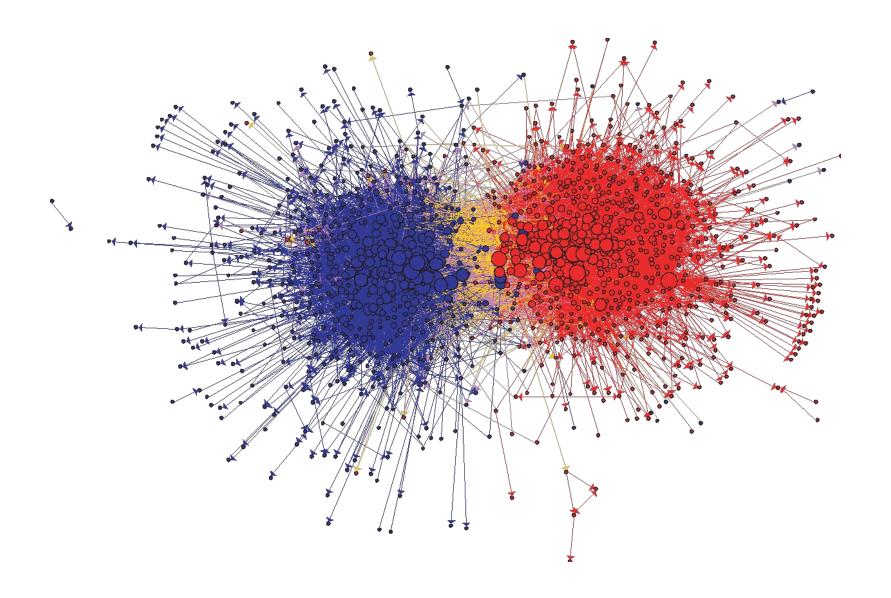
UNIDADE 5 Análise de Redes

Disciplina: Tópicos Especiais III (DEC7553)

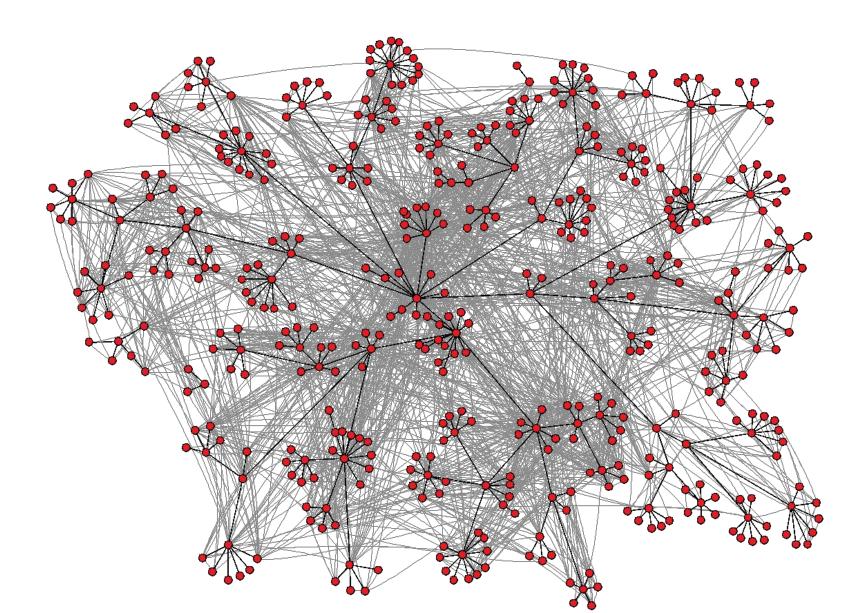
Prof. Alexandre L. Gonçalves

E-mail: a.l.goncalves@ufsc.br

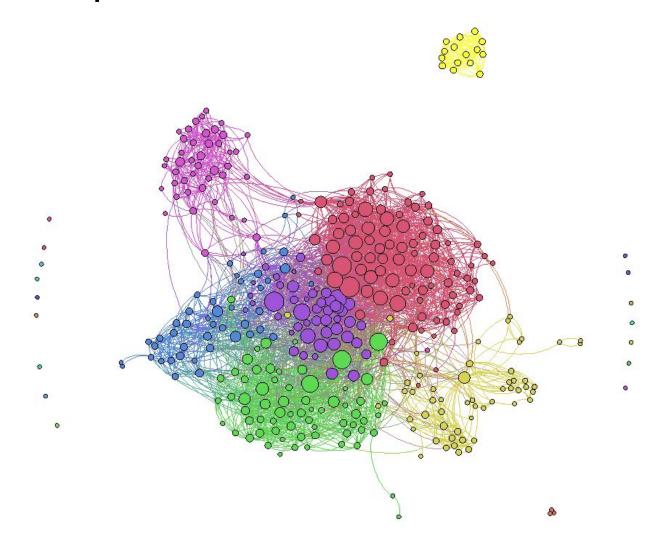
Blogs Políticos



Organizações

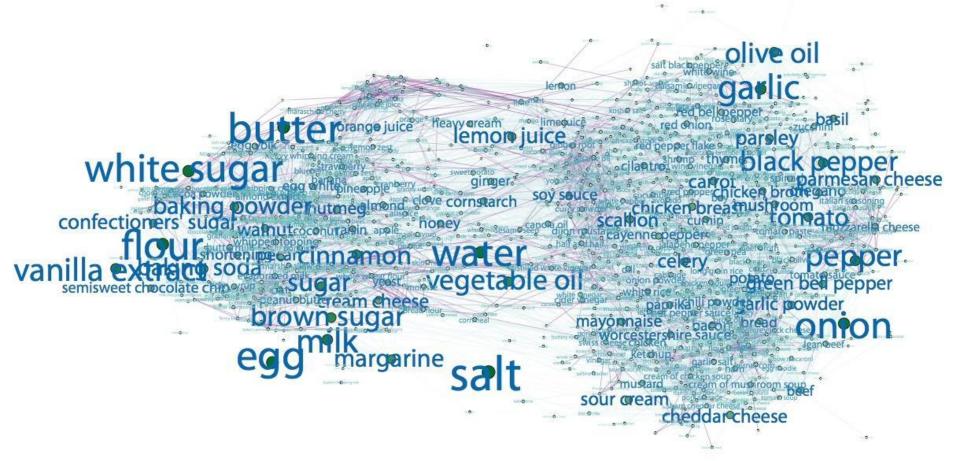


Rede a partir do Facebook[®]



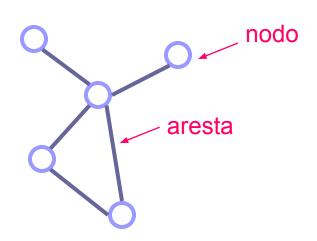


Rede de Ingredientes



O que é uma rede?

Rede é um conjunto de nodos conectados por links



"Network" ≡ "Graph"

Pontos	Linhas	
vertices	edges, arcs	Matemática
nodes	links	Ciência da Computação
sites	bonds	Física
actors	ties, relations	Sociologia

Elementos de rede: arestas

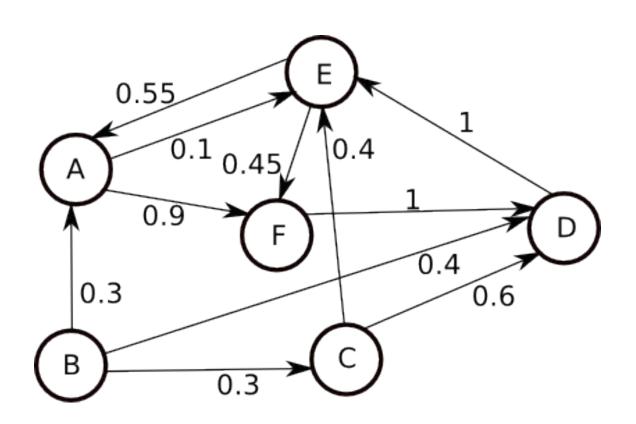
- Direcionado (também chamado arcos, links)
 - □ A -> B
 - A gosta de B,
 - A presenteou B,
 - A é filho de B
- Não direcionado
 - $\neg A < -> B \text{ or } A B$
 - A e B trabalham na mesma organização
 - A e B são irmãos
 - A e B são coautores

Atributos da aresta

Exemplos

- Peso (frequência de comunicação, número de artigos publicados conjuntamente, numero de ligações...)
- ranking (melhor amigo, segundo melhor amigo...)
- tipo (amigo, parente, coautor...)

Grafo Direcionado

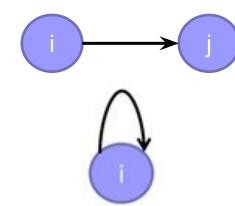


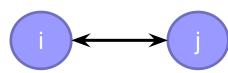
Representação dos dados

- Matriz adjacente
- Lista de arestas
- Lista de adjacência

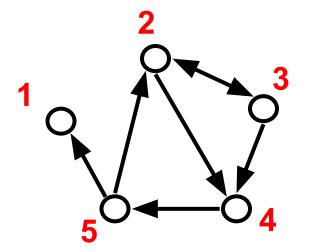
Matriz Adjacente (Adjacent Matrix)

- Representa arestas (quem é adjacente a quem) como uma matriz
 - A_{ij} = 1 se o nodo *i* tem uma aresta para o nodo *j* = 0 se o nodo *i* não possui uma aresta para *j*
 - A_{ii} = 0 a menos que o nodo possua um apontamento (aresta) para si mesmo (self-loop)
 - A_{ij} = A_{ji} se a rede é não direcionada, ou se *i* e *j* possuem uma aresta recíproca





Exemplo de matriz adjacente

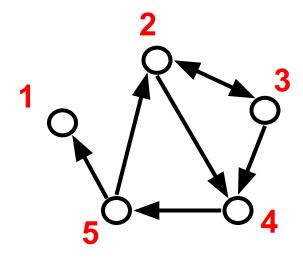


$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Lista de Arestas (Edge List)

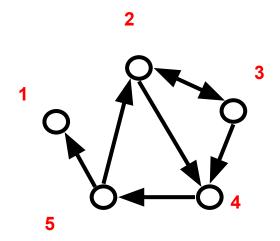
Lista de arestas

- **2, 3**
- **2, 4**
- **3, 2**
- **3, 4**
- **4, 5**
- □ 5, 2
- □ 5, 1



Lista de Adjacência (Adjacency List)

- Lista de Adjacência
 - É mais fácil de trabalhar se a rede é:
 - grande
 - esparsa
 - Rapidamente recupera todos os vizinhos para um nodo:
 - 1:
 - **2**: 3 4
 - **3**: 24
 - **4**: 5
 - **5**: 12



Análise de Rede Social

- A análise de redes sociais está focada em descobrir os padrões de interações das pessoas ou unidades de análise;
- A análise de redes sociais é baseada na noção intuitiva que estes padrões são importantes características da vida dos indivíduos.
- Analistas de redes acreditam que a maneira como um indivíduo vive, depende em grande parte de como o indivíduo está conectado dentro de uma rede de conexões sociais;
- Procura descobrir os vários tipos de padrões tentando determinar as condições em que esses padrões surgem para então descobrir suas conseqüências.

Origens

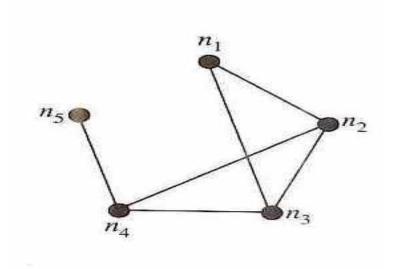
- Augusto Conte, século XIX
 - Propõe uma forma de olhar para a sociedade em termos de interconexões entre os atores sociais.
- Jacob Moreno, 1930
 - Criador da Sociometria
 - Primeiras referências ao termo "rede"
 - uso de grafo = sociograma
- "Idade das trevas" entre 1940 a 1970
- Harrison White, 1970
 - Renascimento da área em Harvard.
 - Aplicou a perspectiva estrutural para uma gama de problemas.
 - Maior contribuição foi a formação de uma geração de estudantes.

Métricas de ARS

- Analisar a rede
 - Diâmetro, densidade, etc.
 - Componentes
- Analisar nodos individualmente
 - Centralidade de grau, de intermediação e de proximidade
- Analisar grupos
 - Cliques, n-cliques, k-plex
 - Comunidades

Diâmetro

 O diâmetro de um grafo é a maior distância geodésica entre todos os pares de nodos presentes no grafo.



Distâncias geodésicas

$$d(1,2) = 1$$

$$d(1,3) = 1$$

$$d(1,4) = 2$$

$$d(1,5) = 3$$

$$d(2,3) = 1$$

$$d(2,4) = 1$$

$$d(2,5) = 2$$

$$d(3,4) = 1$$

$$d(3,5) = 2$$

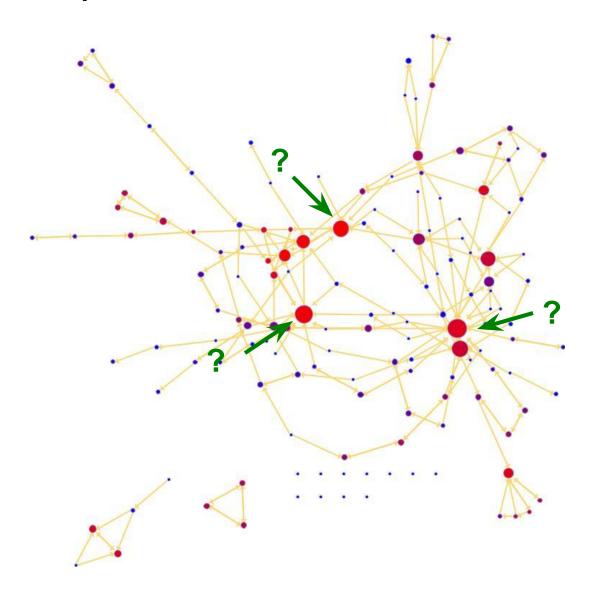
$$d(4,5) = 1$$

Diâmetro do grafo = max d(i,j) = d(1,5) = 3

Densidade

- Descreve o nível geral de ligação entre os pontos de um grafo
 - Quanto mais pontos estão ligados diretamente uns aos outros, mais denso será o grafo.

Grau: qual nodo tem a maioria dos links?

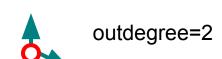


Propriedades de Nodo

- A partir de conexões imediatas
 - indegree
 quantas arestas direcionadas (arcos)
 incidem sobre um nodo



outdegree
 quantas arestas direcionadas (arcos)
 originam de um nodo



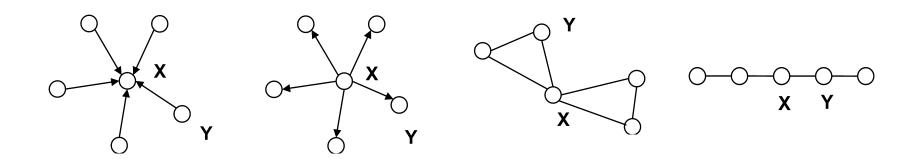
degree (in ou out)
 número de arestas incidentes em um nodo



- A partir do grafo completo
 - Centralidade (betweenness, closeness)

Diferentes noções de centralidade

 Para cada uma das redes abaixo, X possui uma centralidade maior do que Y de acordo com determinada medida.



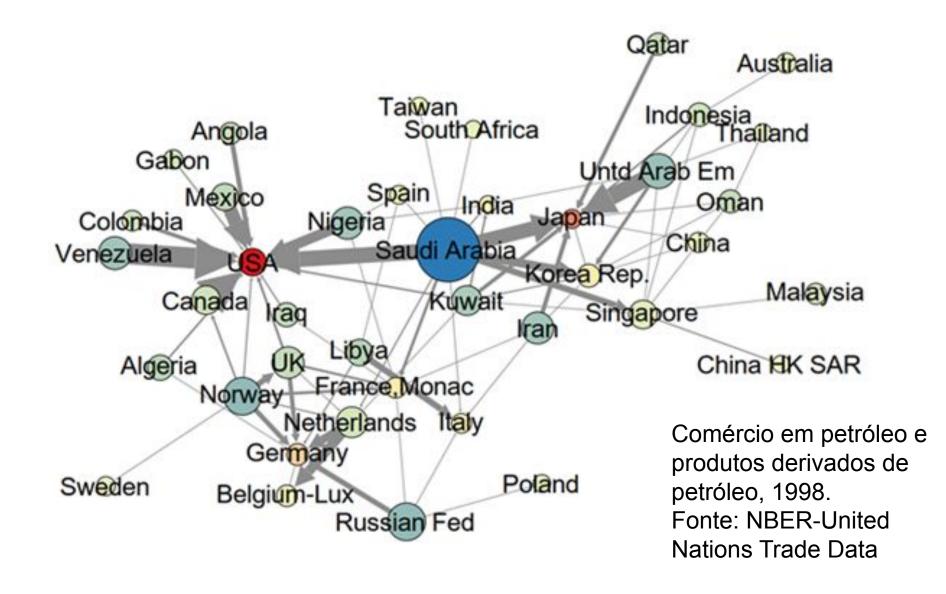
indegree

outdegree

betweenness

closeness

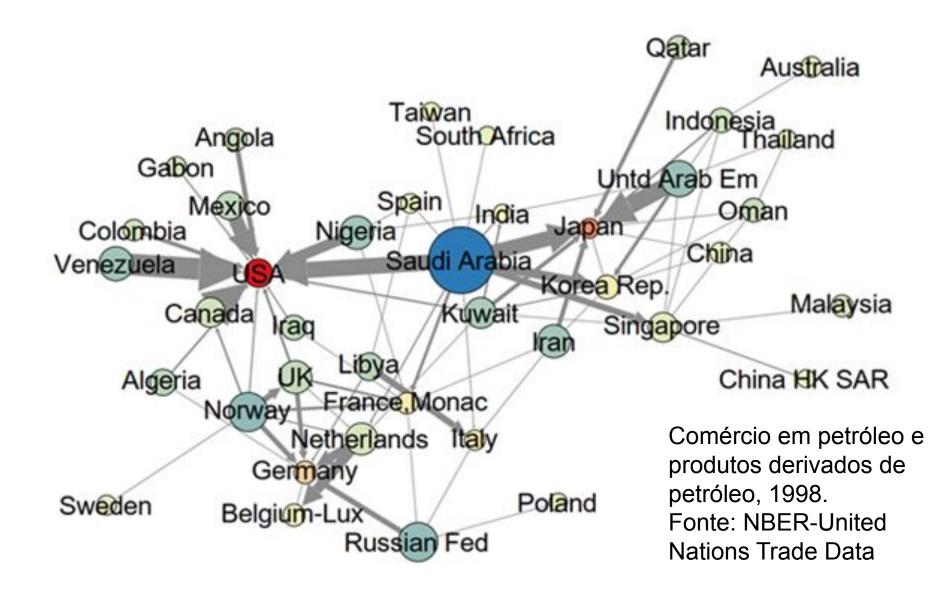
Exemplo



Pergunta?

- Que país possui um alto indegree (importa petróleo e derivados de petróleo de vários outros países)?
 - Arábia Saudita
 - Japão
 - Iraque
 - EUA
 - Venezuela

Exemplo



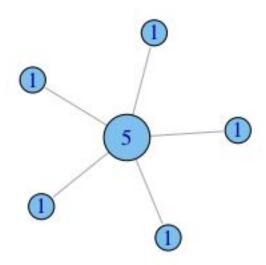
Pergunta?

- Que país possui baixo outdegree, mas exporta uma quantidade significativa (a espessura das arestas representam o valor \$\$ de exportação) de produtos derivados do petróleo
 - Arábia Saudita
 - Japão
 - Iraque
 - EUA
 - Venezuela

Centralidade de Grau

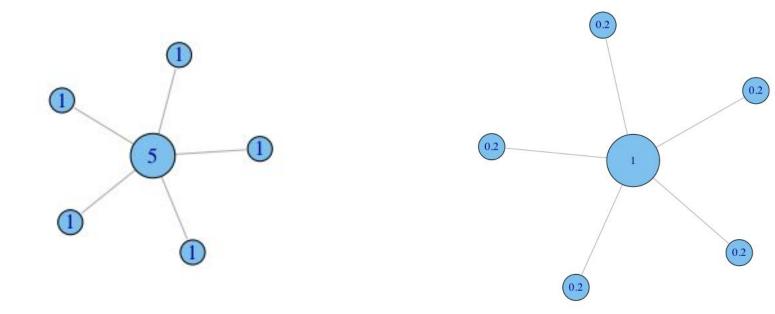
Grau não direcionado

Exemplo: nodos com mais conexões são mais centrais.



Normalização

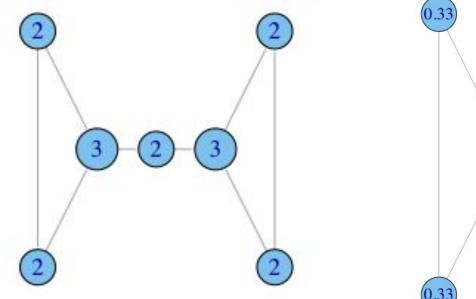
Divide cada grau pelo número de nodos menos 1, ou seja, (N-1)

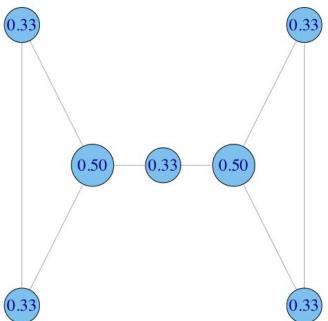


$$Gn = G / (N - 1)$$

Normalização

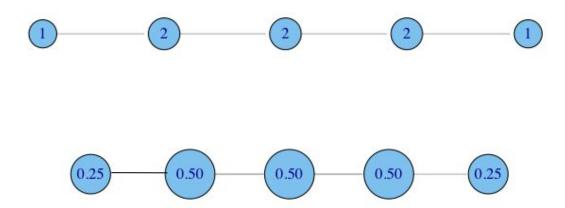
Divide cada grau pelo número de nodos menos 1, ou seja, (N-1)





Normalização

Divide cada grau pelo número de nodos menos 1, ou seja, (N-1)



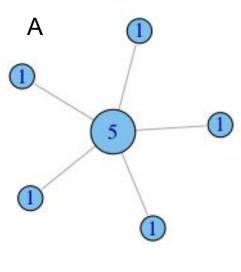
Centralização

Mede até que ponto um grafo está centralizado em relação ao seu nodo com maior centralidade de grau.

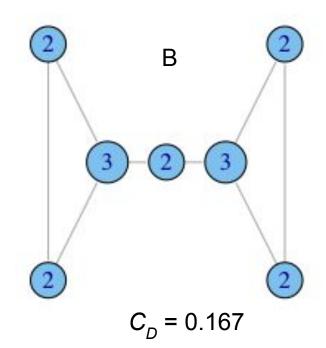
Fórmula Geral de Freeman:

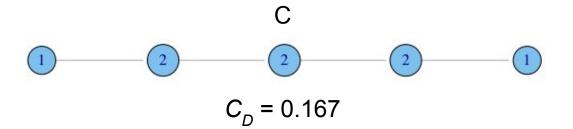
$$C_D = \frac{\sum_{i=1}^g \left[C_D(n^*) - C_D(i)\right]}{\left[(N-1)(N-2)\right]}$$

Centralização



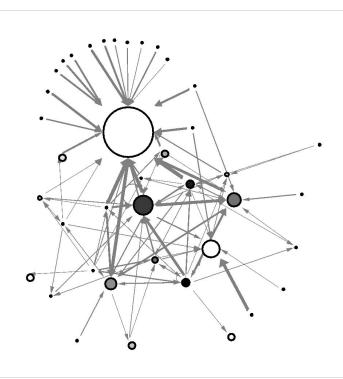
$$C_D = 1.0$$

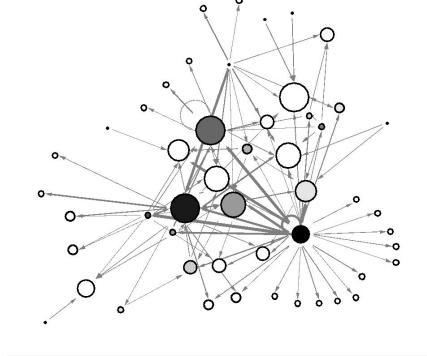




Centralização: Exemplo

Redes de Comércio



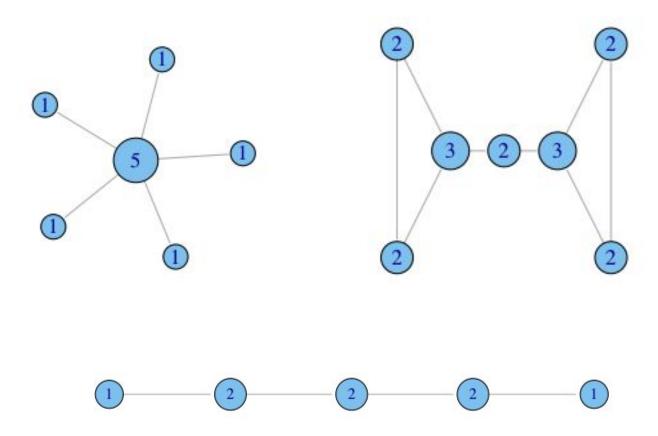


Alta centralização interna (in-centralization): um nodo compra de muitos outros

Baixa centralização interna: compras são mais igualmente distribuídas

O quê o grau não captura ?

De que maneira o grau falha em capturar a centralidade nos grafos abaixo?

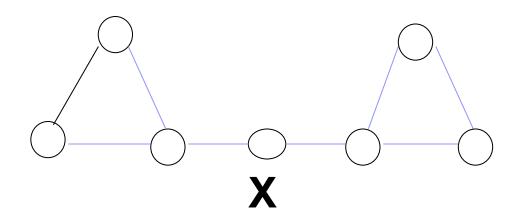


Grau - Restrições

- Grau é uma medida de centralidade local
 - "nodos podem ser bem conectados com sua vizinhança imediata, mas fazem parte de um clique relativamente isolado"* (Liu, 2008)
- Dois nodos com o mesmo grau podem não ter a mesma habilidade para influenciar outros nodos
 - Se o grau é usado para medir a influência local, então a influência do nodo depende de quem são seus vizinhos e o tipo de interação
 - Por exemplo, o número de seguidores no Twitter[®] não reflete diretamente a influência dentro da rede (deve-se considerar o número de retwittes)

^{* &}quot;Grupo de atores em que cada um está direta e fortemente conectado a todos os outros" (EMYRBAYER, 1994)

Brokerage não é capturado pelo grau



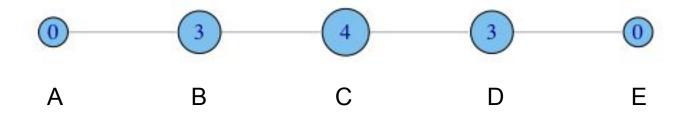
Um ponto com grau relativamente baixo pode desempenhar um papel 'intermediário' importante e ser central para a rede (Scott, 2004)

 Centralidade de Intermediação (Betweenness): capturando o brokerage

Mede a frequência com que determinado nodo aparece no caminho mais curto entre dois nodos quaisquer.

Betweenness

Versão não normalizada:

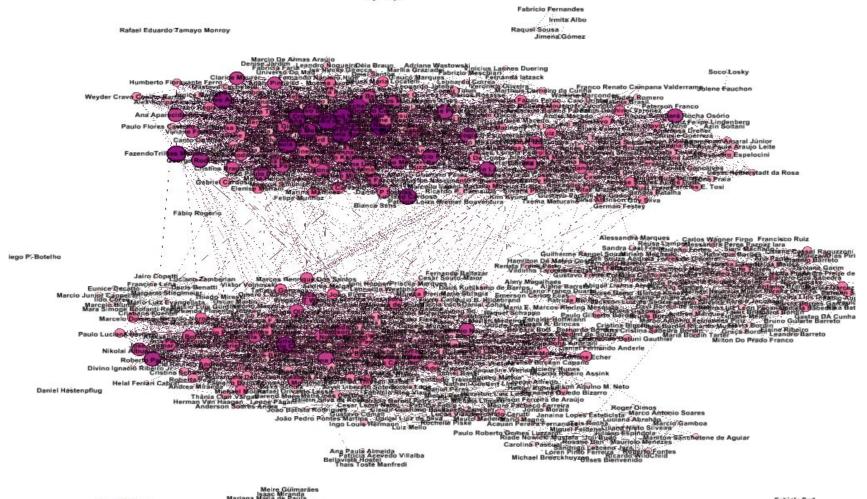


- A não possui qualquer intermediação
- B está entre A e 3 outros vértices: C, D, e E
- C está entre 4 pares de vértices (A,D),(A,E),(B,D),(B,E)
- Note que não existem caminhos alternativos para esses pares, deste modo, C obtém a agregação máxima

Normalização:
$$C_B(i) = C_B(i)/[(n-1)(n-2)/2]$$

Betweenness: Exemplo

Rede de um perfil do Facebook: o tamanho dos nodos é determinado pelo grau e a cor por betweenness.



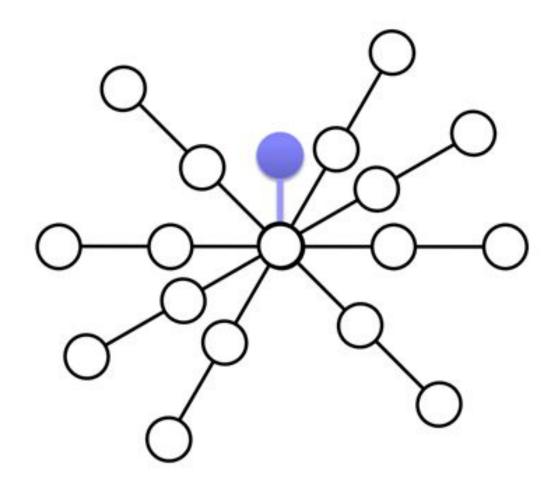
Centralidade de Proximidade (Closeness Centrality)

- E se não for tão importante possuir muitas conexões diretas?
- Talvez fosse interessante estar no "meio" das coisas, não tão longe do centro.

Representa a distância natural entre todos os pares de nodos, definida pelo soma dos caminhos mais curtos.

Centralidade de Proximidade

Não é necessário estar em uma posição de brokerage.



Centralidade de Proximidade

A **distância** de um nodo *s* é definido como a soma das distâncias para todos os outros nodos, e a sua proximidade como a inversa dessa **distância**.

Assim, quanto mais central é um nodo, menor é a distância total para todos os outros nodos.

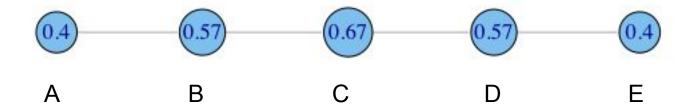
$$C_c(i) = \left[\sum_{j=1}^{N} d(i,j)\right]^{-1}$$

Centralidade de Proximidade Normalizada

$$C_C'(i) = (C_C(i))/(N-1)$$

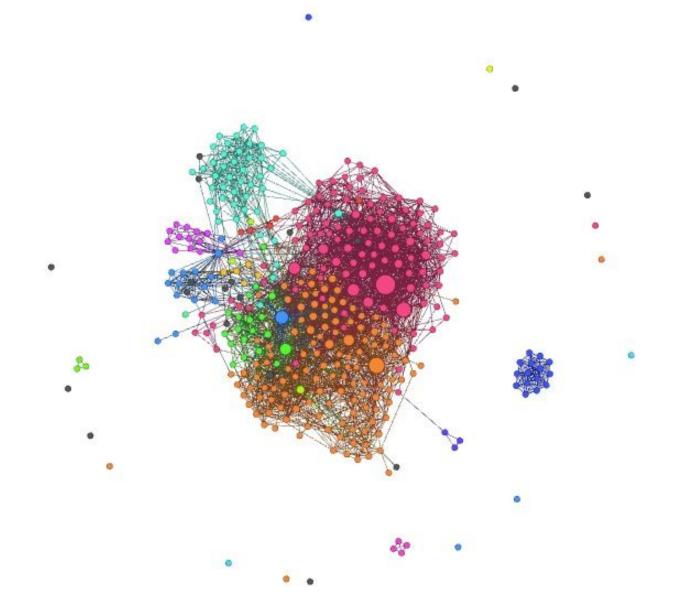
Closeness pode ser relacionada como uma medida que avalia a velocidade que uma informação levará para se espalhar para todos os outros nodos sequencialmente.

Centralidade de Proximidade: Exemplo



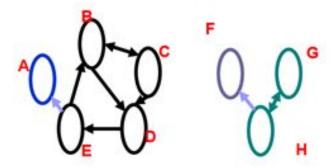
$$C'_{c}(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1}\right]^{-1} = \left[\frac{1+2+3+4}{4}\right]^{-1} = \left[\frac{10}{4}\right]^{-1} = 0.4$$

Está tudo conectado?



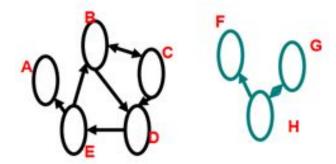
Componentes Conectados

- Componentes fortemente conectados
 - Cada nodo dentro do componente pode ser alcançado pelos demais nodos no componente seguindo os links direcionados
 - Componentes fortemente conectados
 - BCDE
 - A
 - GH
 - F



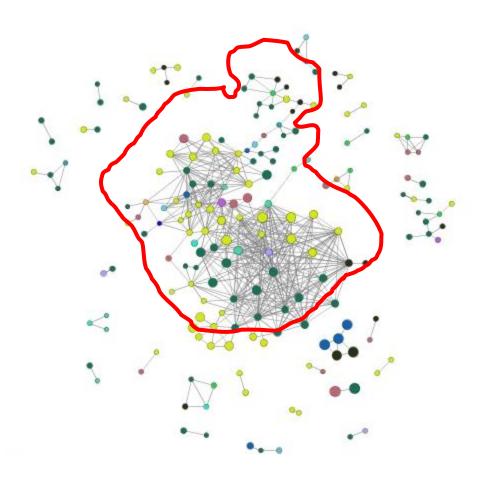
- Componentes fracamente conectados
 - Cada nodo pode ser alcançado a partir dos demais nodos seguindo links direcionados
 - Componentes fracamente conectados
 - ABCDE
 - GHF

 Em uma rede não direcionada utiliza-se simplesmente o termo 'componentes conectados'



Componente gigante

Se o maior componente engloba uma fração significativa do grafo, este é chamado de componente gigante.

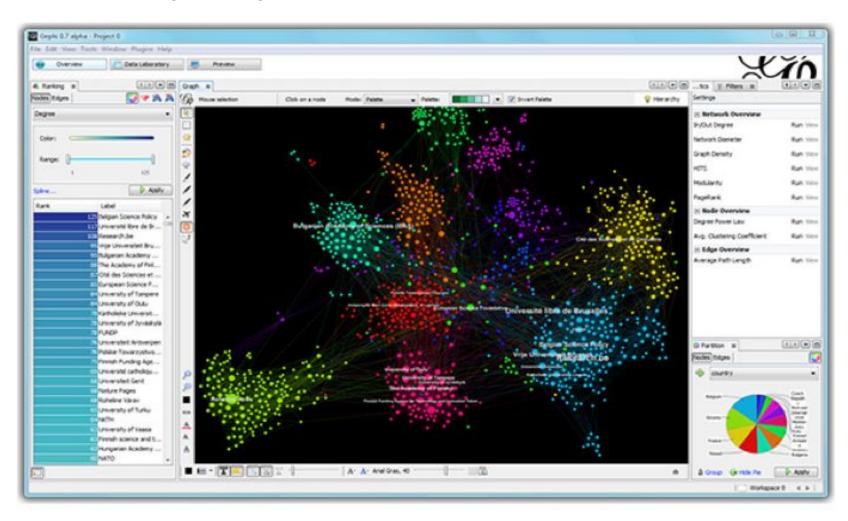


Softwares de ARS

- Pajek (http://pajek.imfm.si/doku.php)
- UCINet (https://sites.google.com/site/ucinetsoftware/)
- NodeXL (http://nodexl.codeplex.com/)
- NetLogo (https://ccl.northwestern.edu/netlogo/)
- Gephi (https://gephi.org/)
- SNA package for R

Gephi

Link: http://gephi.org/



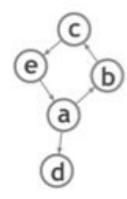
Por que Gephi?

- Um grande comunidade
- Java
- Plug-ins
 - Baseado na Netbeans Platform
- Excelente desempenho (OpenGL)
- Importação de Bancos de Dados Relacionais
 - Postgres, SQL Server, MySQL
- Qualidade gráfica
- Capacidade de lidar com milhares de nodos

Suporte a múltiplos formatos

	a	b	C	d	e
a	-	1		1	-
b	-	-	1	-	-
c	-	-		-	1
d	-	-			-
e	1	-		-	-

Tabular



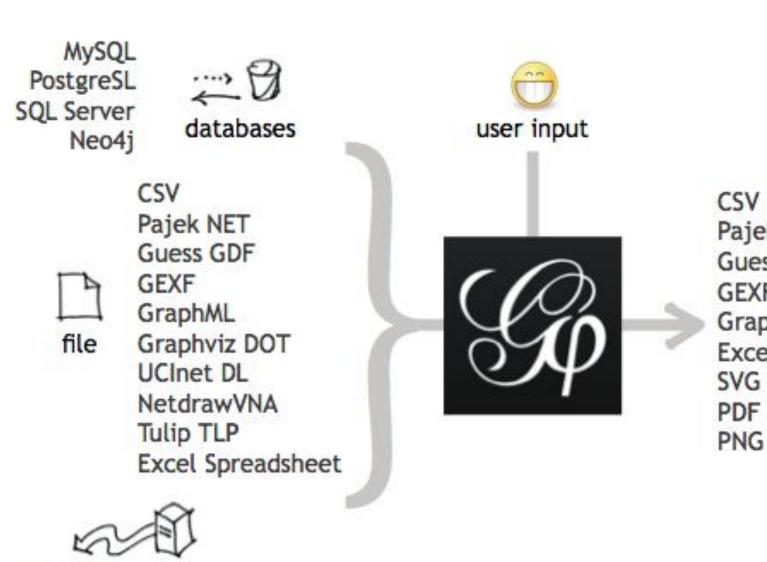
Graphical

```
<graph>
     <nodes>
       <node id="a" />
       <node id="b" />
       <node id="c" />
       <node id="d" />
       <node id="e" />
     </nodes>
     <edges>
       <edge source="a" target="b" />
       <edge source="a" target="d" />
       <edge source="b" target="c" />
       <edge source="e" target="a" />
       <edge source="c" target="e" />
     </edges>
</graph>
```

XML

and many others...

Fluxo



CSV
Pajek NET
Guess GDF
GEXF
GraphML
Excel Spreadsheet
SVG
PDF

graph streaming

Biblioteca de ARS

- networkx
- Escrita em Python
- https://networkx.github.io/

Referências

- SCOTT, J. Social Network Analysis. A Handbook. 2nd edition. SAGE Publications: London, 2000.
- WASSERMAN, Stanley; FAUST, Katherine. Social Network Analysis: methods and applications. Cambridge University Press. Structural analysis in social the social sciences series, v. 8, (1994) 1999. 857 p. ISBN 0-521-38707-8.
- HANNEMAN, R. A.; RIDDLE, M. (2005). Introduction to social network methods. Riverside, California: University of California, Riverside. Disponível em: http://faculty.ucr.edu/~hanneman/nettext/
- Freeman, L. 1979. Centrality in Social Networks: Conceptual Clarification. Social Networks, 1(3), 215–239.
- EMIRBAYER, M.; GOODWIN, J. Network analysis, culture and the problem of agency. American Journal of Sociology, v. 99, n. 6, p. 1411-1454, 1994.
- Parte do material foi baseado no trabalho de Lada Adamic.

Bons Estudos!