Redes Complexas
Aula 5

Aula passada

- Centralidade
- Grau, betweeness, closeness
- Autovetor, Katz, PageRank, PPR

Aula de hoje

- Padrões de mixagem (*mixing patterns*)
- Correlação entre graus
- Similaridade entre vértices

Padrões de Mixagem

- Vértices possuem tipos diferentes
  - propriedade do vértice, não estrutural
- Redes sociais: gênero, idade, nacionalidade, time
- Outros: composição química, velocidade, idioma, função, etc

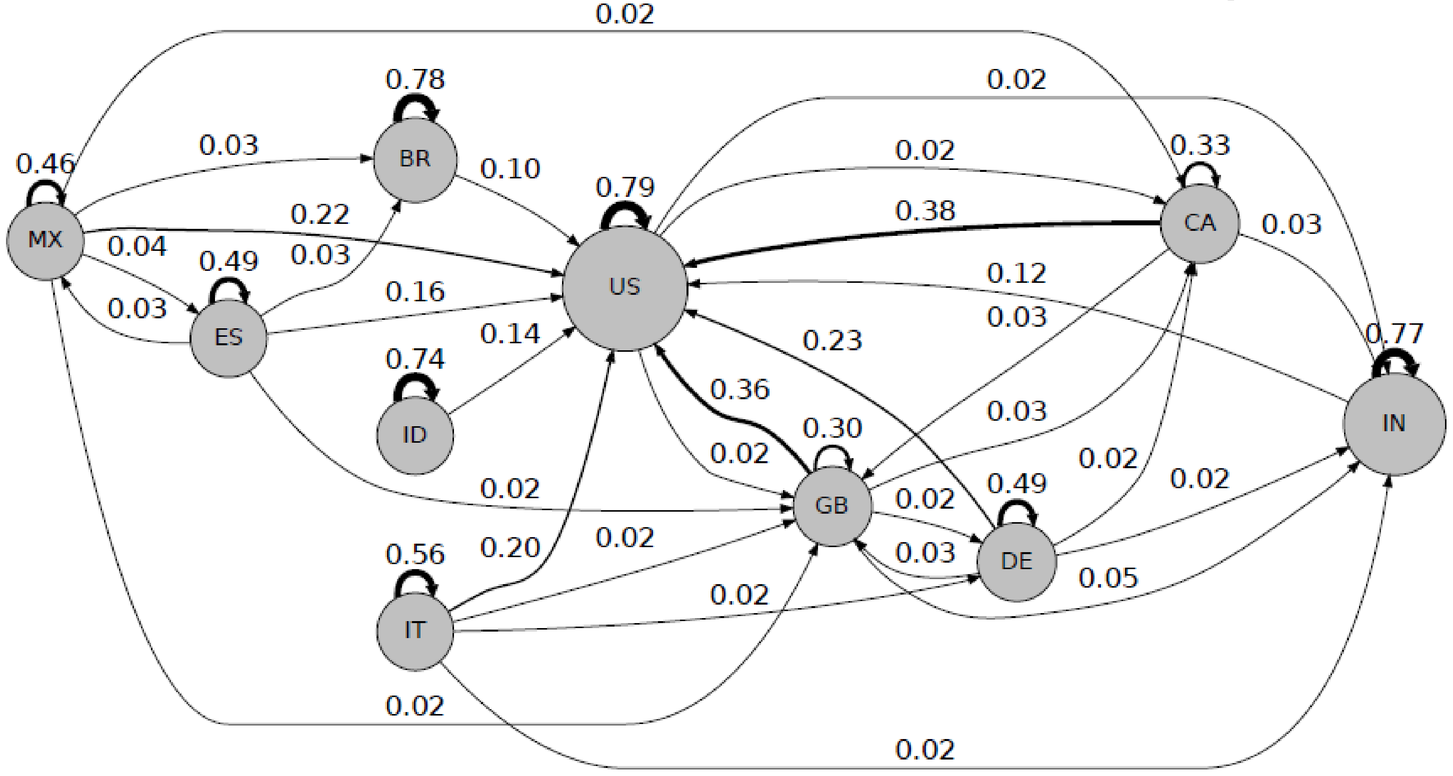
Como vértices se *misturam*?

  - Quais são as características na pontas das arestas?

- **Problema**: definir uma métrica para capturar este fenômeno

Exemplo

- Tese de mestrado (2014 na UFMG)
  - estudo sobre Google+ por Gabriel Silva
- Pessoas e amizades (direcionada), pessoas em países
  - peso é fração de arestas saindo do tipo (>0.01)



- Quem são os *hubs*? Quais outras características?

Exemplo

- Rede social, casamento entre grupos étnicos diferentes

		women			
		black	hispanic	white	other
men	black	0.258	0.016	0.035	0.013
	hispanic	0.012	0.157	0.058	0.019
	white	0.013	0.023	0.306	0.035
	other	0.005	0.007	0.024	0.016

TABLE I: The mixing matrix e_{ij} and the values of a_i and b_i for sexual partnerships in the study of Catania *et al.* [23]. After Morris [24].

- Tendência de relacionamento entre vértices do mesmo tipo
- Como representar (medir) fenômeno?

Coeficiente de Assortatividade

- Assortative coefficient
- e_{ij} : fração de arestas entre vértices do tipo i e j
  - $e_{ij} = n_{ij} / m$ ← m é o número total de arestas da rede

$$\sum_{ij} e_{ij} = 1, \qquad \sum_{j} e_{ij} = a_i, \qquad \sum_{i} e_{ij} = b_j,$$

- a_i : fração de arestas que incidem sobre tipo i
- b_j : fração de arestas que incidem sobre tipo j
- Na maioria dos casos $a_i = b_j$, para i = j
- Mas não quando grafo é bipartido (caso anterior)

Coeficiente de Assortatividade

- Assortative coefficient

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

$a_i b_i$ é o valor esperado de e_{ii} se fosse aleatório

Fator de normalização

- r = 0, relacionamentos totalmente aleatórios
- r = 1 : relacionamentos somente entre iguais
- r < 0 : relacionamento entre diferentes

Exemplo

- Rede social, casamento entre grupos étnicos diferentes

		women				
		black	hispanic	white	other	a_i
men	black	0.258	0.016	0.035	0.013	0.323
	hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	other	0.005	0.007	0.024	0.016	0.053
	b_i	0.289	0.204	0.423	0.084	

TABLE I: The mixing matrix e_{ij} and the values of a_i and b_i for sexual partnerships in the study of Catania *et al.* [23]. After Morris [24].

- r = 0.621

Mixagem em Função da Estrutura

- Exemplo anterior: vértices tinham tipo
- Definir mixagem em função da estrutura

Idéias?

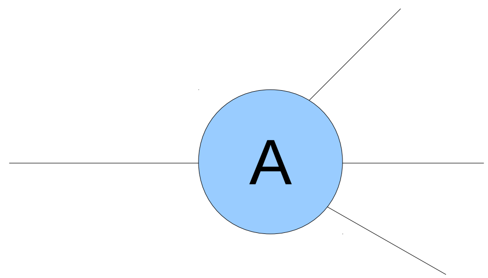
- Característica topológica de vizinhos que possa ser quantificada
- Ex. grau do vértice como “tipo”
  - correlação entre graus de vértices

Correlação entre Grau

- Métrica para medir correlação entre graus de vértices vizinhos
- Correlação positiva: graus altos são vizinhos
- Correlação negativa: grau alto é vizinho de grau baixo
- Métrica: Coeficiente de correlação de Pearson
  - Medida estatística de correlação empírica
  - Aplicada ao grau de vértices vizinhos

Correlação entre Grau

- p_k : fração de vértices com grau k
- q_k : fração de vértices com *restante de grau* igual a k



$$q_k = \frac{(k+1)p_{k+1}}{z},$$

grau médio do grafo

- Relação entre q_k e notação anterior:

$$\sum_j e_{jk} = q_k.$$

fração de arestas entre vértices de grau j e k

Correlação de Pearson

- Aplicada a grau de vértices vizinhos

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2},$$

valor esperado de e_{jk} se fosse aleatório

Variância (empírica) de q_k

- r = -1, perfeitamente dissociados
- r = 1 : perfeitamente associados
- r = 0 : associação aleatória

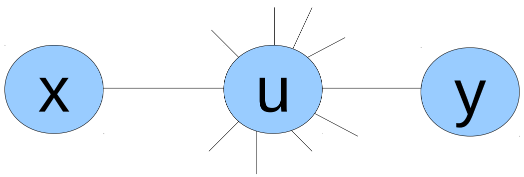
Medida de Similaridade

- Como definir uma medida de similaridade entre os vértices?
- **Ideia:** número de vizinhos em comum
  - n_{ij} : número de vizinhos em comum entre vértices i e j
- Normalização?
  - dividir por n penaliza vértices de grau baixo
- **Jaccard Similarity**
  - medida de similaridade entre dois conjuntos (os vizinhos dos vértices)
  - J(A,B) = |interseção(A,B)| / |união(A,B)|

Coeficiente Adamic/Adar

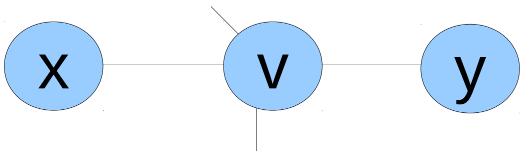
- Jaccard não penaliza vértices de grau grande
  - grau grande está em muitas interseções
- **Ideia:** dar mais peso para interseção com vértices de grau menor

u tem grau 10



Menos indicativo da
similaridade entre x e y

v tem grau 4



Mais indicativo da
similaridade entre x e y

- Adamic/Adar: soma ponderada inversamente pelo grau dos vizinhos em comum

$$A(x,y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$$

Grau do vértice u,
na interseção de x e y

mais de 2300 citações!

Adamic, L; Adar, E., "Friends and neighbors on the web", *Social networks* (2003)

Similaridade de Coseno

- Usar ideia de espaço vetorial
- Vértice é um vetor no espaço vetorial de n dimensões (uma dimensão por vértice)
  - vizinhos possuem valores 1 no vetor
- Usar geometria
  - coseno do ângulo entre vetores (dois vértices)
  - cos(theta) = x . y / |x| |y|
- Fazendo as contas...
  - x . y = número de vizinhos em comum
  - |x| |y| = média geométrica do no. de vizinhos
- $s_{ij} = n_{ij} / \mathrm{sqrt}(d_i)\ \mathrm{sqrt}(d_j)$

Similaridade Entre Vizinhos

- **Fato**: vértices vizinhos em redes reais possuem muitas similaridades
  - ex. maior chance de ter vizinho em comum

Como medir similaridade?

- Utilizando apenas a estrutura

Muitas maneiras!

- **Ideia:** Comparar com aleatório para ter referência