# Ciência das Redes Centralidades

Ricardo Luders Thiago H Silva



#### Centralidades



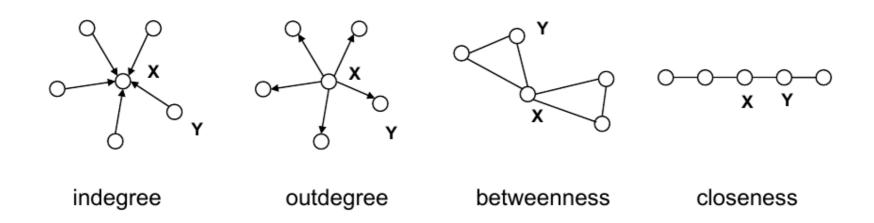
#### Descobrir qual é o nó mais central é importante:

- Pode ajudar a disseminar informações na rede mais rapidamente
- Pode ajudar a parar epidemias
- Pode ajudar a proteger a rede contra quebras

### Centralidades



#### Centralidade pode ter vários significados:



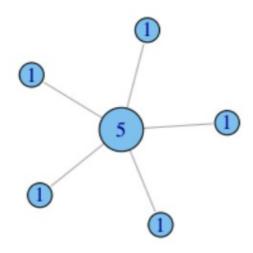
### Centralidade de grau



Quando o número de conexões é a melhor medida de centralidade?

- pessoas que farão favores a você

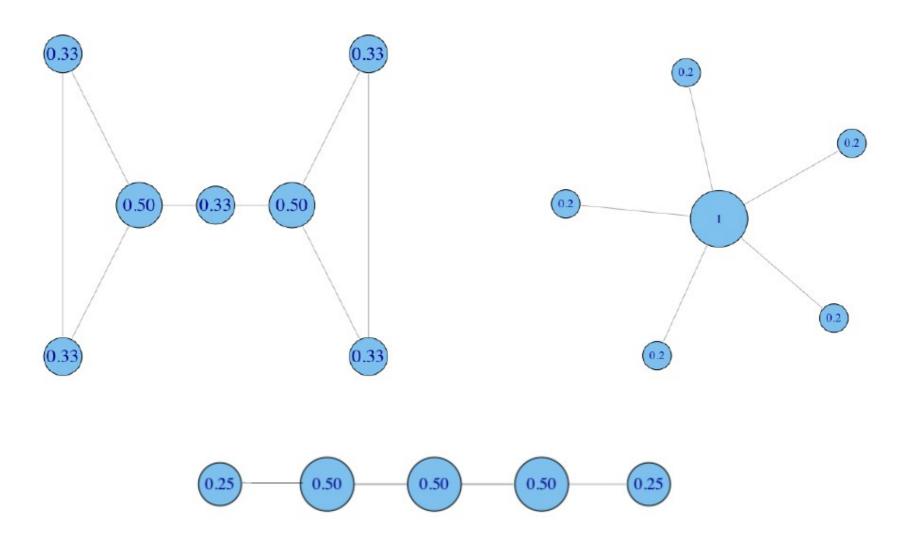
- pessoas com quem você pode conversar / tomar uma cerveja



## Normalização



Dividir pelo número máximo de nós (N-1)

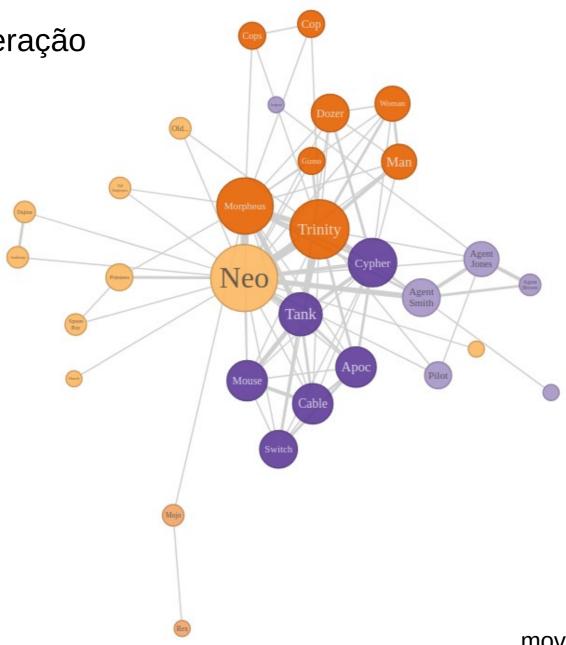


## Centralidade de grau



Rede de interação de atores

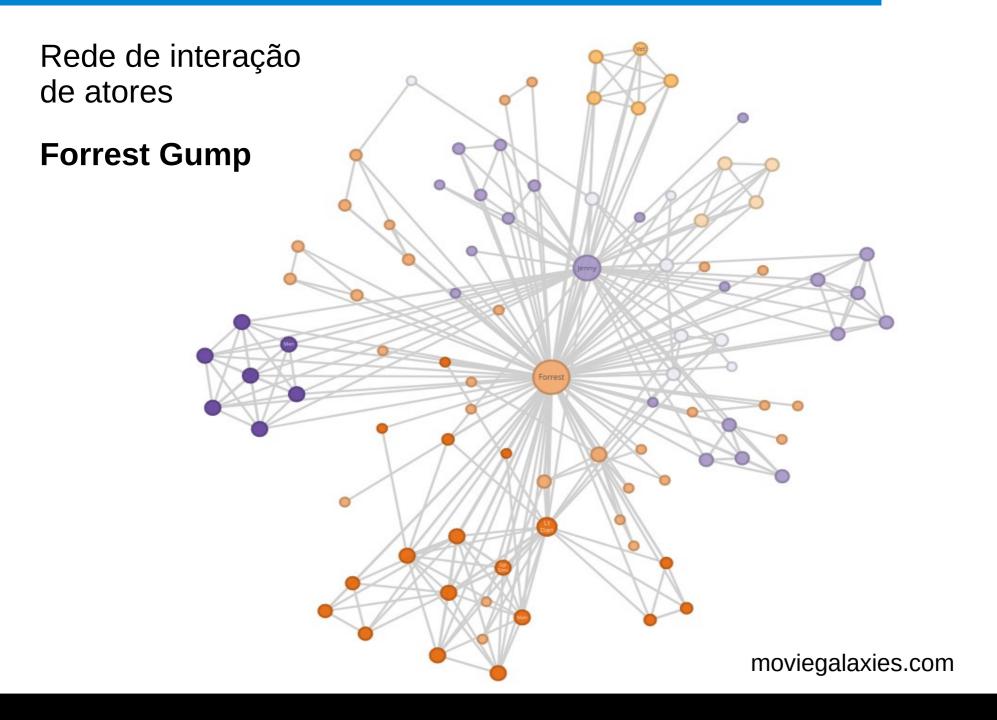
#### **Matriz**



moviegalaxies.com

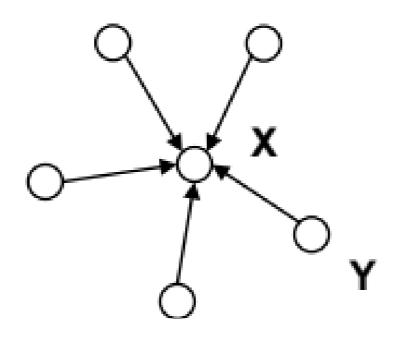
## Centralidade de grau





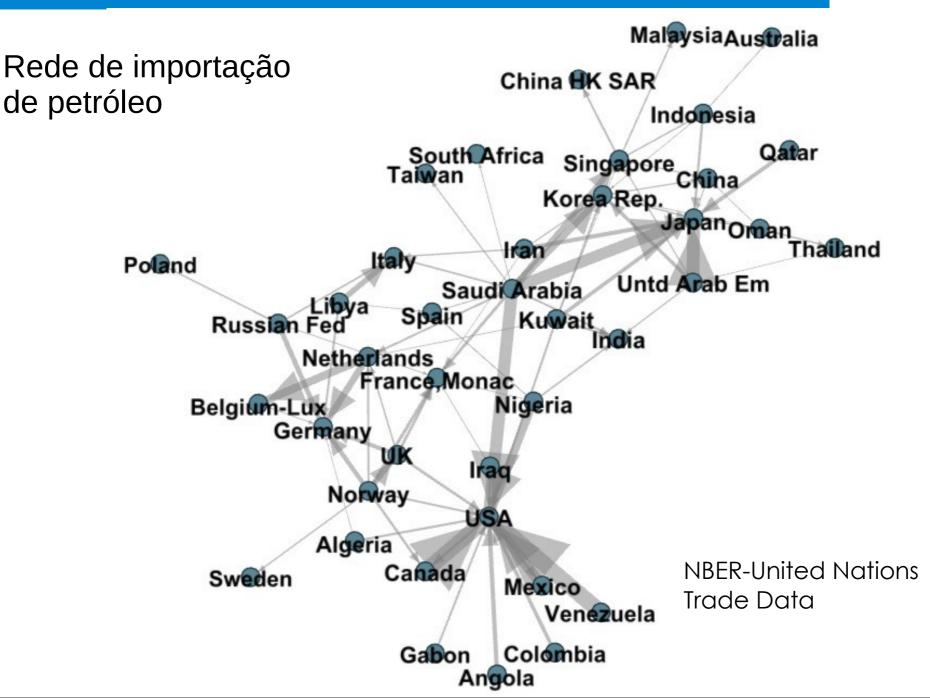
## Grau de entrada





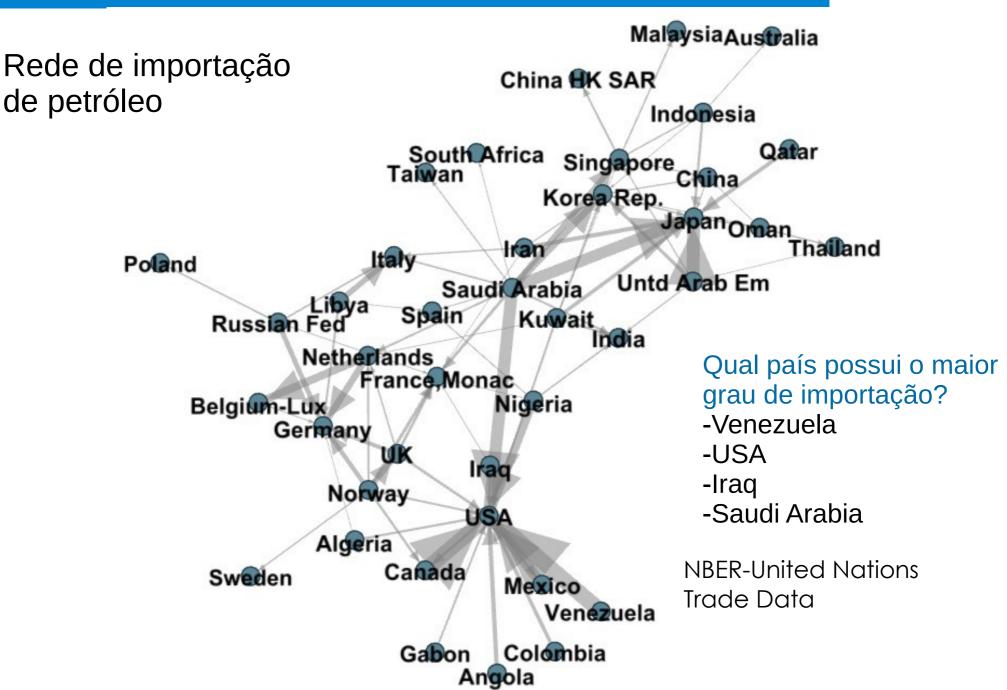
#### Grau de entrada



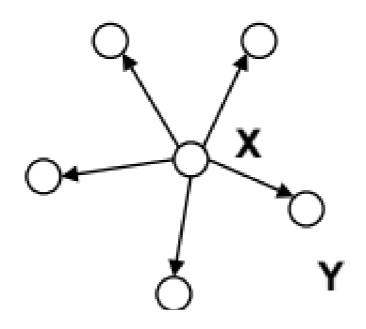


#### Grau de entrada





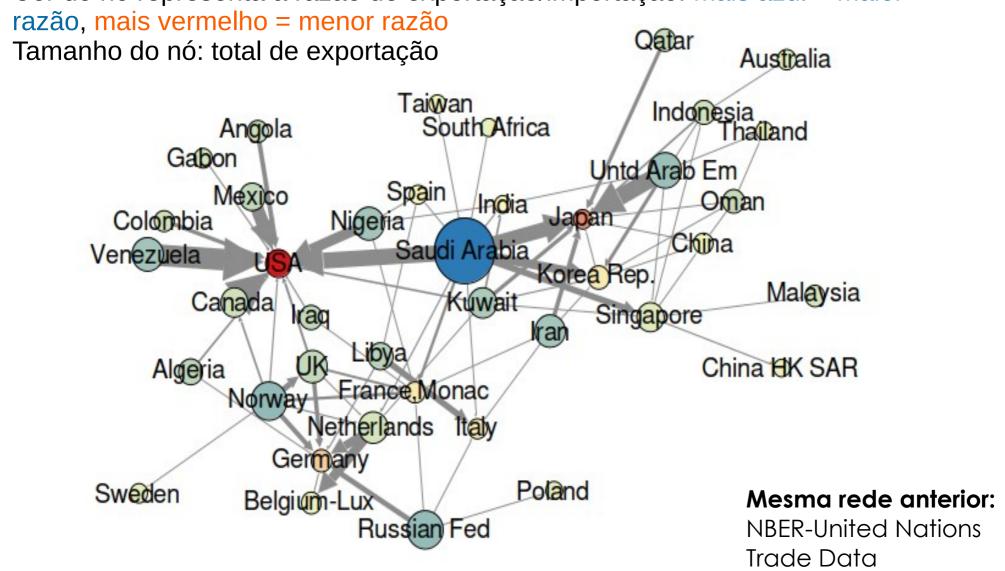






Largura da aresta representa \$

Cor do nó representa a razão de exportação/importação: mais azul = maior





Qual país tem um grau de saída baixo mas exporta uma quantidade significante?

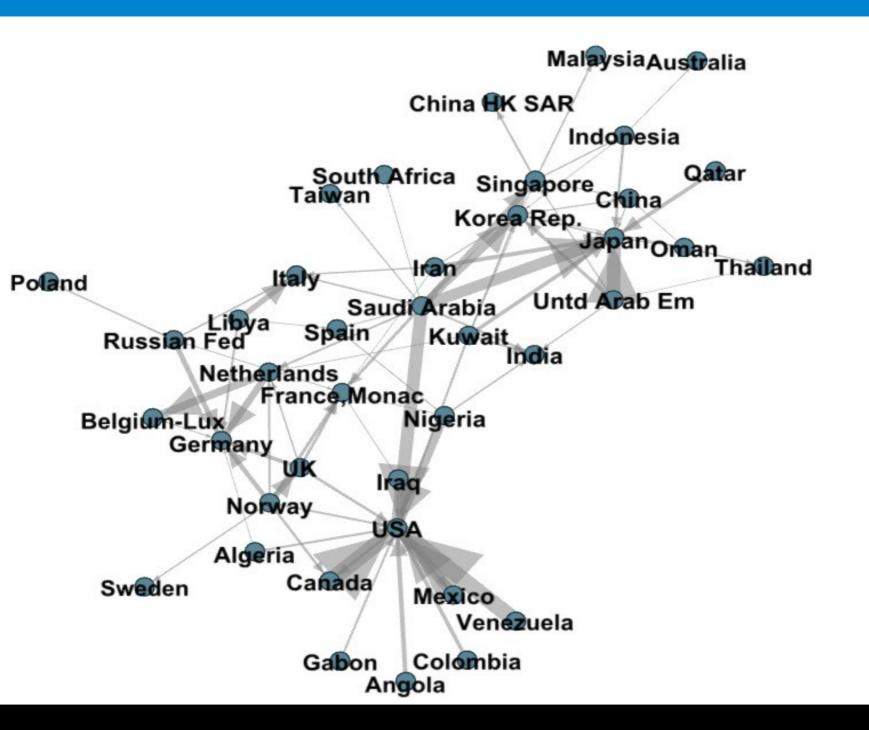
- Arábia saudita
- Japão
- Iraq
- USA
- Venezuela



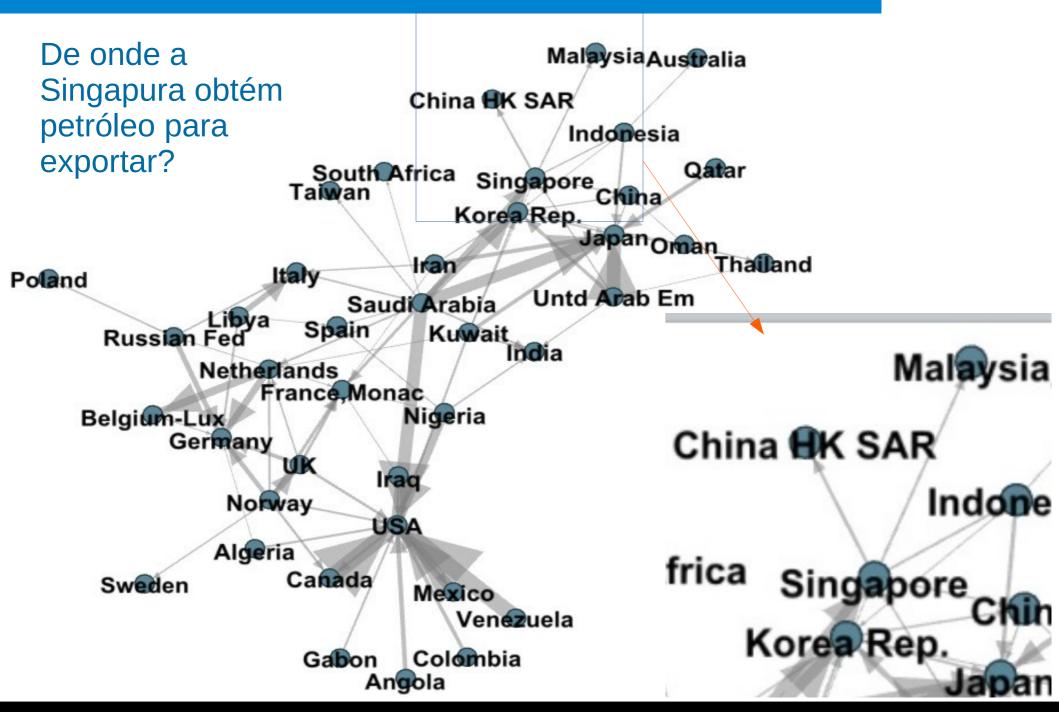
Qual país tem um grau de saída baixo mas exporta uma quantidade significante?

- Arábia saudita **→**
- Japão
- Iraq
- USA
- Venezuela

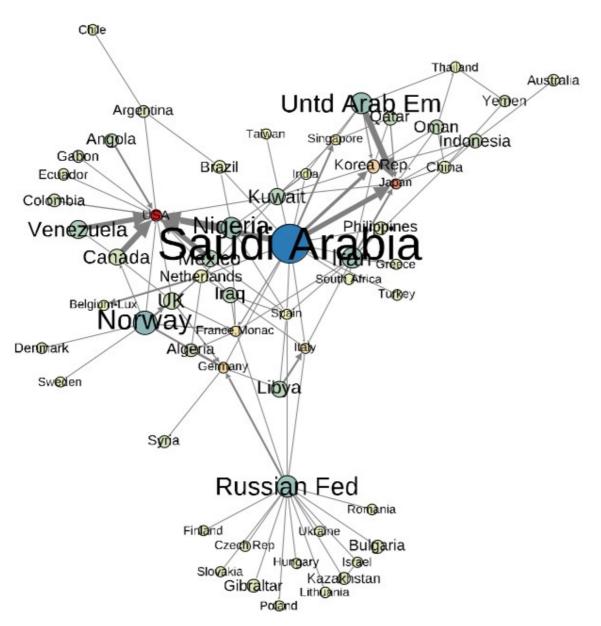












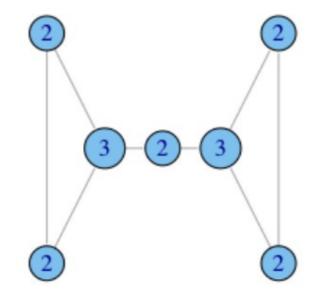
Rede de exportação considerando somente o material bruto

Singapura agora é primariamente um importador

#### Quando a Centralidade de Graus não é tão boa?





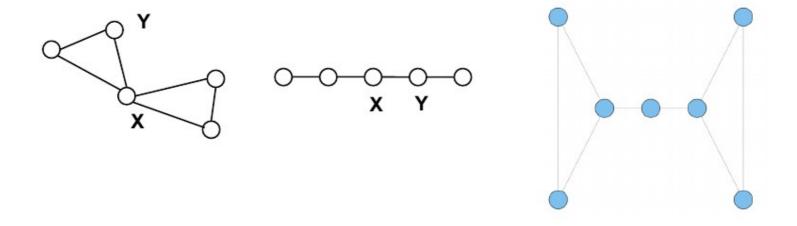


- Capacidade de intermediar entre grupos
- Probabilidade de que as informações originadas em qualquer lugar da rede cheguem até você ...



**Intuição:** quantos pares de indivíduos teriam que passar por você para se atingirem com um número mínimo de saltos?

Quem tem intermediação mais alta, X ou Y?





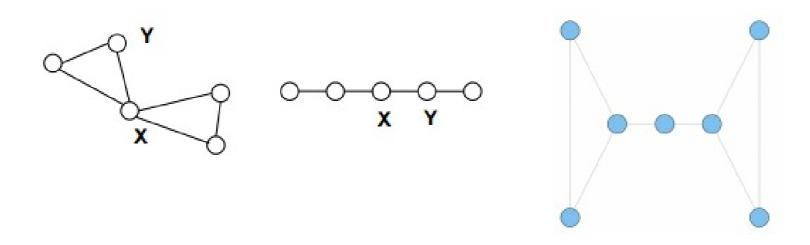
$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

 $\sigma_{st}$  é o número total de menores caminhos do nó s para o nó t

 $\sigma_{st}(v)$  é o número de menores caminhos que passam por v

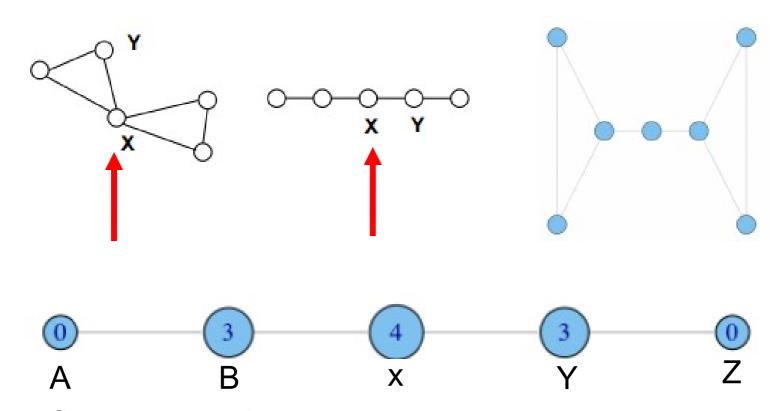


Quem tem o maior valor de intermediação x ou y?





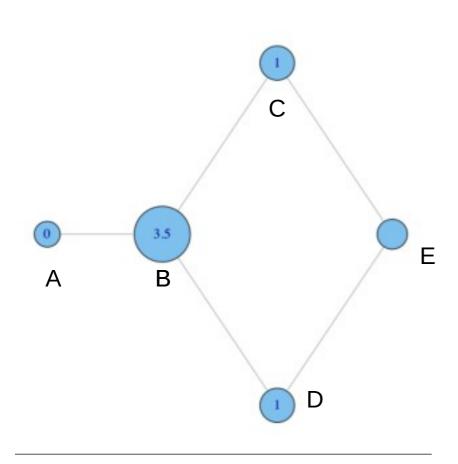
#### Quem tem o maior valor de intermediação x ou y?



A não fica entre nenhum outro par B fica entre (A-X), (A-Y), (A-Z) X fica entre (A-Y), (A-Z), (B-Y), (B-Z)

Note que não há caminhos alternativos para esses pares tomarem, então x recebe todo o crédito





Sem normalização

C e D possuem betweenness 1 porque ambos são caminhos mais curtos para (A,E) e (B,E), assim compartilham o crédito:

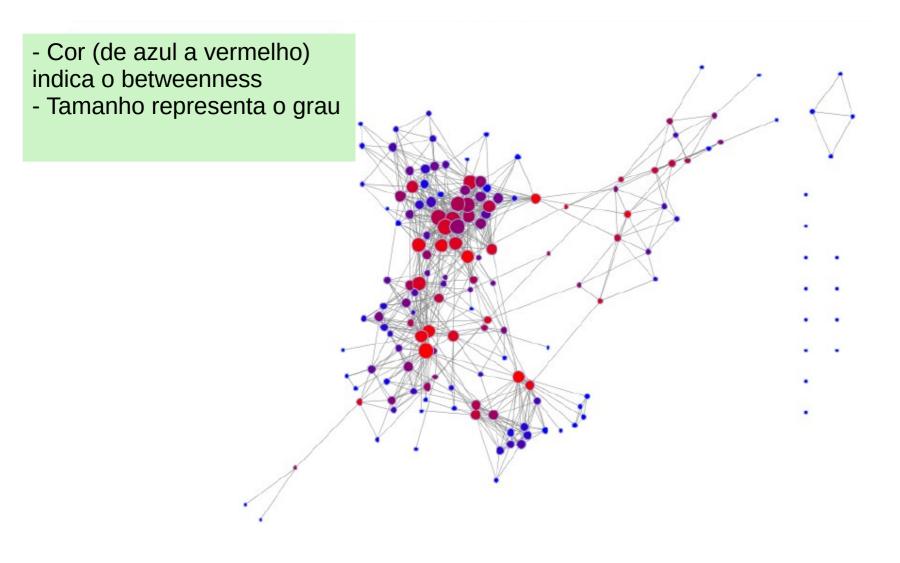
$$\frac{1}{2} + \frac{1}{2} = 1$$

Qual é o betweenness para E?

#### Intermediação / Betweenness centrality

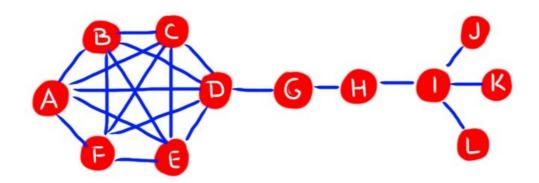


#### Exemplo do Facebook (acadêmico – um usuário específico)



#### Intermediação / Betweenness centrality





Encontre um nó que tenha alta intermediação, mas baixo grau Encontre um nó que tenha baixa intermediação, mas alto grau

#### Proximidade / Closeness centrality



- E se não for tão importante ter muitos amigos diretos?
- Ou estar "entre" outros

Mas ainda querer estar no "meio" das coisas, não muito longe do centro

#### Proximidade / Closeness centrality



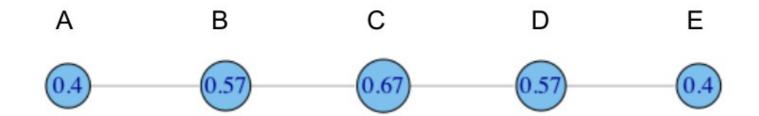
A centralidade de proximidade de um nó **u** é o recíproco da distância média do caminho mais curto para **u** em todos os **n-1** nós alcançáveis.

$$C(u) = rac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$$

onde d (v, u) é a distância do caminho mais curto entre v e u, e n é o número de nós que podem alcançar u.

### Proximidade / Closeness centrality - Exemplo

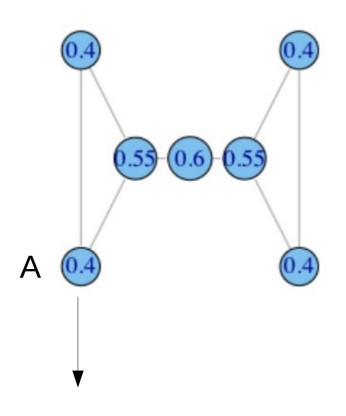


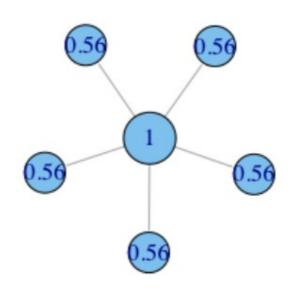


$$C(A) = 4 / (1+2+3+4) = 0,4$$

#### Proximidade / Closeness centrality - Exemplo



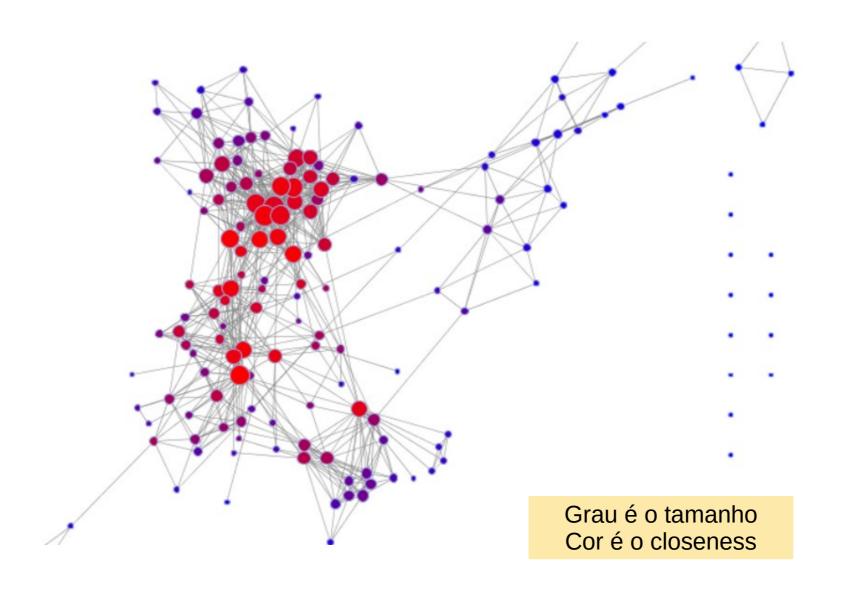




$$C(A) = 6 / (1+1+2+3+4+4) = 0,4$$

### Exemplo - Facebook

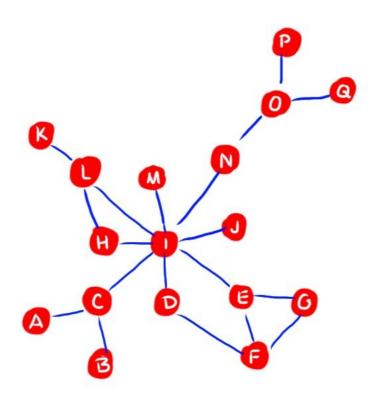




### Exemplo - Facebook



Qual nó tem grau relativamente alto, mas baixa proximidade?



#### Centralidade de autovetor (eigenvector) e variantes



Eigenvector centrality:

$$x_i = \sum_j A_{ij} x_j$$

Katz centrality:

$$x_i = lpha \sum_j A_{ij} x_j + eta$$

Page Rank centrality:

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

#### Autovetor / Eigenvector centrality



A centralidade de grau depende de haver muitas conexões: mas e se essas conexões estiverem bem isoladas?

Um nó central deve ser aquele conectado a nós poderosos

A centralidade de autovetor para o nó i é o i-ésimo elemento do autovetor x do maior autovalor  $\lambda$ , definido pela equação:

$$Ax = \lambda x$$

onde  $\bf A$  é a matriz de adjacência do grafo  $\bf G$  com autovalor  $\bf \lambda$ 

Em virtude do teorema de Perron-Frobenius, há uma solução única, cujas entradas são todas positivas, se  $\lambda$  for o maior autovalor da matriz de adjacência A

#### Autovetor / Eigenvector centrality



Suponha que x\_i é a centralidade de autovetor do nó i. A equação abaixo traduz a ideia de que o nó i tem maior centralidade se está conectado (vizinhos) a nós de maior centralidade de autovetor:

$$x_i' = \sum_j A_{ij} x_j$$
, ou na forma matricial:  $\mathbf{x'} = \mathbf{A} \mathbf{x}$ ,

Suponha um procedimento iterativo x(t)=A A ... (Ax); t vezes:  $\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0)$ .

Escrevendo x(0) como combinação linear dos autovetores:

$$\mathbf{x}(t) = \mathbf{A}^t \sum_i c_i \mathbf{v}_i = \sum_i c_i \kappa_i^t \mathbf{v}_i = \kappa_1^t \sum_i c_i \left[ \frac{\kappa_i}{\kappa_1} \right]^t \mathbf{v}_i$$
 autovetor i

Pergunta: x(t) converge para  $t \to \infty$ ?

Sim, sobra apenas o autovetor do maior autovalor:  $\mathbf{x}(t) 
ightarrow c_1 \kappa_1^t \mathbf{v}_1$ 

Logo, a centralidade de autovetor do nó i é proporcional ao autovetor do maior autovalor

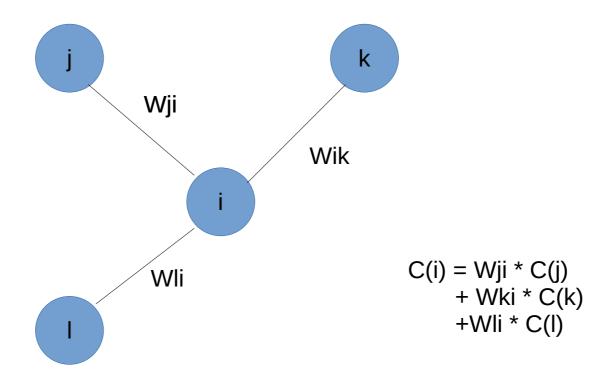
$$\mathbf{A}\mathbf{x} = \kappa_1\mathbf{x}.$$

major autovalor

#### Autovetor / Eigenvector centrality



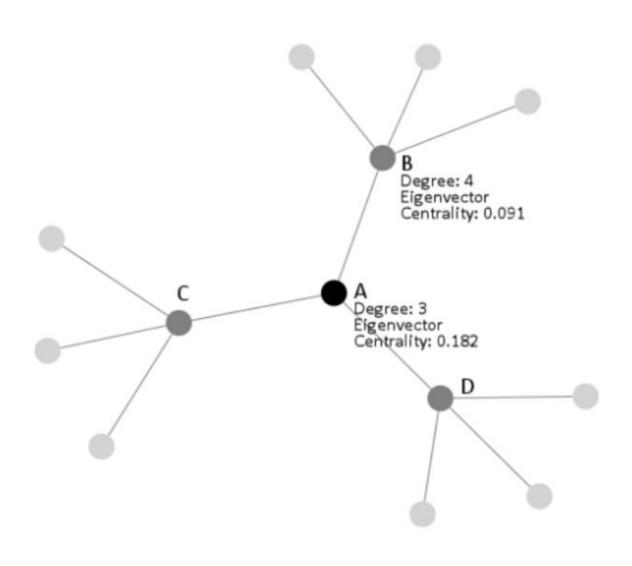
O quão central você é depende do quão central os seus vizinhos são



A ideia é repetir esse processo várias vezes até que não tenha alteração significativa

### Autovetor / Eigenvector centrality - Exemplo





#### Autovetor / Eigenvector centrality

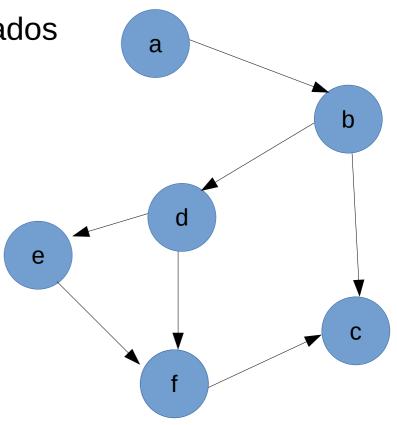


Devemos ter cautela com o uso em grafos direcionados

Funciona para grafos fortemente conectados

Problema para grafos fracamente conectados

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



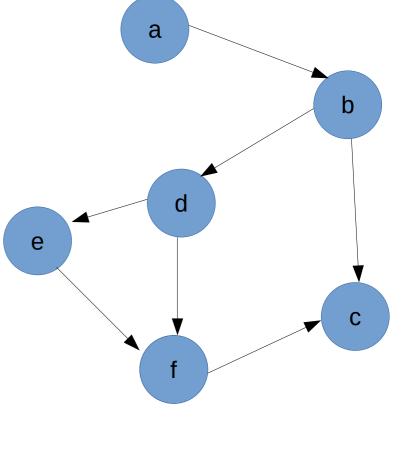
#### Power leak



#### **Calcular:**

Novo poder = soma do poder dos meus vizinhos.

	Iterações							
nó	0	1	2	3	4	5	6	
a	1							
b	1							
С	1							
d	1							
е	1							
f	1							



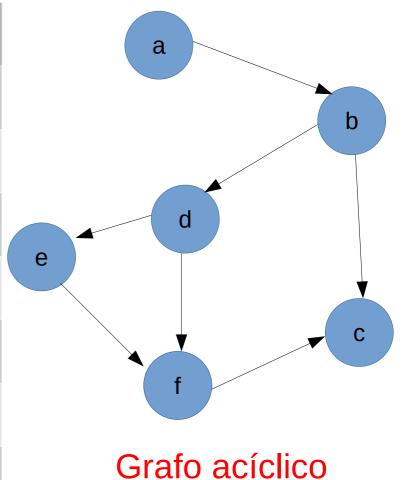
#### Power leak



#### Calcular:

Novo poder = soma do poder dos meus vizinhos.

	Iterações							
nó	0	1	2	3	4	5	6	
a	1	0	0	0	0	0	0	
b	1	1	0	0	0	0	0	
С	1	2	3	2	2	1	0	
d	1	1	1	0	0	0	0	
е	1	1	1	1	0	0	0	
f	1	2	2	2	1	0	0	



#### Power leak



#### **Calcular:**

Novo poder = soma do poder dos meus vizinhos.

		Iterações						a
nó	0	1	2	3	4	5	6	b
a	1	0	0	0	0	0	0	
	Vá	rias r	edes	reais	pos	suem	essa	a característica,
b								itações
								3
С	1	2	3	2	2	1	0	
d	1	1	1	0	0	0	0	- C
C.	_	_	_					
е	1	1	1	1	0	0	0	T
f	1	2	2	2	1	0	0	Grafo acíclico
								Graio aciciico

#### Adaptações para o Power leak – Katz centrality



Dar a cada vértice um pequeno valor "grátis". Precisamos escolher constantes de  $\alpha$  e  $\beta$ 

A katz centrality do nó i é:

$$x_i = lpha \sum_j A_{ij} x_j + eta$$
 Garante centralidade não zero

Termo de atualização do Eigenvector

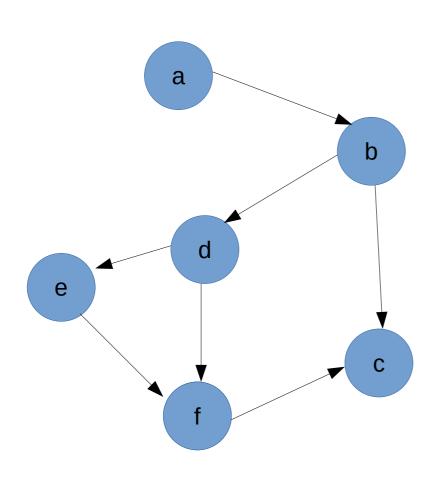
Por conveniência  $\beta=1$ 

$$lpha < rac{1}{\lambda_{ ext{max}}}$$

para que a centralidade de Katz seja calculada corretamente

# Katz centrality - exemplo





0,31 0,47 1 0,55 0,59

# Análise de link: PageRank centrality



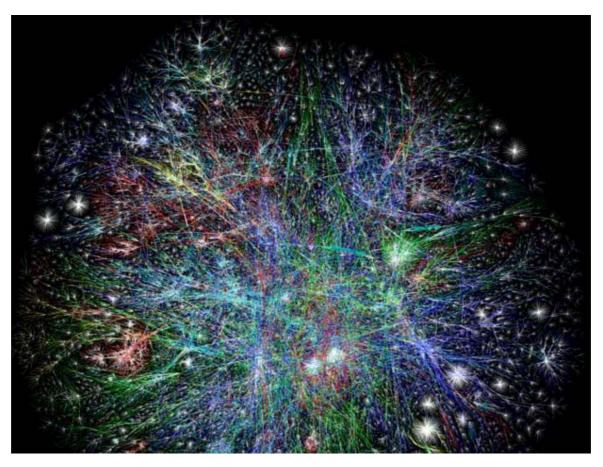
Antes de falar de Pagerank, vamos dar um passo para trás no tempo

### Como organizar a WWW?



A Web indexada contém pelo menos 3,82 bilhões de páginas (setembro de 2021)

Grande desafio é achar a informação que queremos (e a sua reputação!)



https://www.worldwidewebsize.com/

#### Na década de 1990



As pessoas estavam se acostumando com o novo mundo da Web







#### Primeira tentativa de organizar a Web



#### Vai de acordo com a vida antes da web



Catálogos telefônicos



Propaganda no Cadê? | Centro de Informações | Política de Privacidade

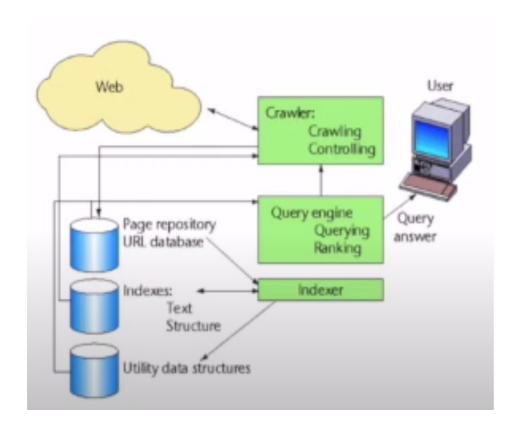
Informação organizada por humanos manualmente em diretórios Não escala! Uso de diretórios não é eficiente de usar.

#### Segunda tentativa de organizar a Web



Recuperação de informação

Similar ao modelo usado por bibliotecas



Existem problemas.

Em um livro, se tem muitas menções a Chico Buarque, então é provavelmente de interesse de pessoas interessadas em Chico Buarque.

Web é competitiva por atenção!

Na Web, indivíduos podem fazer "mutretas" para gerar atenção para seus sites, como introduzir conteúdo escondido, etc.

#### Terceira tentativa de organizar a Web

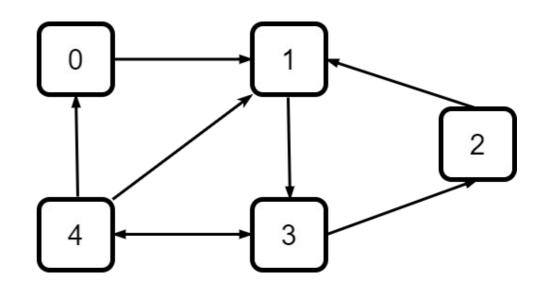


Como determinar se uma página Web é de boa reputação?

Um link é implicitamente um atestado de qualidade para a página alvo

Ranquear as páginas usando a estrutura de links no grafo da Web

Links de páginas "importantes" devem contar mais



#### PageRank Centrality





Copyright @1998 Google Inc.

#### Fruto de uma tese de doutorado

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30(1-7), 107-117.

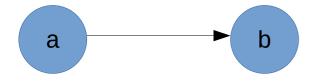
#### PageRank Centrality



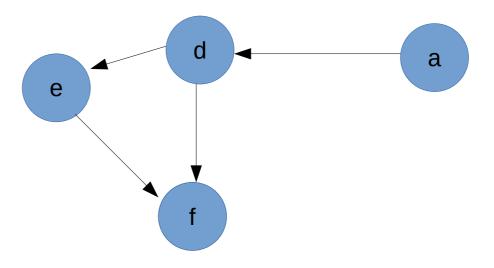
#### Simples modelo de votação:

- Uma página Web é a soma dos votos dos seus links de entrada
- Cada voto por um link é proporcional à importância de sua página
- A importância da fonte é dividida igualmente entre as suas páginas alvo

#### **Problemas**



Sink vértices (votos vazam)



Armadilhas (previnem convergência)

#### PageRank Centrality



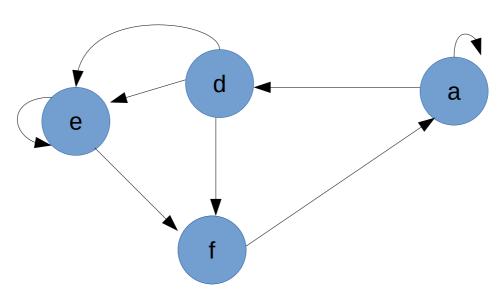
Em vez de assumir apenas um sistema de votação simples

PageRank assume a existência de um surfista aleatório

Que clica em links aleatoriamente

As vezes fica entediado e vai para outra página aleatória (teleportação)

A probabilidade de que o surfista aleatório visite uma página é o seu PageRank

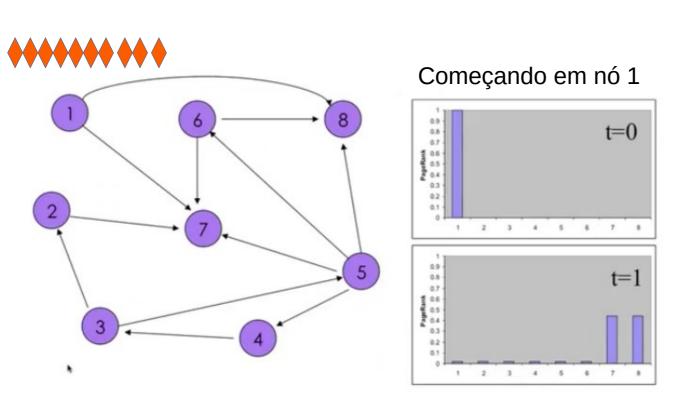


### PageRank Centrality - Exemplo



Situação provável no tempo 1

Probabilidade de teletransporte = 0,2



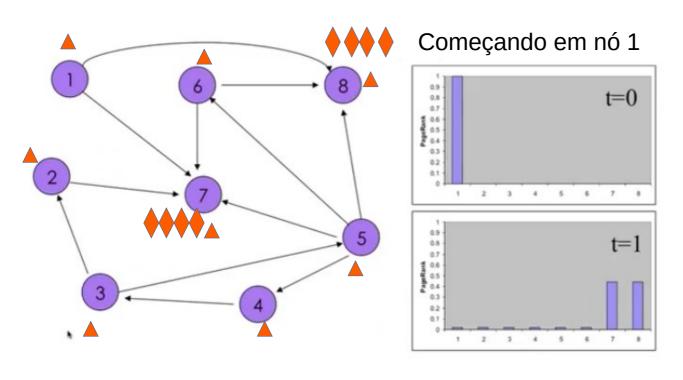
0,8 de probabilidade vai ser distribuída entre 8 e 7, o resto vai para os outros.

#### PageRank Centrality - Exemplo



Situação provável no tempo 1

Probabilidade de teleporte = 0,2



0,8 de probabilidade vai ser distribuída entre 8 e 7, o resto vai para os outros.

#### PageRank Centrality - Exemplo



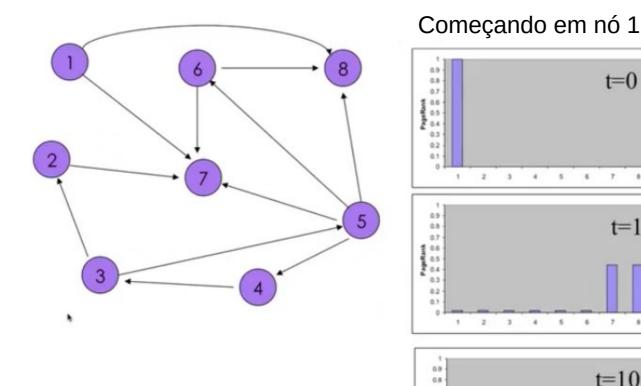
Situação provável no tempo 1

Probabilidade de teleporte = 0,2

t=0

t=1

t = 10



0,8 de probabilidade vai ser distribuída entre 8 e 7, o resto vai para os outros.

Faz sentido o 7 ser mais importante?

### PageRank intuição



Google aponta para milhões de páginas. Isso faz com que todas as páginas sejam importantes?

#### Solução:

Uma página deve compartilhar uma fração de sua importância para cada vizinho.

Dividir pelo grau de saída durante a atualização

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

### PageRank intuição



Google aponta para milhões de páginas. Isso faz com que todas as páginas sejam importantes?

#### Solução:

Uma página deve compartilhar uma fração de sua importância para cada vizinho.

Dividir pelo grau de saída durante a atualização

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

### PageRank intuição



Google aponta para milhões de páginas. Isso faz com que todas as páginas sejam importantes?

#### Solução:

Uma página deve compartilhar uma fração de sua importância para cada vizinho.

Dividir pelo grau de saída durante a atualização

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$
 
$$(1 - \alpha) \frac{1}{\tau}$$
 Ideia do teletransporte

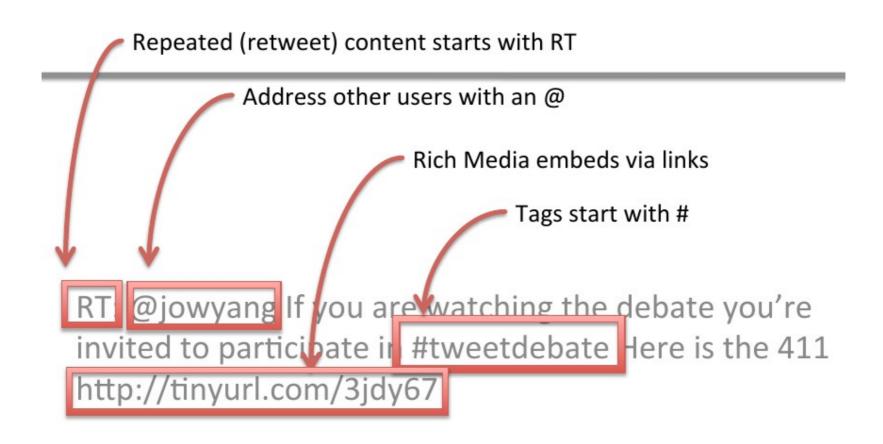
# Exemplo geral



#### As pessoas "tweetam" enquanto assistem

The second second	
re you doing?	140
ac mini = :-) – boxee about 18 hours ago	update
Raratequila @BigRed561 oh nah, where is dat? less than 20 seconds ago from web in reply to BigRed561	
tabligherooz VPS از 12.000توبان/اکیک هاست 12.000 از 12.000 انوبان/اکیک هاست less than 20 seconds ago from FriendFeed	ff.im/GahAB
jon_cohen @henryb35 no thanks less than 20 seconds ago from web in reply to henryb35	
AndreJayP @HASstyle420 @duecebigz26 @andre @gogoquedawg @jrdafuture7 @silkybutmilky @cl #4thqtr- aint nuffin star @magicdoom04 ?? less than 20 seconds ago from UberTwitter	
lauramdempsey don't forget, kids; throw your si	upport behind
	tabligherooz VPS از 12.000 توبان/اکیک هاست 12.000 الوجان/اکیک هاست 12.000 الوک الوک الوک الوک الوک الوک الوک الوک





Anatomia de um tweet



Três hashtags: #current # debate08 #tweetdebate

97 minutos de debate + 53 minutos a seguir = 2,5 horas no total.

3.238 tweets de 1.160 pessoas.

- 1.824 tweets de 647 pessoas durante o debate.
- 1.414 tweets de 738 pessoas postam debate.

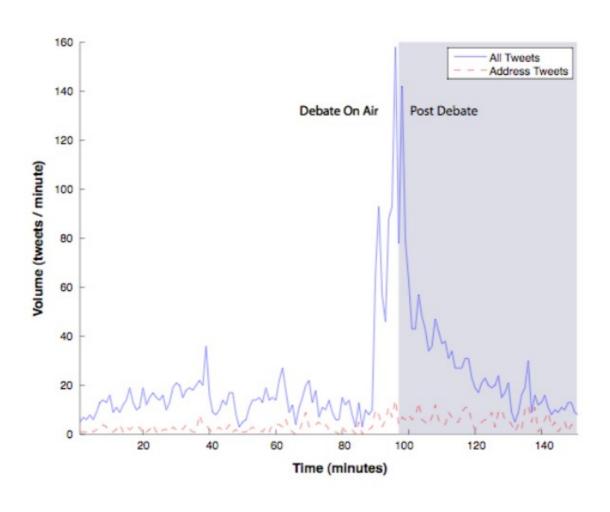
577 @ menções (reciprocidade!)

- 266 menções durante o debate
- 311 depois.

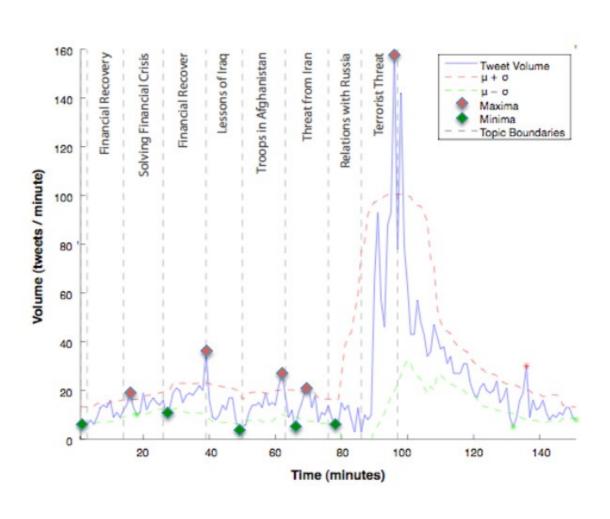
RT baixo: 24 retuítes no total

- 6 durante
- 18 depois.



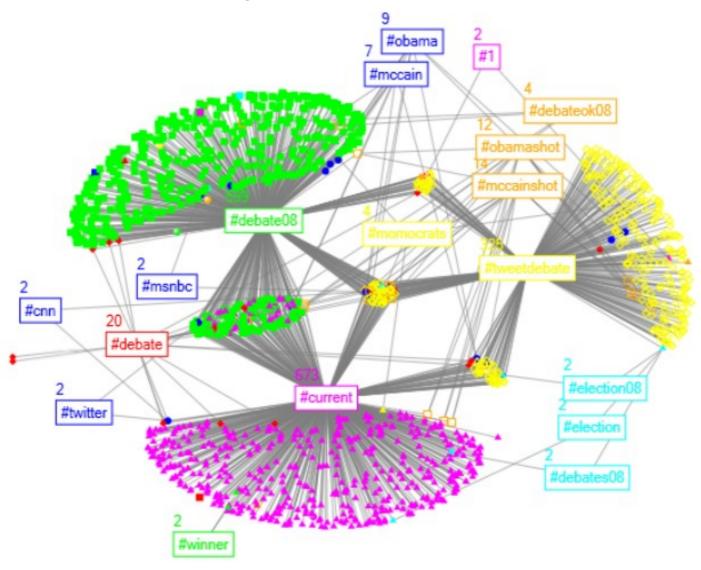






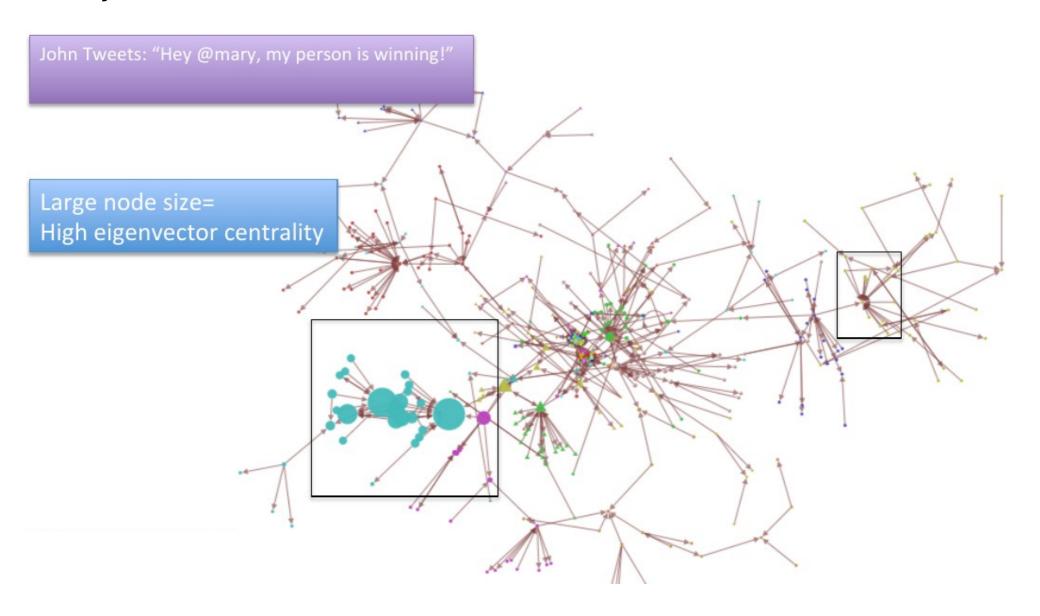


#### Clusters de usuários e tags





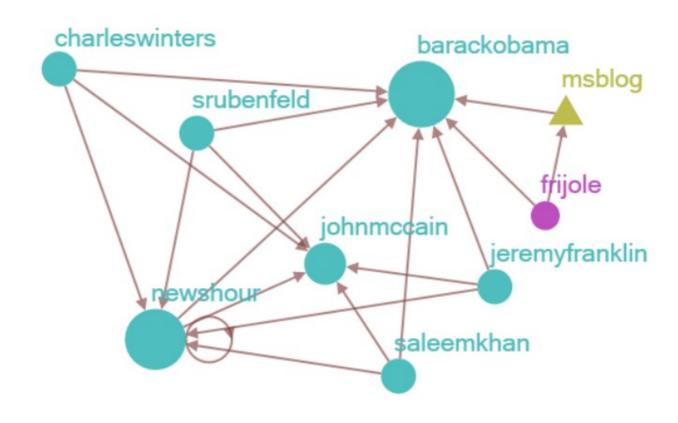
#### Menções no Twitter





# Descoberta automática através da centralidade

High Eigenvector Centrality Figures on Twitter from the First US Presidential Debate of 2008.





Twitter User	Eigenvector Centrality	In Degree	Out Degree
@barackobama	0.472	15	0
@newshour	0.427	11	5
@johnmccain	0.277	6	0
@charleswinters	0.223	0	3
@jeremyfranklin	0.223	0	3
@saleemkhan	0.223	0	3
@srubenfeld	0.223	0	3
@msblog	0.221	5	6
@frijole	0.175	0	7

#### Pontos finais



Várias métricas de centralidade (outras também não mencionadas)

O que o certa métrica traz de informação útil em uma determinada rede?

Em redes direcionadas: grau de entrada, grau de saída, page rank

#### Pontos finais



Várias métricas de centralidade (outras não mencionadas)

**Uma análise é importante:** o que o certa métrica representa no seu problema?

### Agradecimentos/referências



Kleinberg's book: Chapter 3.

Agradecimentos L. Adamic

M. Newmann. Networks. Oxford University Press. April 2010

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009, October). Tweet the debates: understanding community annotation of uncollected sources. In Proceedings of the first SIGMM workshop on Social media (pp. 3-10).