar atalho para próxima fase
  - log n
- Logo, $T <= c (\log n)^2$

Generalizações

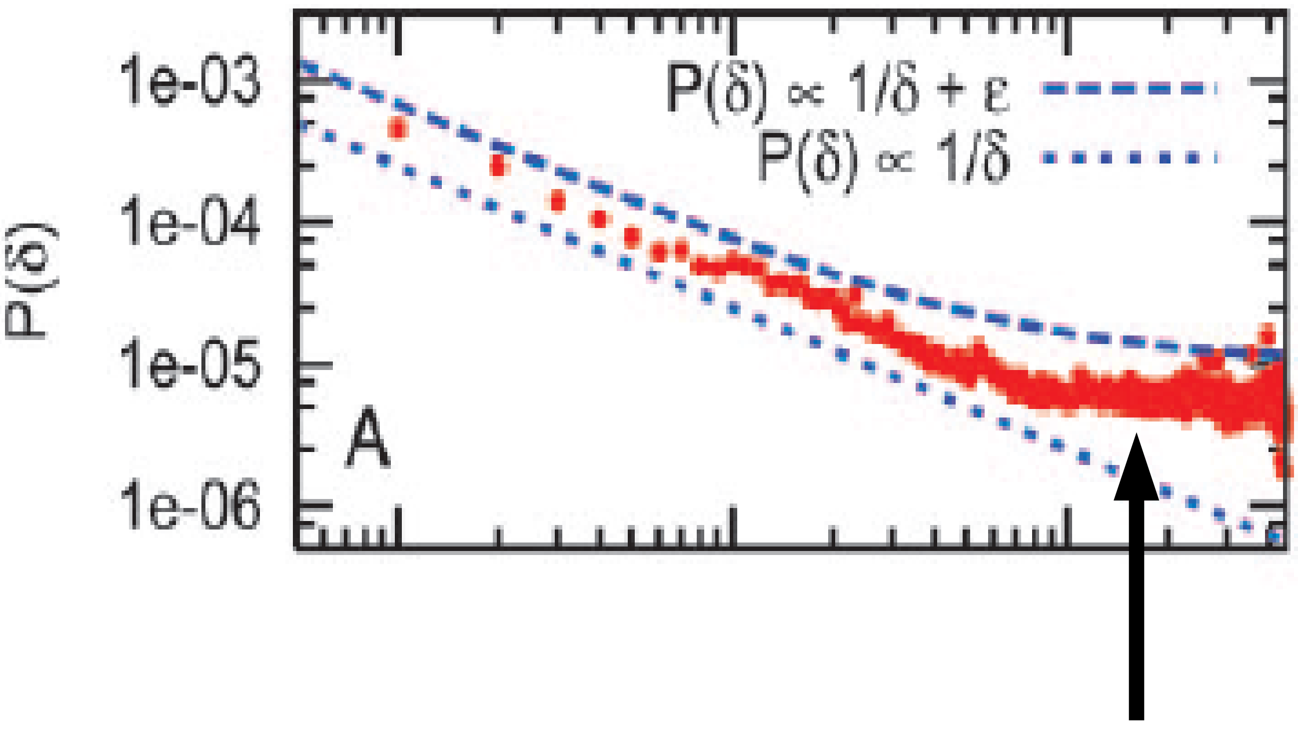
- Generalizado para qualquer número constante de atalhos
  - não muda complexidade
- Generalizado para qualquer algoritmo que utiliza apenas informação local
  - guloso não usa o passado
- Generalizado para grids com d dimensões
- Ponto crítico ocorre quando $\alpha = d$
  - guloso encontra caminhos curtos apenas nestes caso
  - mesma intuição

Redes Sociais na Prática

- Amizades estão correlacionadas a distância geográfica?
  - distância entre cidades onde pessoas moram
- Estudo empírico de rede social online (2004)
- LiveJournal
  - blogeiros tem identidade, endereço
  - declaram amizade por outros blogueiros
- ~500K usuários que declararam endereço válido nos EUA

Distribuição de Distância

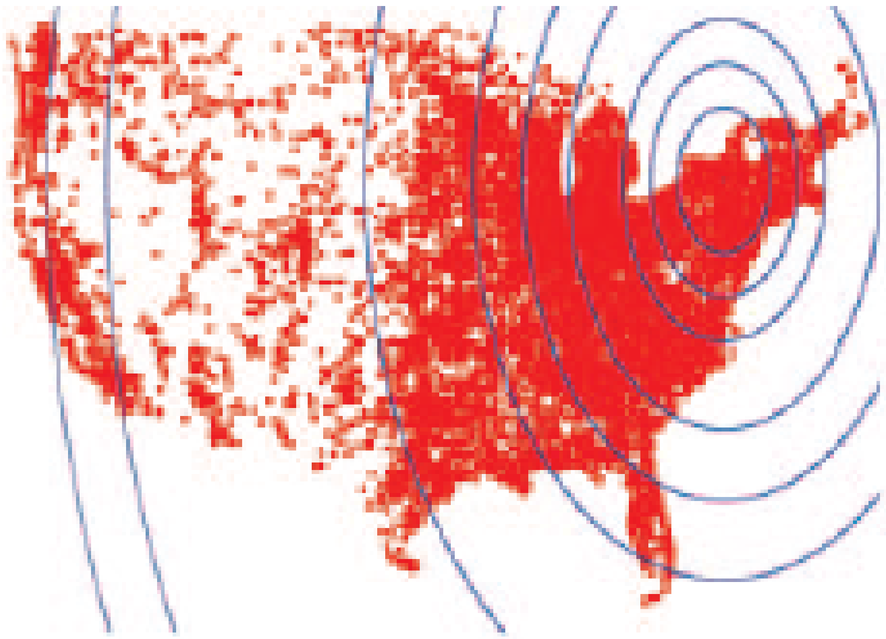
- Distância entre amigos, u e v (aresta)
  - usando endereço postal
- d(u,v) = δ, arrendondado para múltimplos de 10Km
- P(δ) : fração de amigos com distância δ



- Segue lei de potência até certo ponto
- Não influencia para distâncias muito grandes
- P(δ) converge para $5x10^{-6}$

Densidade Populacional

- Qual relação entre tamanho da vizinhança (em pessoas) e distância geográfica?
  - Exponencial? Igualmente distribuída?
- Cada ponto representa uma pessoa (*blog*)



- Círculos concêntricos, centro Ithaca, NY
- Cada círculo representa 50K pessoas

Amizade, Distâncias, Navegação

- Duas pessoas u e v moram a 500m
  - Em Xapuri, certamente são conhecidos
  - Em Copacabana, muito pouco provável
- Probabilidade de amizade depende de distância e da densidade populacional
- Novo modelo para capturar como amizades ocorrem
  - extensão de modelo de Kleinberg
- Desempenho do guloso é bom (poly-log)

"In a lamentably imperfect world, it is remarkable that people form friendships so close to the perfect distribution for navigating their social structures."

- Ao acaso? Ou existe algo por trás?