

PADRÕES LOCAIS DA REDE

Prof. Fabrício Olivetti de França
folivetti@ufabc.edu.br

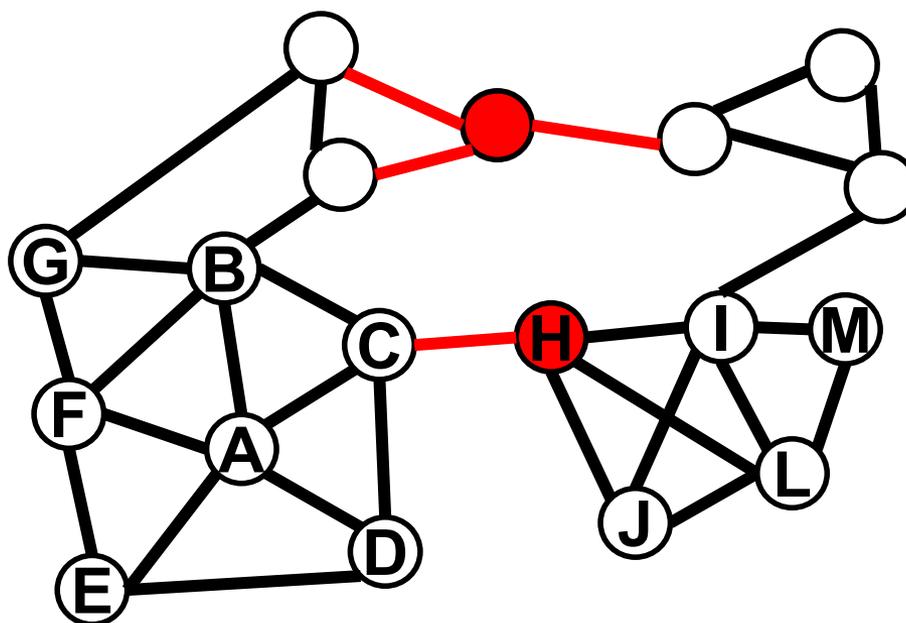


Universidade Federal do ABC

CENTRALIDADE

Importância dos nós

Até então vimos que cada nó da rede cumpre seu papel na transmissão de informação.



Sem eles a informação poderia **demorar mais para chegar** ao destino ou **nem mesmo chegar.**



Importância dos nós

Porém, alguns nós são mais importantes do que outros nessa transmissão!



Importância dos nós



Centralidade

Essa importância é mensurada através da **CENTRALIDADE** do nó, e ela indica:

- ❑ **Influência de uma pessoa** em uma rede social
- ❑ **Importância de certas avenidas** para o fluxo de veículos
- ❑ Qual a ordem de importância de **páginas web** para serem apresentadas no **resultado de busca**



Centralidade

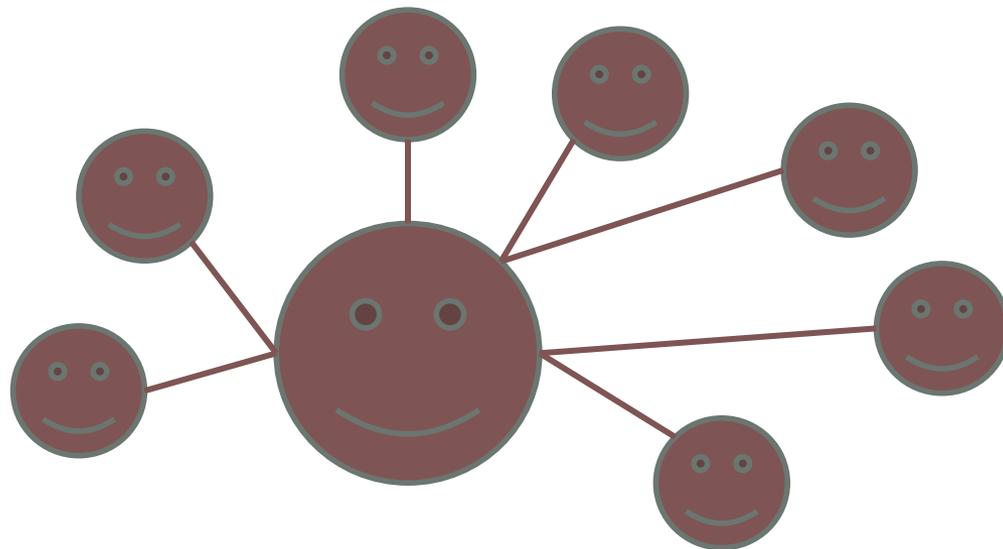
Dá para perceber que **identificar esses nós** de maior importância pode ser vital para a **estabilidade ou controle da sua rede.**

Também notem que a **forma com que essa importância é mensurada** pode ser diferente dependendo da aplicação.



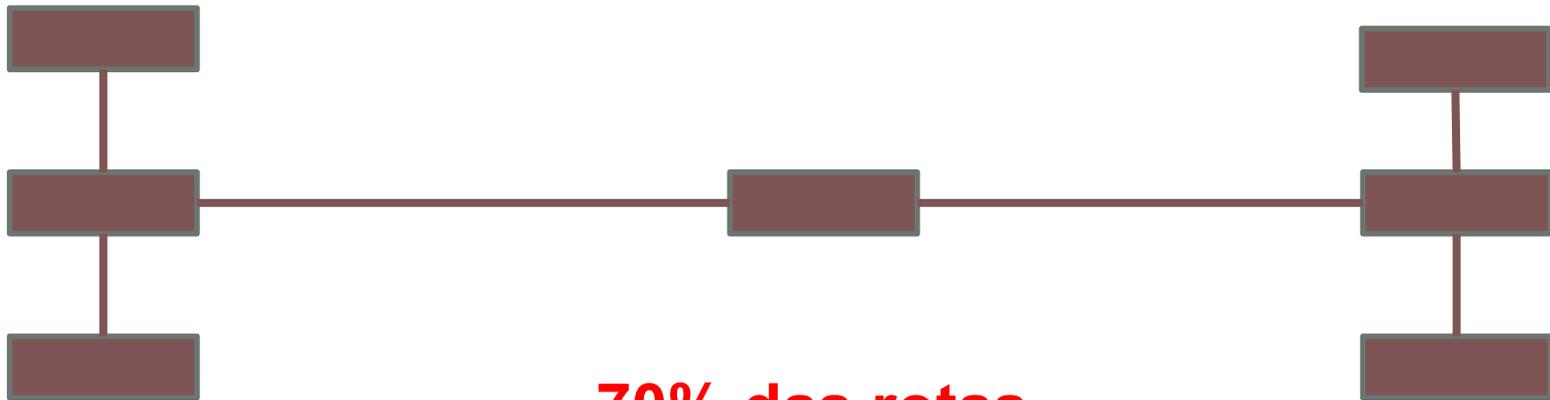
Exemplos

Em uma **rede social** o nó central é aquele que tem mais conexões, pois influencia mais pessoas ao mesmo tempo.



Exemplos

Em uma **rede de transporte** o nó importante é aquele pelo qual mais pessoas passam para ir de um local ao outro



**70% das rotas
passam por aqui!!**



Centralidade

A grande maioria das medidas de centralidades tem origem em **estudos de sociologia**.

Nessa aula veremos os conceitos mais conhecidos, as centralidades de:

- ❑ Grau
- ❑ Proximidade
- ❑ Betweenness
- ❑ Pagerank (utilizado pelo Google)



Centralidade de Grau

A ideia mais simples de um nó central é aquele que está **mais conectado que os demais**, ou seja, aquele que tem um maior grau.

Se ele tem mais conexões, significa que:

- ❑ Ele deve ser importante (para os **outros nós** **“desejarem” se conectar a ele**)
- ❑ Uma informação partindo dele é **transmitida diretamente para muito mais nós**



Centralidade de Grau

A centralidade de grau foi a **primeira medida** de centralidade já utilizada (e é também a **mais simples**)

Simplesmente é calculada como a relação do grau do nó com o valor máximo de grau que tal nó poderia ter:

$$C_g(v_i) = \text{grau}(v_i) / (|V|-1)$$

Grau máximo possível: número total de nós menos um (exceto quando é possível conectá-lo a ele mesmo)



Centralidade de Grau

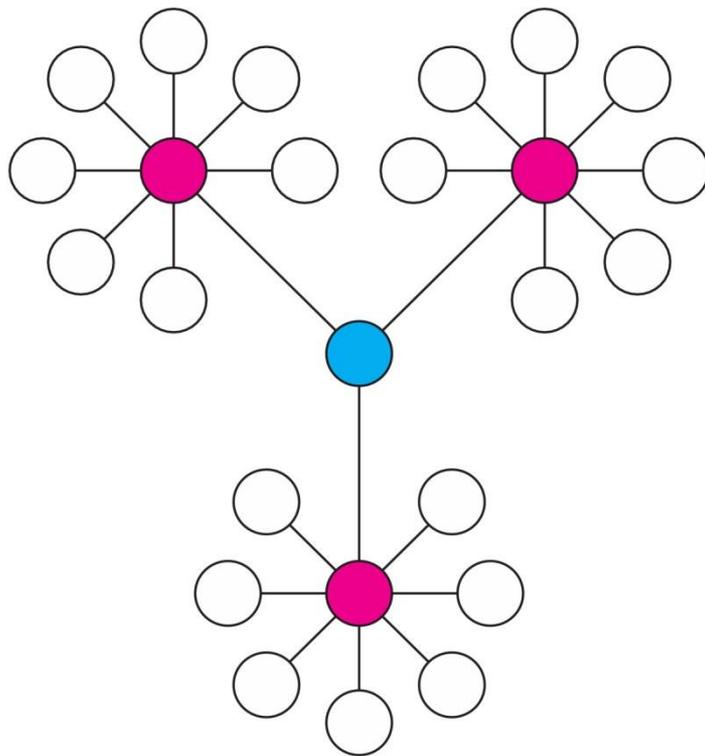
Em resumo, essa medida pode ser interpretada como “o quão importante os outros nós julgam que ele é” e/ou **“por quantas vias o nó pode receber ou emitir informação”**

Aplicações:

- ❑ Quantas pessoas a **opinião de uma certa pessoa atinge**, de acordo com seu número de amigos;
- ❑ Quantas pessoas alguém doente vai **infectar**;
- ❑ Qual **roteador** um hacker deve atacar para parar um serviço.



Centralidade de Grau



$C_g = 0,042$



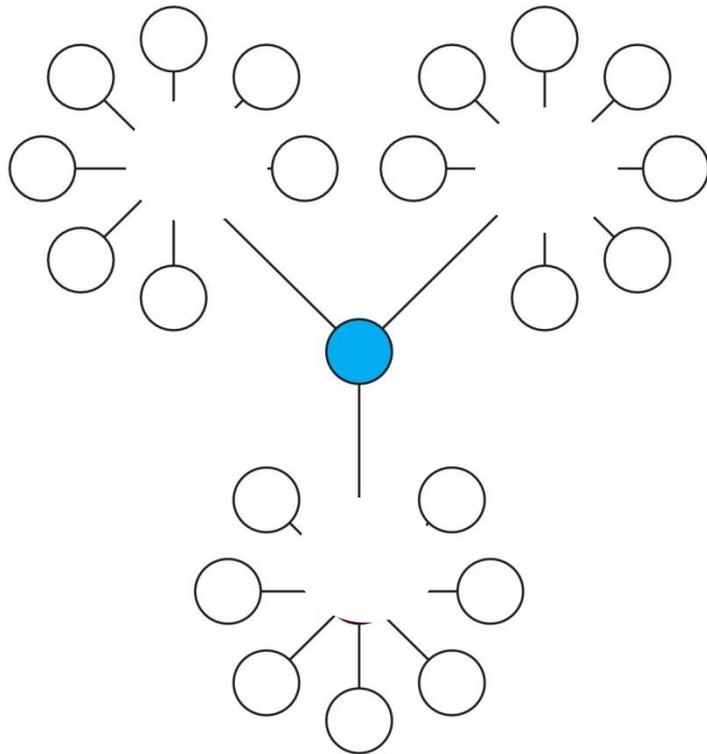
$C_g = 0,333$



$C_g = 0,125$



Centralidade de Grau



$C_g = 0,042$



$C_g = 0,333$

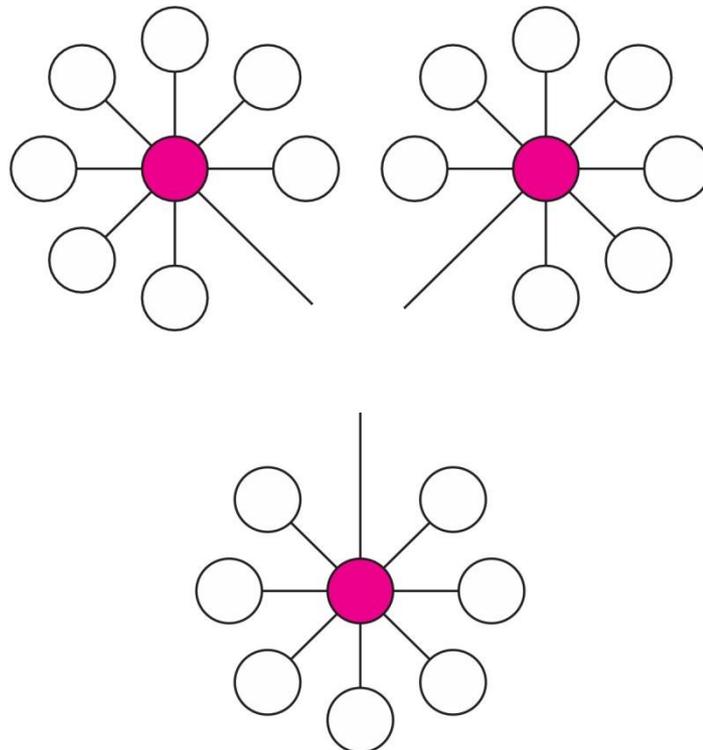


$C_g = 0,125$



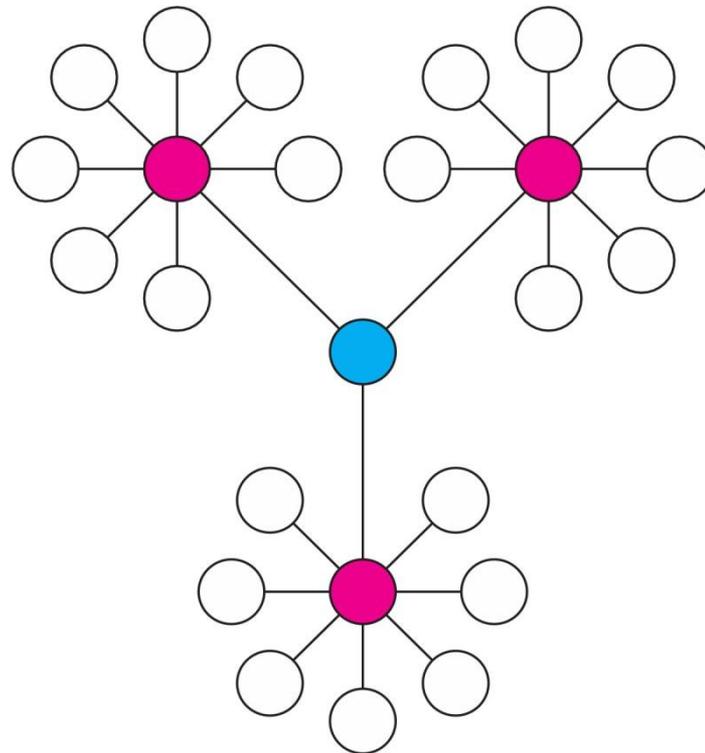
Considere Isso!

Se escolhermos atacar somente o nó azul, também desconectamos a rede



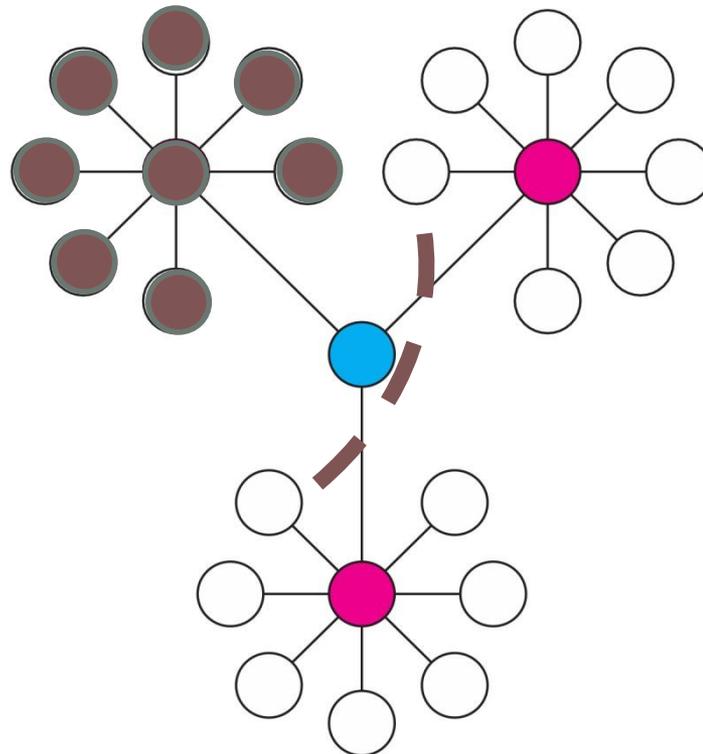
Considere Isso!

O envio de informação através do nó azul é, em média, mais rápida que a partir dos nós rosas



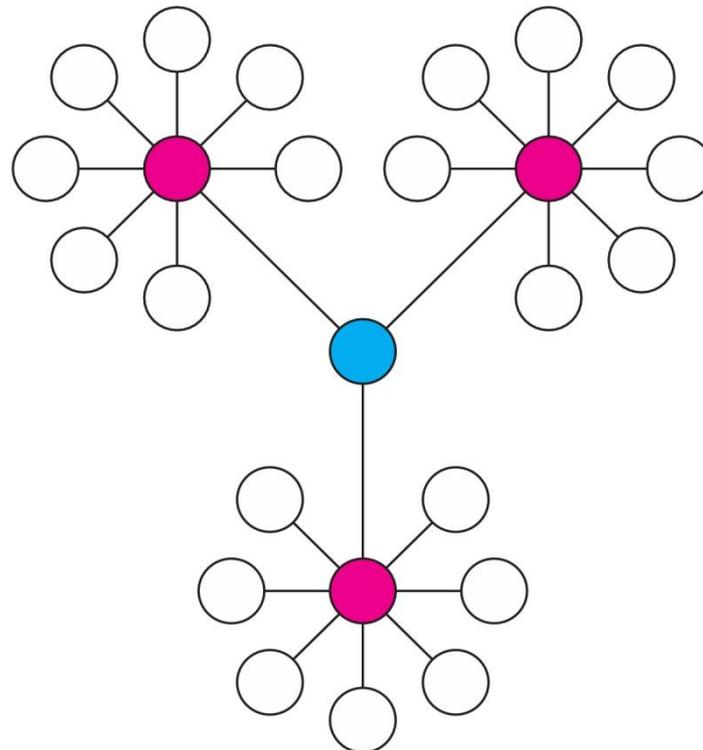
Considere Isso!

Talvez a melhor forma de lidar com a transmissão de uma doença seja isolando o nó azul

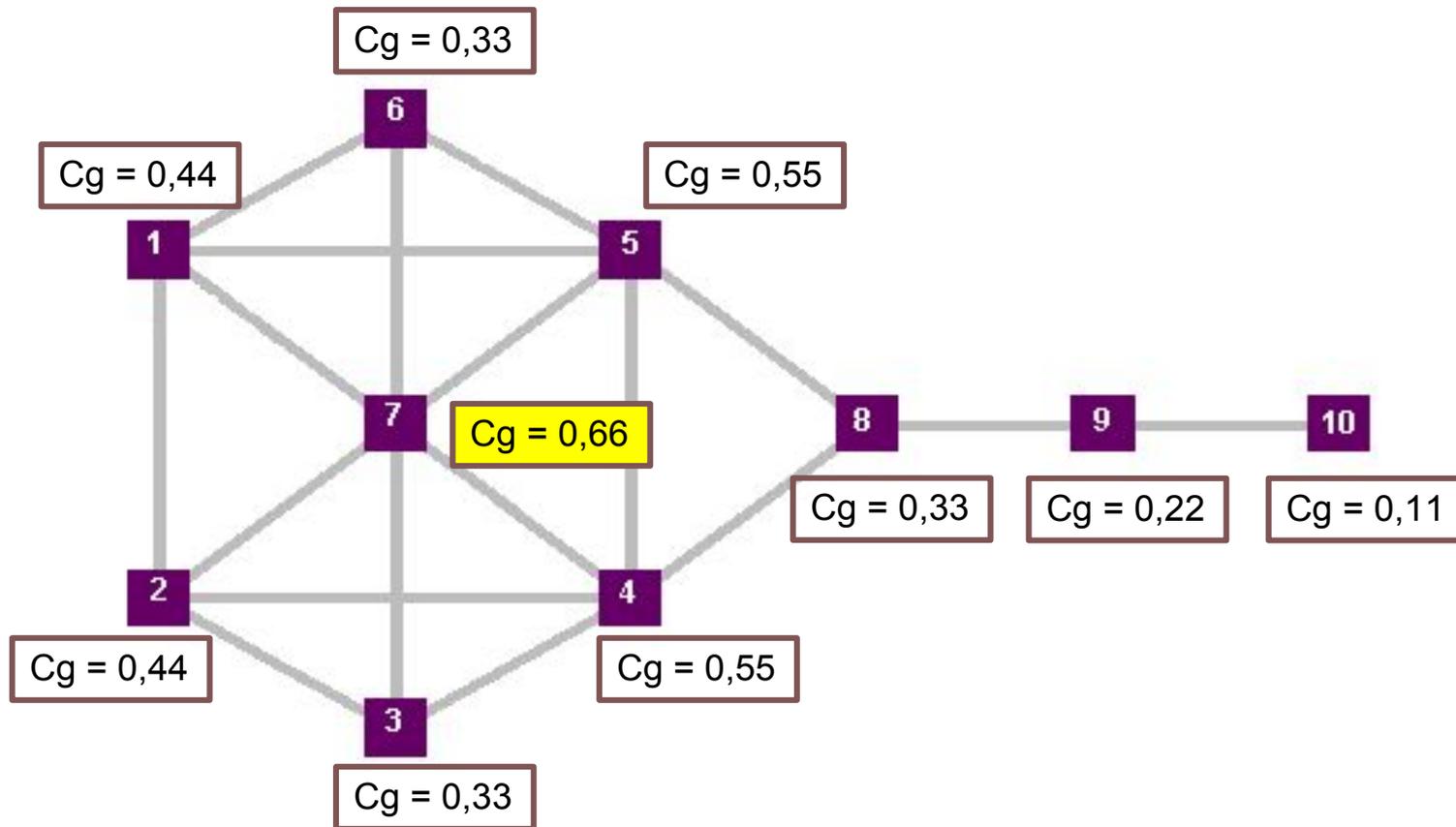


Considere Isso!

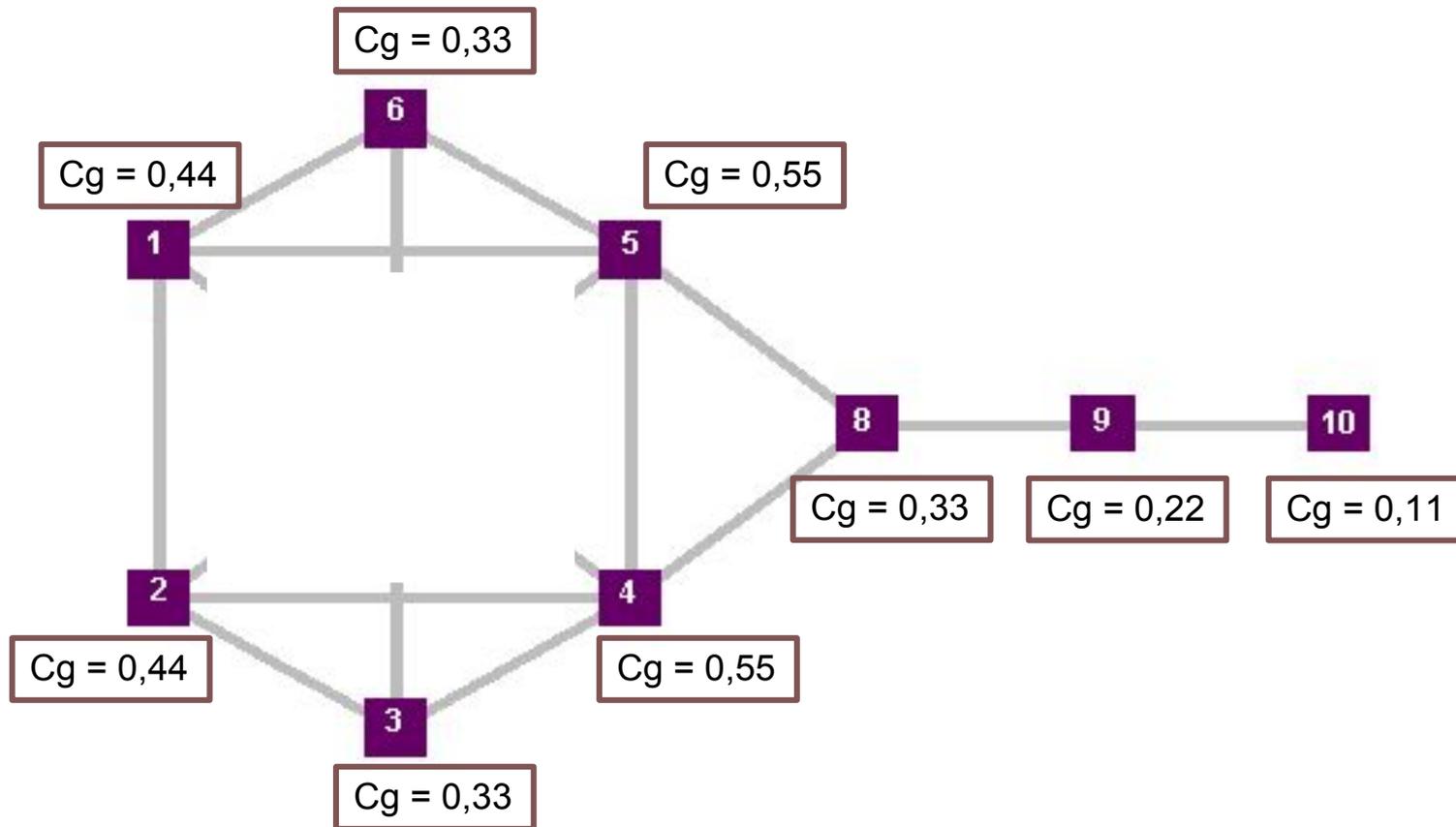
Por outro lado, os nós rosas tem vantagem de escolha de múltiplos caminhos para seu objetivo, pode ser um concentrador de riquezas!



Centralidade de Grau



Centralidade de Grau



Considere Isso!

A remoção do nó central **não desconectou** a rede!

Porém, a remoção dele pode causar um **congestionamento** de informação, pois ele era um nó “centralizador”, provavelmente com maior capacidade de receber e transmitir informação.



Centralidade de Proximidade

Uma outra forma de pensar em centralidade de um nó é imaginando que a rede representa um espaço físico e o nó mais central é aquele que se **localiza no centro da rede.**

Em um **espaço Euclidiano** o ponto central é aquele que minimiza a distância média entre todos os outros pontos (**medóide**).

Em uma rede podemos pensar que o nó central é aquele nó que minimiza a distância média entre todos os outros nós.



Centralidade de Proximidade

Um nó que tem a distância média menor pode **espalhar uma informação mais rapidamente** e de forma mais eficiente para todos os nós do que um nó que simplesmente tem muitas conexões.



Centralidade de Proximidade

Em uma rede a distância entre dois nós pode ser calculada pelo **caminho mais curto entre eles**

Para medir o quão **distante** um nó se encontra no grafo podemos somar a distância do caminho mais curto entre esse nó e todos os outros

Logo, para verificar a **proximidade** do nó, basta calcular a inversa dessa distância



Centralidade de Proximidade

A Centralidade de Proximidade de um nó é geralmente calculada como:

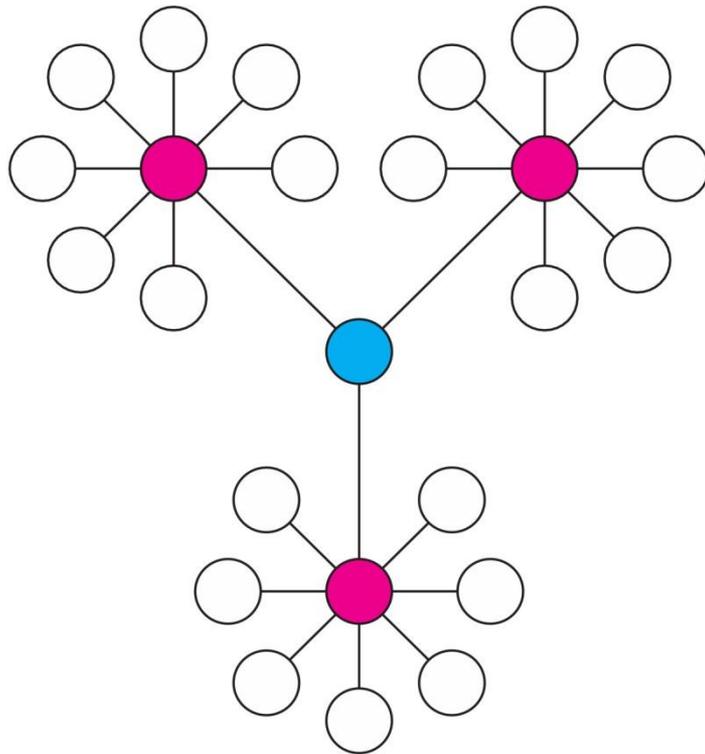
$$Cp(v_i) = \frac{n - 1}{\sum_{v_j \in V/v_i} d(v_i, v_j)}$$

Ela deve ser interpretada como “em quanto tempo a informação de um nó se propaga para todos os outros nós da rede”

Isso pode ser relevante tanto para ataque a uma rede (espalhar boatos, divulgar promoções) como defesa da rede (imunidade a vírus, ataque de hackers)



Centralidade de Proximidade



Cp = 0,31



Cp = 0,44



Cp = 0,53



Centralidade de Betweenness

Em alguns estudos o nó que temos interesse, e consideramos importante, é aquele que **mais contribui para o funcionamento da rede.**

Queremos saber qual nó causará maior impacto caso seja removido da rede.

Essa medida foi introduzida para verificar o **controle/influência de uma pessoa para transmitir uma informação entre grupos sociais diferentes.**

Freeman, Linton (1977). "A set of measures of centrality based upon betweenness". *Sociometry* **40**: 35–41.



Centralidade de Betweenness

Essa medida, conhecida como centralidade betweenness, mede a **“carga” de informação que um nó tende a receber**

Ela é calculada como a **quantidade de caminhos mais curtos da rede do qual esse nó faz parte**

Se a informação sempre viaja pelo caminho mais curto, quanto esse nó servirá de intermediário em diversas situações



Centralidade de Betweenness

○ cálculo é definido como:

$$Cb(v_i) = \sum_{v_j \neq v_i \neq v_k} \frac{\sigma_{v_j, v_k}(v_i)}{\sigma_{v_j, v_k}}$$

onde σ_{v_j, v_k} é o número de caminhos mais curtos entre esses nós e $\sigma_{v_j, v_k}(v_i)$ é o número de caminhos mais curtos entre esses nós que passa por v_i



Centralidade de Betweenness

$$Cb'(v_i) = \frac{Cb(v_i)}{(n-1)(n-2)}$$

Medida normalizada para comparação entre redes

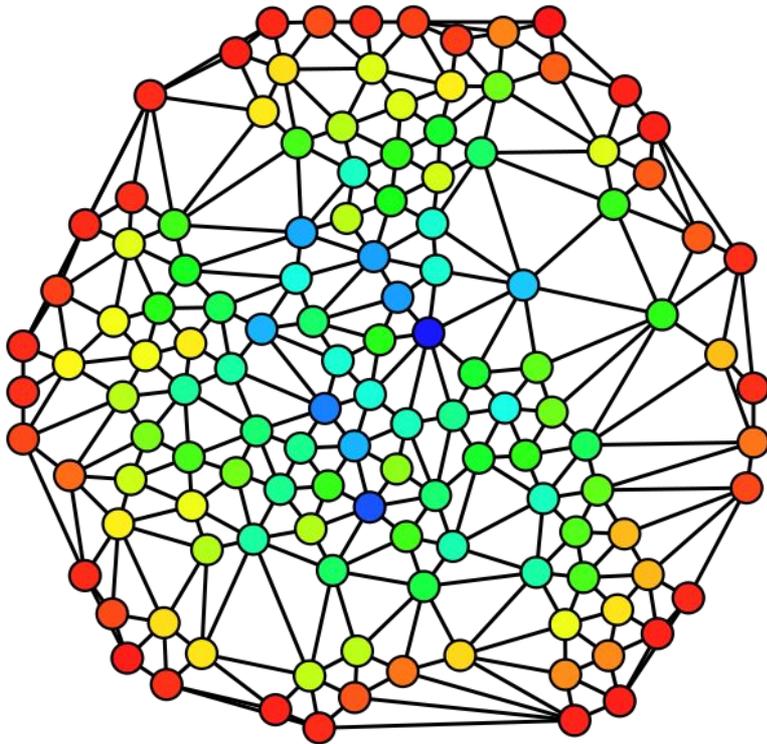
$$(n-1)(n-2)$$

Grau de betweenness do nó mais central em uma rede em formato estrela. Se não for direcionada, dividir por 2



Centralidade de Betweenness

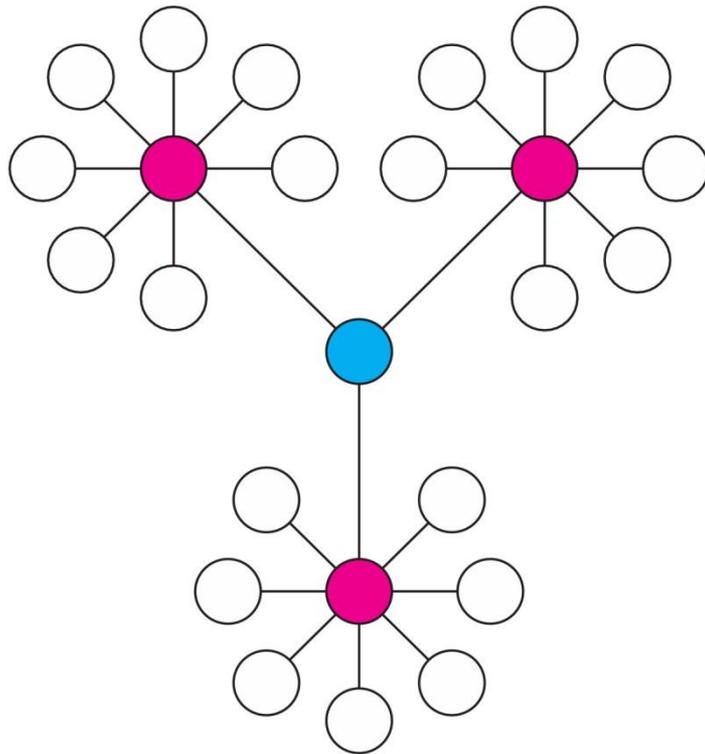
Essa métrica pode ser “custosa” para calcular pois necessita conhecer todos os possíveis caminhos mais curtos entre cada par de nós



Em redes com **ORDEM** e **TAMANHO** elevados, o valor dessa centralidade é estimado utilizando uma amostragem dos caminhos.



Centralidade de Betweenness



Cb = 0,00



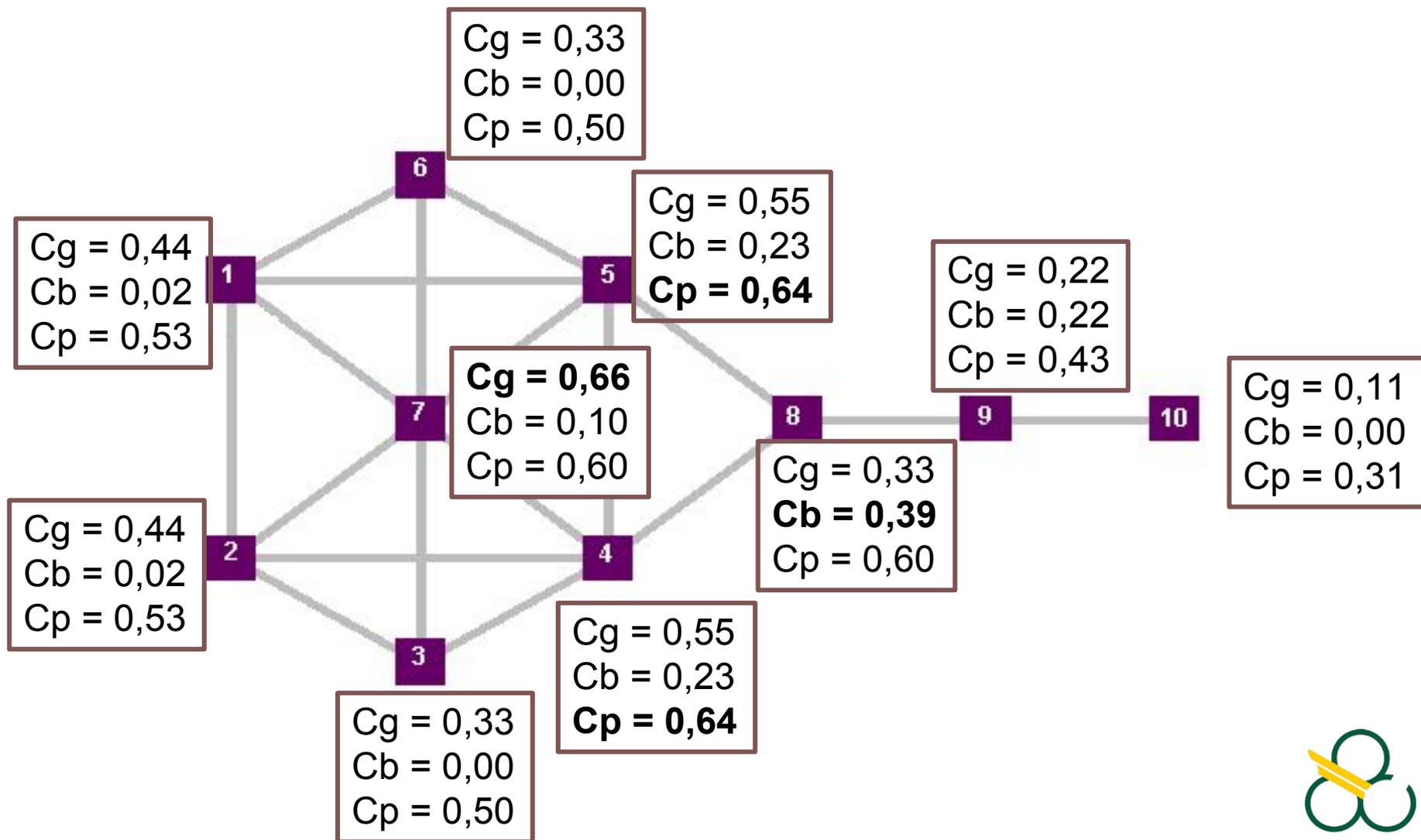
Cb = 0,51



Cb = 0,70

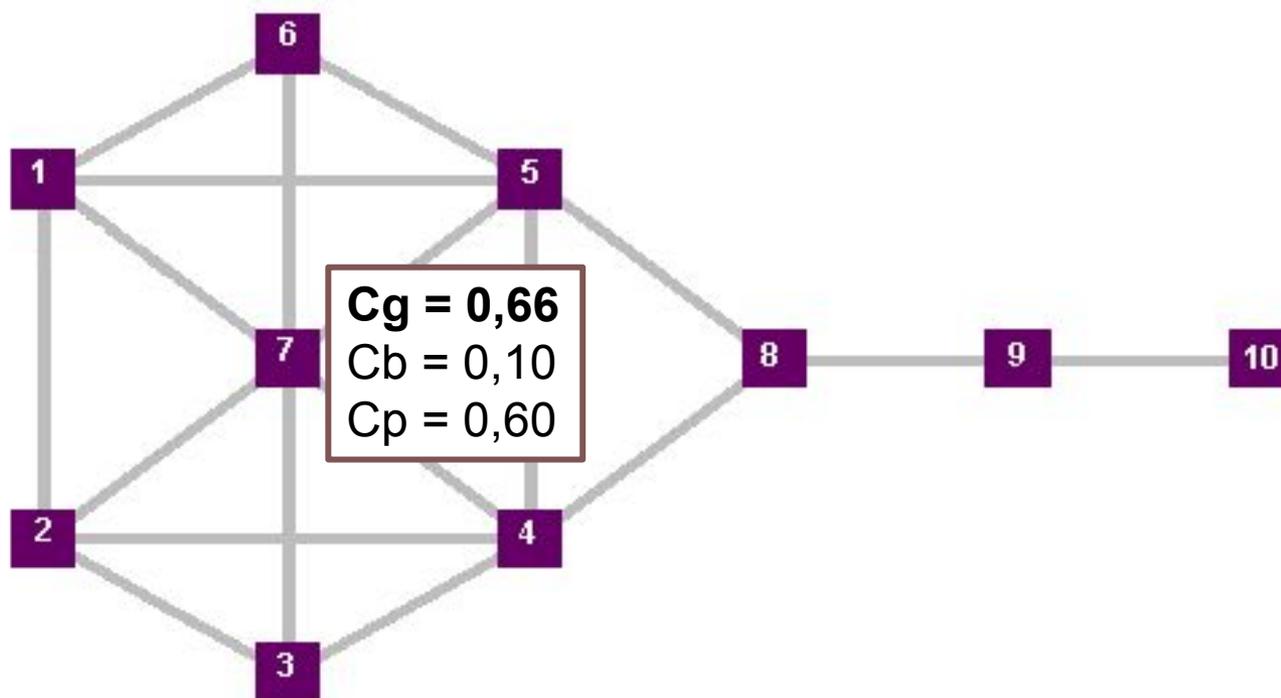


Centralidade até agora



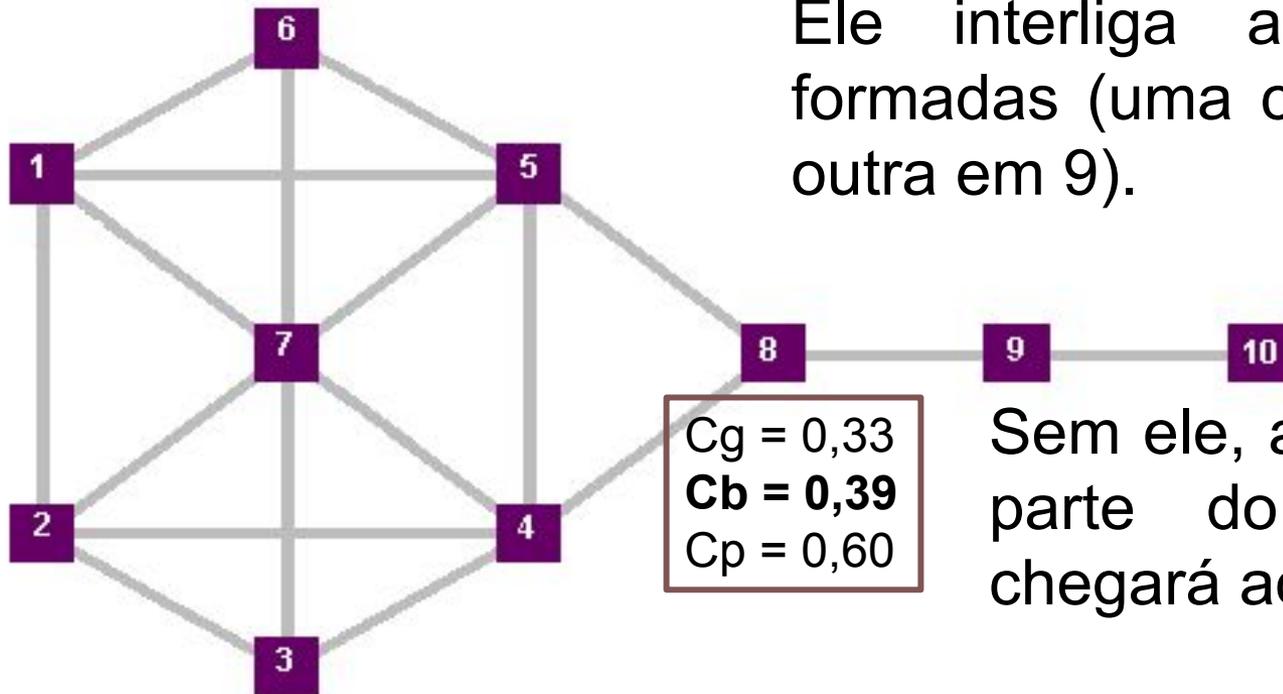
Centralidade até agora

O **nó 7 tem o maior número de conexões**, é o mais “popular” levando em conta a centralidade de grau, pois conhece mais pessoas.



Centralidade até agora

Porém, não é ele que está mais bem localizado na rede. **O nó 8, que tem maior grau de centralidade betweenness**, está em um ponto vital da rede.



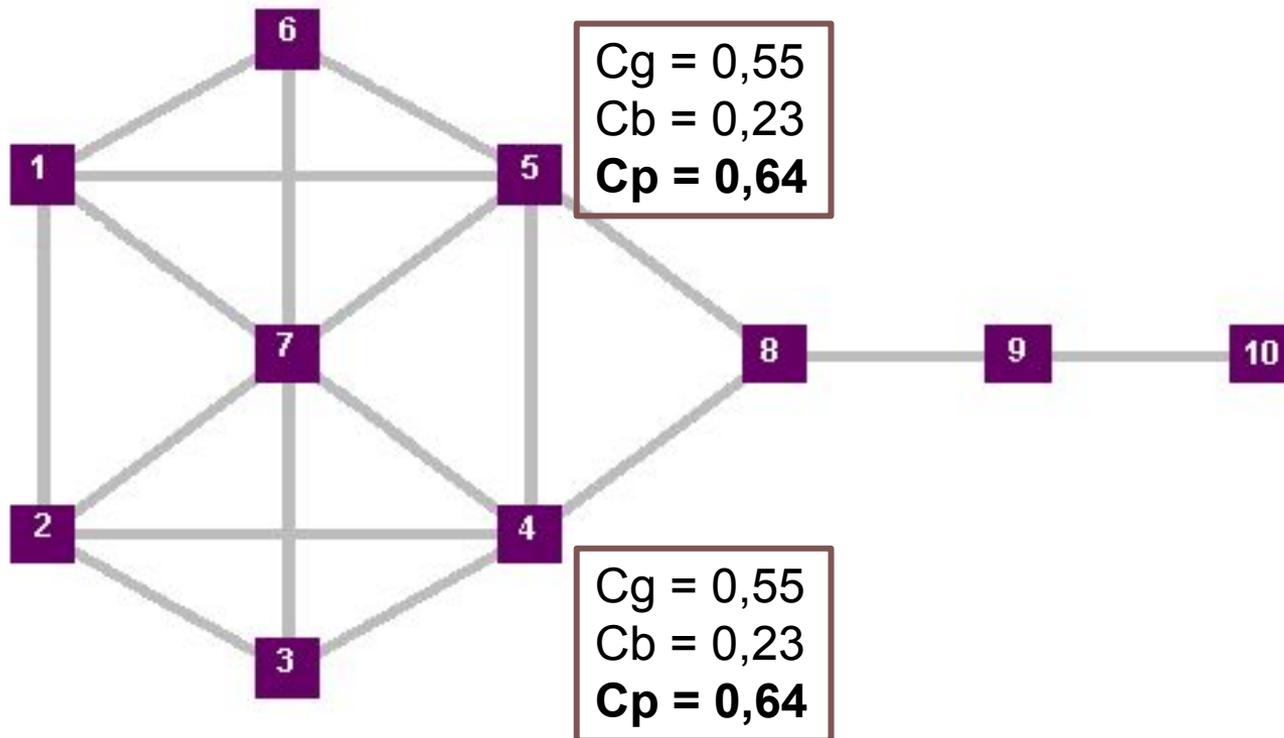
Ele interliga as duas redes formadas (uma centrada em 7 e outra em 9).

Sem ele, a informação que parte do nó 1 nunca chegará ao nó 10.



Centralidade até agora

Os nós 4 e 5 tem menos conexões que o nó 7, mas a posição na rede em que eles se encontram é privilegiada. **Eles conseguem ver tudo que acontece de forma mais rápida.**



Centralidade de Autovetor

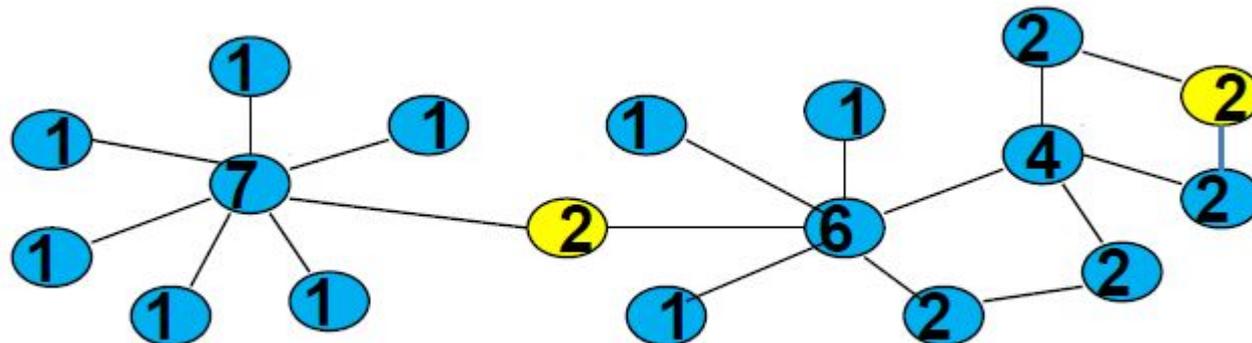
A centralidade de grau permite calcular o quão conectado é um nó.

No contexto de rede social isso pode ser um determinante de influência.



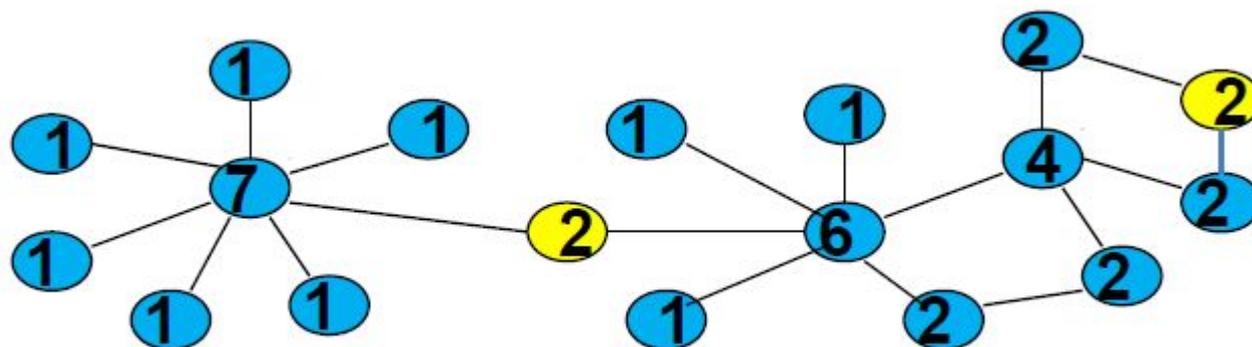
Centralidade de Autovetor

Porém, analisemos o seguinte caso:



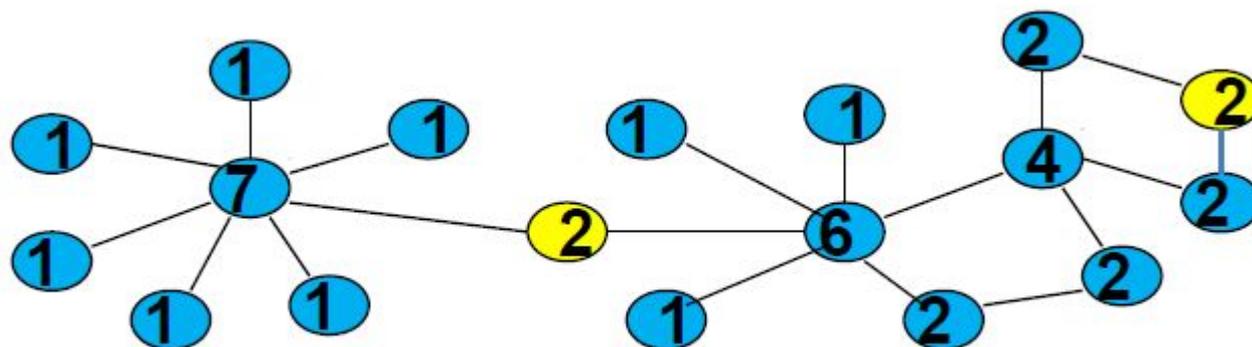
Centralidade de Autovetor

Apesar dos dois nós em amarelo apresentarem um grau de influência "2", eles não apresentam efetivamente a mesma influência.



Centralidade de Autovetor

Um deles está conectado a pessoas mais importantes (com grau 6 e 7), enquanto o outro com pessoas menos importante (outros de grau 2).



Centralidade de Autovetor

Uma forma de melhorar esse conceito foi definido com a centralidade de autovetor.

Essa centralidade calcula a importância do vértice baseado em sua influência na rede.



Centralidade de Autovetor

- Essa centralidade é proporcional à centralidade dos vizinhos do nó i :

$$C_i = \frac{1}{\lambda} \sum_j a_{ij} C_j$$

com $a_{ij} = 1$ se os nós i e j são adjacentes e λ é uma constante de importância.



Centralidade de Autovetor

-

Temos que:

$$\lambda \cdot C = \sum_i \sum_j a_{ij} C_j = A \cdot C$$

que se resume em encontrar a solução de **autovalores** e **autovetores** de A.



Autovalor e Autovetor

-

Dada uma matriz A e um vetor x e um escalar λ , dizemos que x é um autovetor de A e λ o correspondente autovalor se:

$$A \cdot x = \lambda x$$

Ou seja, se multiplicarmos a matriz A pelo vetor x , ele não mudará de direção, apenas sua intensidade.



Autovalor e Autovetor

Obviamente temos a solução trivial em que $x = 0$, porém é interessante encontrar soluções diferentes.

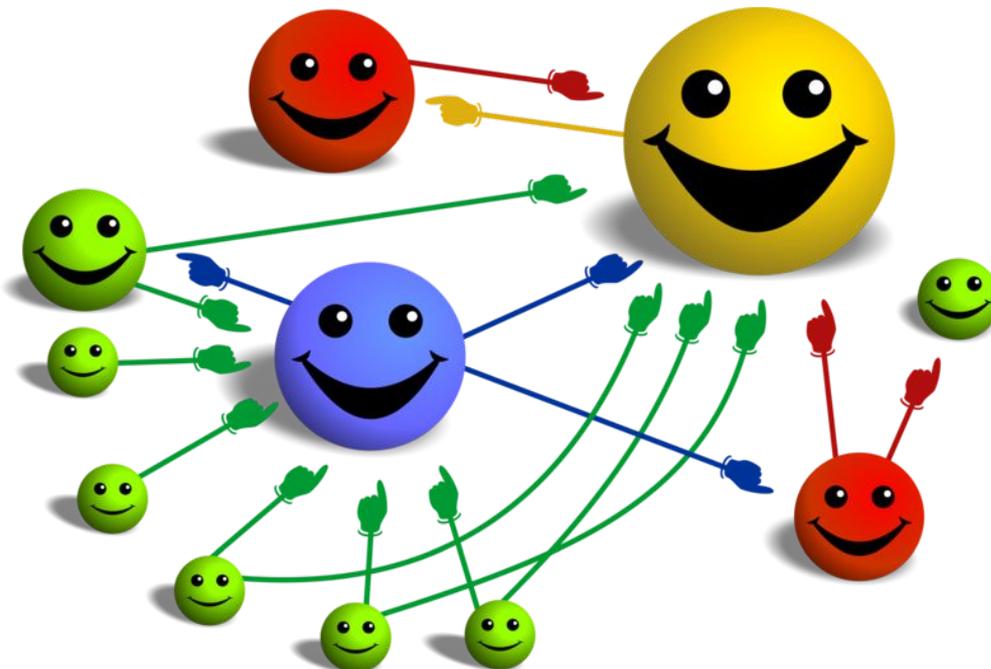
Dada certas condições temos um sistema com n equações e múltiplas soluções.

Como o vetor de centralidades deve ser positivo, e pelo teorema de Perron–Frobenius o maior **autovalor** corresponde ao **autovetor** não-negativo.



PageRank

Partindo da ideia da centralidade de autovetor, pesquisadores do Google pensaram em um algoritmo para determinar as páginas principais para serem mostradas em um resultado de busca.



PageRank

Esse algoritmo, denominado PageRank, determina a importância de um nó (página web) da seguinte forma:

- ❑ Uma página web A é importante se existem várias páginas apontando para ela e,
- ❑ Cada página B que aponta para ela incrementa o grau de centralidade proporcionalmente à quantas páginas apontam para B.



PageRank

A centralidade de PageRank (Cpr) de um nó é calculada da seguinte forma:

$$Cpr(u) = \sum_{v|(v,u) \in E} \frac{Cpr(v)}{grau_saida(v)}$$

Note que para calcular a centralidade de u precisamos conhecer a centralidade de todos os nós ligado à ele.



PageRank

- Vamos reescrever a equação anterior em forma matricial, para todos os nós:

$$C_{pr} = C_{pr} * N$$

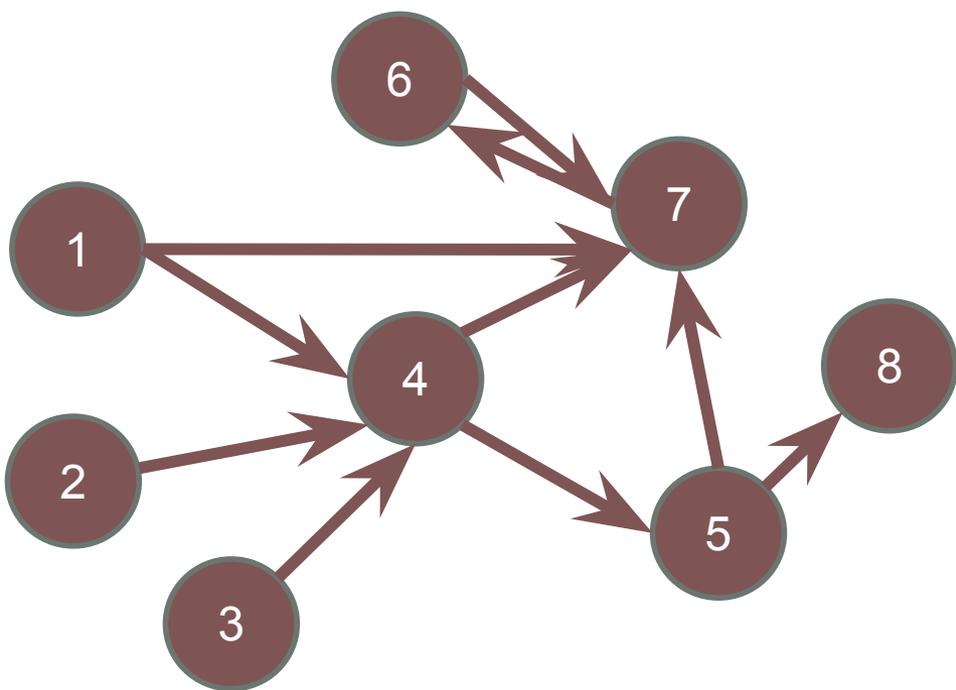
onde n_{ij} é igual a $1/k_i^{saída}$ se (i,j) formam uma aresta e 0 caso contrário; k_i representa o grau de saída do nó i.

Reparem que $1/k_i^{saída}$ representa a probabilidade de você sair do nó i para qualquer outro nó para o qual ele aponta.



PageRank

Para calcular o valor de PageRank, iniciamos o vetor Cpr com valores quaisquer.



			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
1	1	1	1	1	1	1	1



PageRank

Em seguida, aplicamos a equação de Cpr repetidamente até que os valores não se alterem (ou alterem muito pouco).

1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	2,5	0,5	1	2,5	0,5
---	---	---	-----	-----	---	-----	-----



PageRank

Em seguida, aplicamos a equação de Cpr repetidamente até que os valores não se alterem (ou alterem muito pouco).

0	0	0	2,5	0,5	1	2,5	0,5
---	---	---	-----	-----	---	-----	-----

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	1,25	2,5	2,5	0,25
---	---	---	---	------	-----	-----	------



PageRank

Em seguida, aplicamos a equação de Cpr repetidamente até que os valores não se alterem (ou alterem muito pouco).

0	0	0	0	1,25	2,5	2,5	0,25
---	---	---	---	------	-----	-----	------

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	0	2,5	3,125	0,625
---	---	---	---	---	-----	-------	-------



PageRank

Em seguida, aplicamos a equação de Cpr repetidamente até que os valores não se alterem (ou alterem muito pouco).

0	0	0	0	0	2,5	3,125	0,625
---	---	---	---	---	-----	-------	-------

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	0	3,125	2,5	0
---	---	---	---	---	-------	-----	---



PageRank

Em seguida, aplicamos a equação de Cpr repetidamente até que os valores não se alterem (ou alterem muito pouco).

0	0	0	0	0	3,125	2,5	0
---	---	---	---	---	-------	-----	---

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

=

0	0	0	0	0	2,5	3,125	0
---	---	---	---	---	-----	-------	---



PageRank

Epa!!

0	0	0	0	0	2,5	3,125	0
---	---	---	---	---	-----	-------	---

X

			0,5			0,5	
			1				
			1				
				0,5		0,5	
						0,5	0,5
						1	
					1		

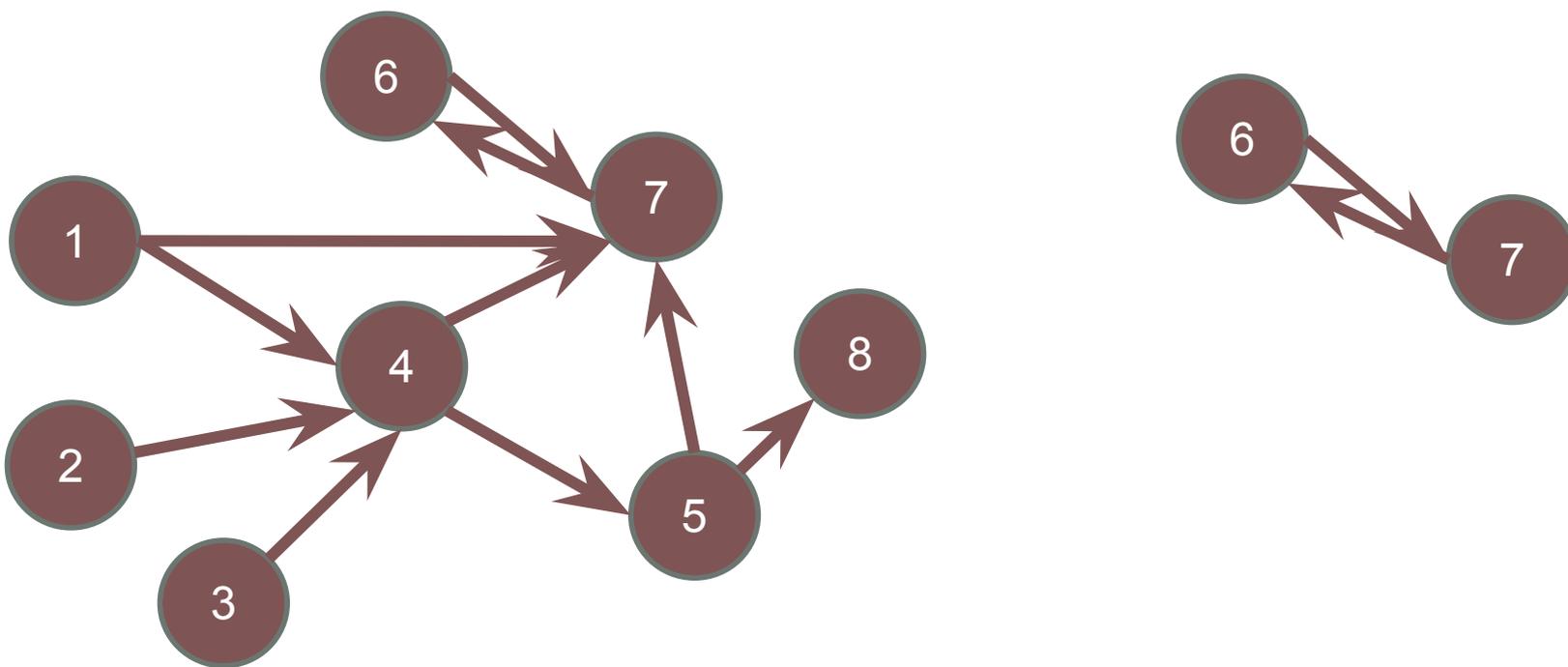
=

0	0	0	0	0	3,125	2,5	0
---	---	---	---	---	-------	-----	---



PageRank

Por causa do nó 6 e do nó 7, o cálculo entra em um laço infinito, pois do 6 só existe possibilidade de ir para o 7 e do 7 apenas para o 6.



PageRank

Para evitar esse ciclo, é adicionado uma pequena probabilidade para sair de um nó para qualquer outra página.

Com isso é construída a matriz **P** da seguinte forma:

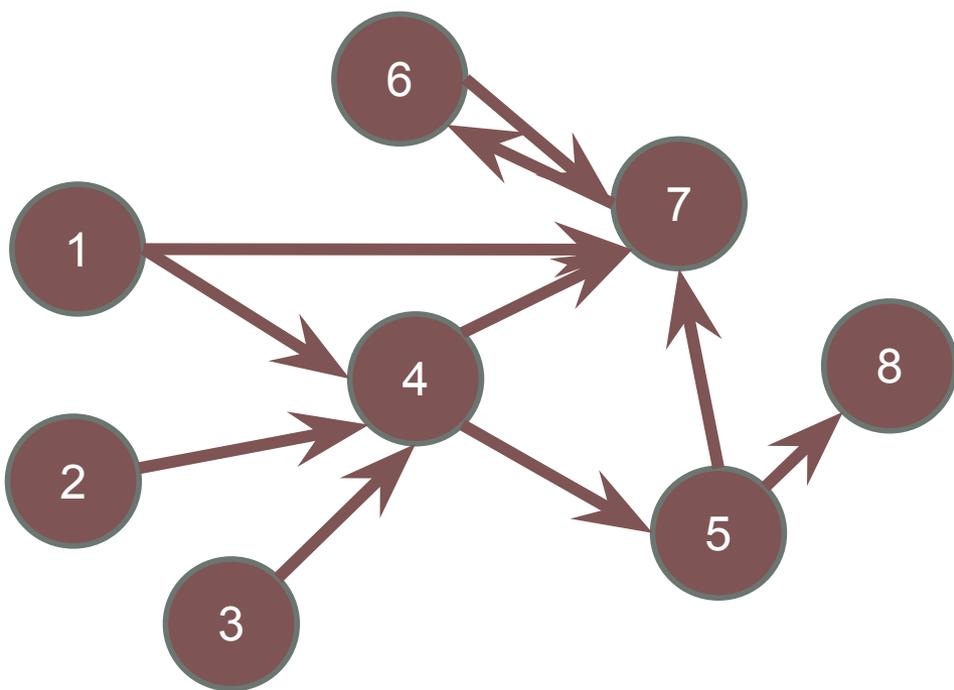
$$P = \alpha N + (1 - \alpha)E$$

A matriz **E** é uma matriz com valores iguais a $1/n$ em todos os elementos e n é a **ORDEM** da rede. O valor de α é definido como 0,85 e serve para amortizar o cálculo.



PageRank

Para essa rede a matriz \mathbf{P} será:



0.02	0.02	0.02	0.44	0.02	0.02	0.44	0.02
0.02	0.02	0.02	0.87	0.02	0.02	0.02	0.02
0.02	0.02	0.02	0.87	0.02	0.02	0.02	0.02
0.02	0.02	0.02	0.02	0.44	0.02	0.44	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.44	0.44
0.02	0.02	0.02	0.02	0.02	0.02	0.87	0.02
0.02	0.02	0.02	0.02	0.02	0.87	0.02	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02



PageRank

Basta utilizar o mesmo método iterativo, só que calculando:

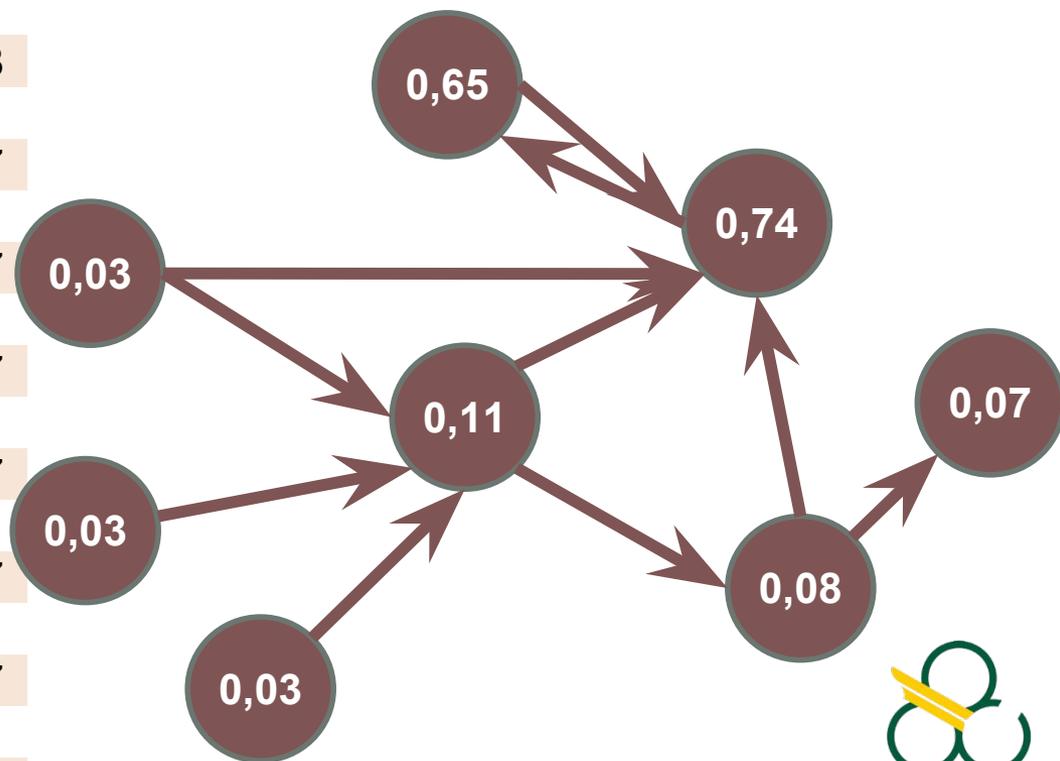
$$C_{pr} = \frac{C_{pr} * P}{\|C_{pr} * P\|}$$

esse método é conhecido como Método das Potências (Power Iteration) e, para o tipo de matriz que utilizamos (esparsas) é um método eficiente para calcular o autovetor com maior autovalor.



PageRank

Resultando na sequência de *Cpr's* à esquerda.



0.04 0.04 0.04 0.65 0.17 0.29 0.65 0.17

0.04 0.04 0.04 0.14 0.33 0.62 0.68 0.11

0.04 0.04 0.04 0.12 0.09 0.6 0.76 0.17

0.04 0.04 0.04 0.12 0.09 0.71 0.68 0.08

0.03 0.03 0.03 0.11 0.09 0.63 0.76 0.07

0.03 0.03 0.03 0.11 0.08 0.71 0.69 0.07

0.03 0.03 0.03 0.11 0.08 0.64 0.75 0.07

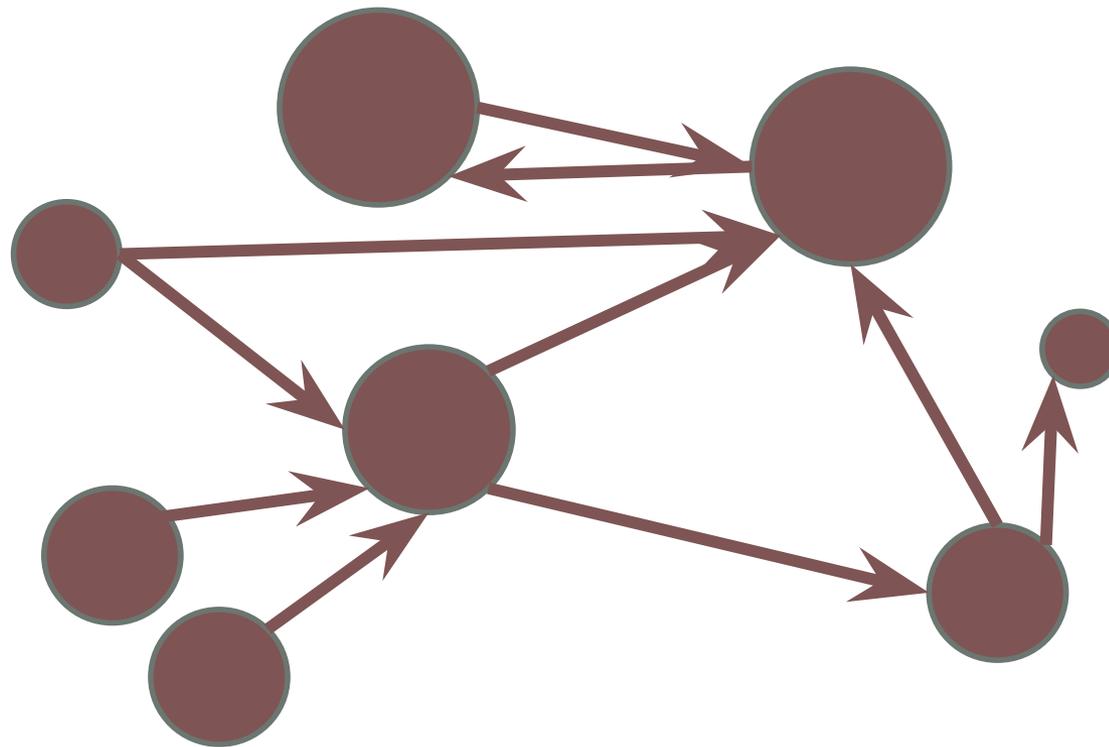
0.03 0.03 0.03 0.11 0.08 0.7 0.7 0.07

0.03 0.03 0.03 0.11 0.08 0.65 0.75 0.07

0.03 0.03 0.03 0.11 0.08 0.69 0.7 0.07

0.03 0.03 0.03 0.11 0.08 0.65 0.74 0.07

PageRank



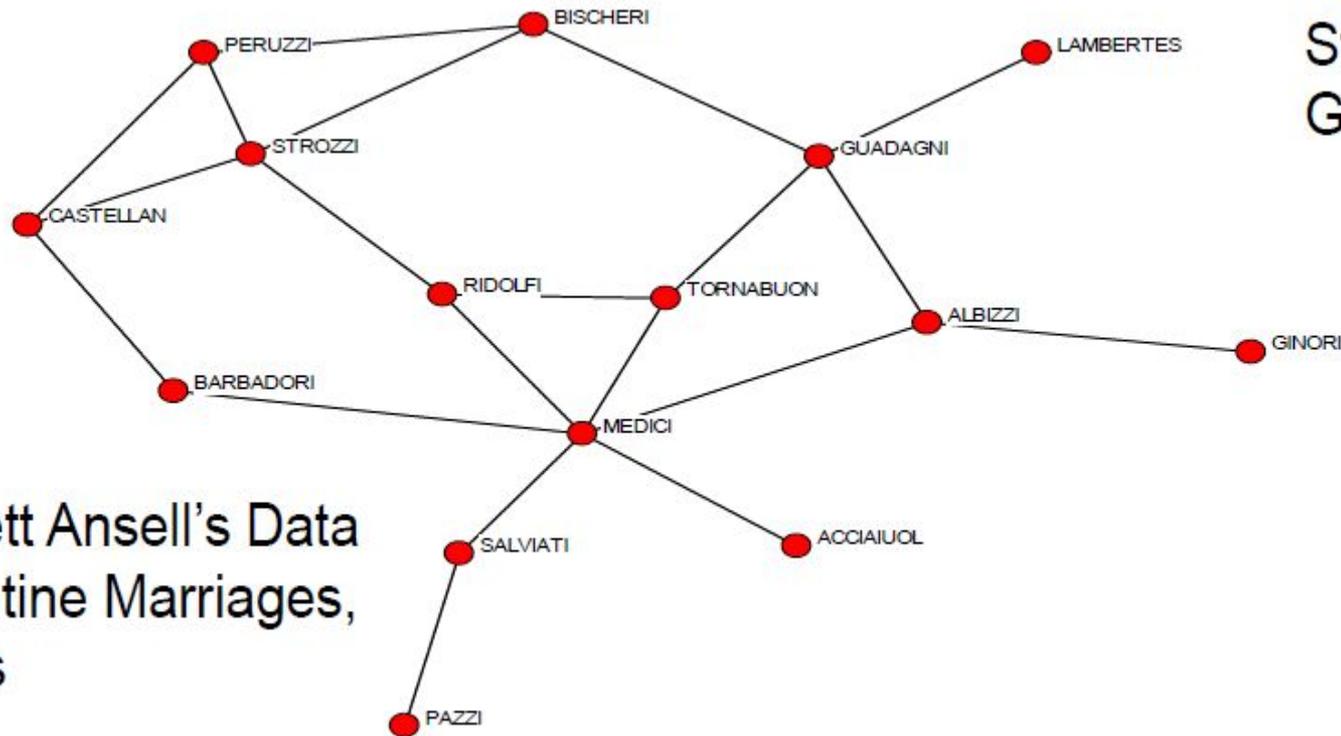


Universidade Federal do ABC

CENTRALIDADE NAS REDES REAIS

Florentina – Cent. Grau

● PUCCI



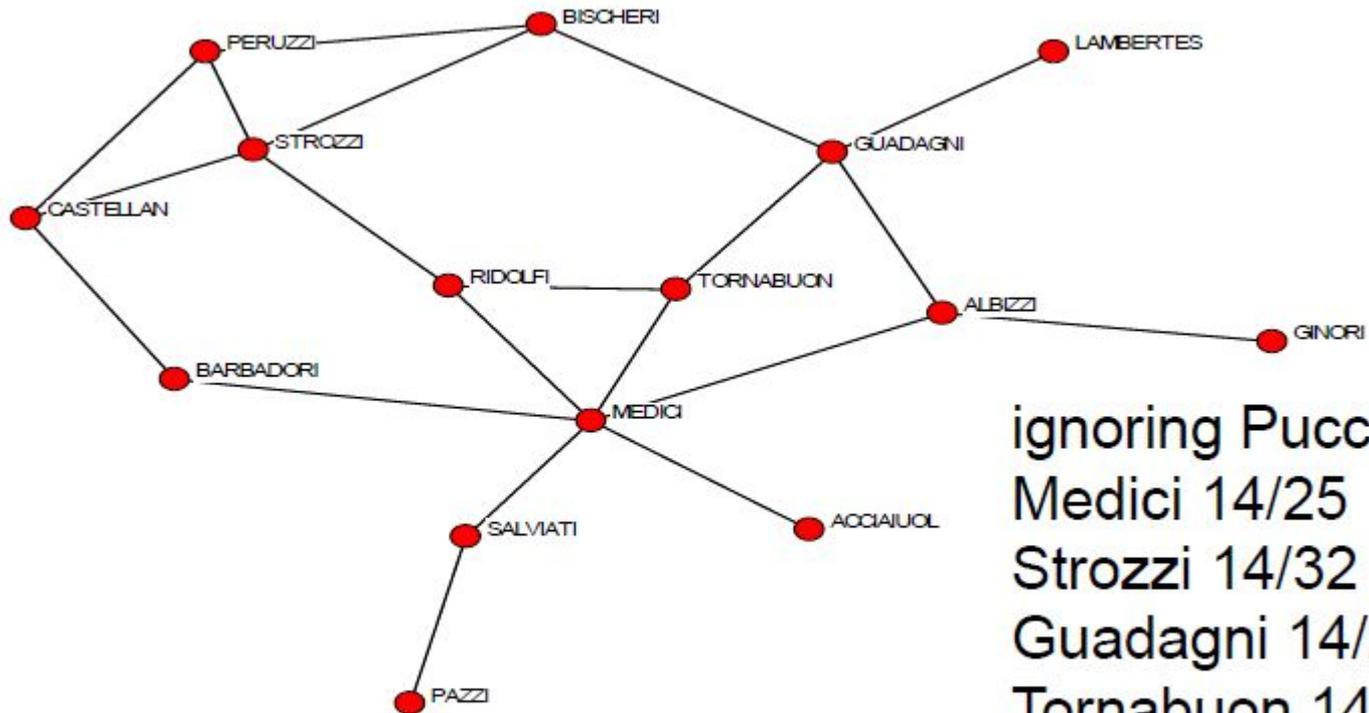
Medici = 6
Strozzi = 4
Guadagni = 4

Padgett Ansell's Data
Florentine Marriages,
1430's



Florentina – Cent. Proximidade

● PUCCI

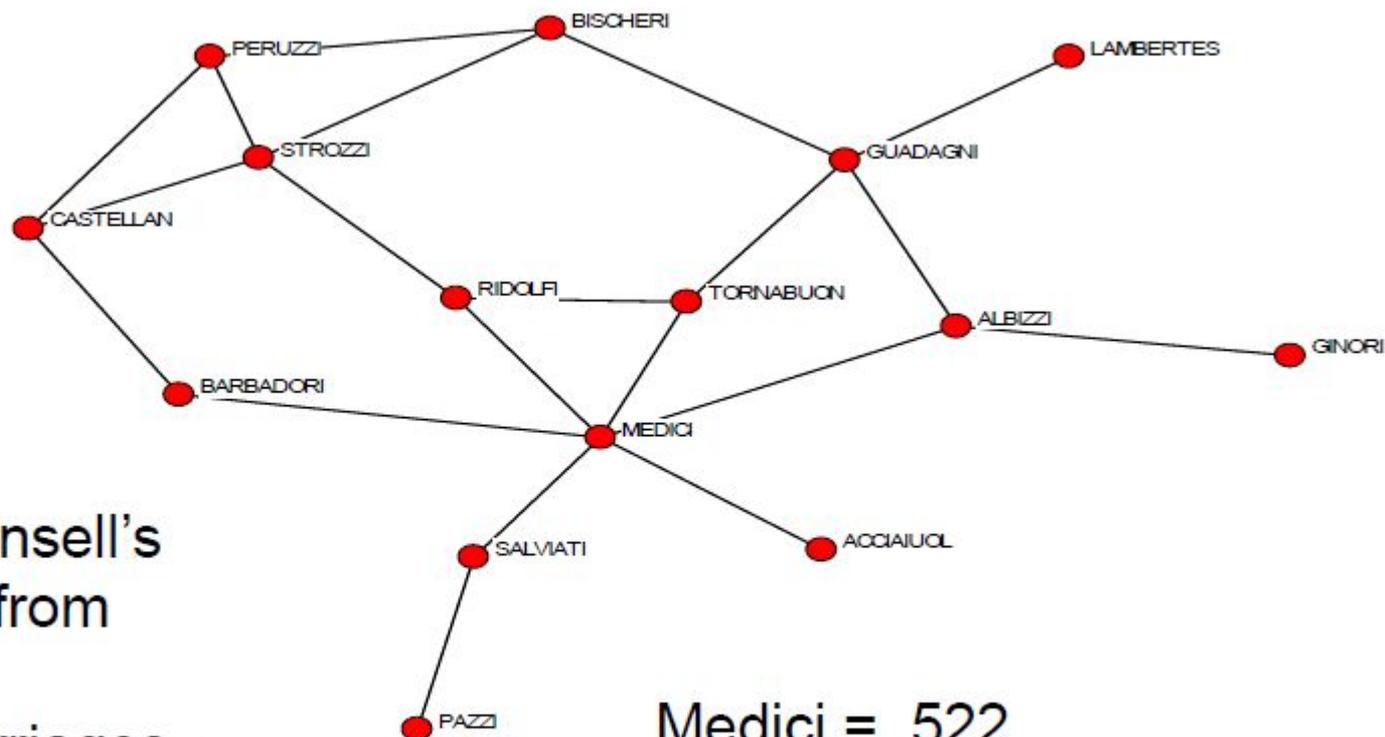


ignoring Pucci:
Medici 14/25
Strozzi 14/32
Guadagni 14/26
Tornabuon 14/29
Ridolfi 14/28



Florentina – Cent. Betweenness

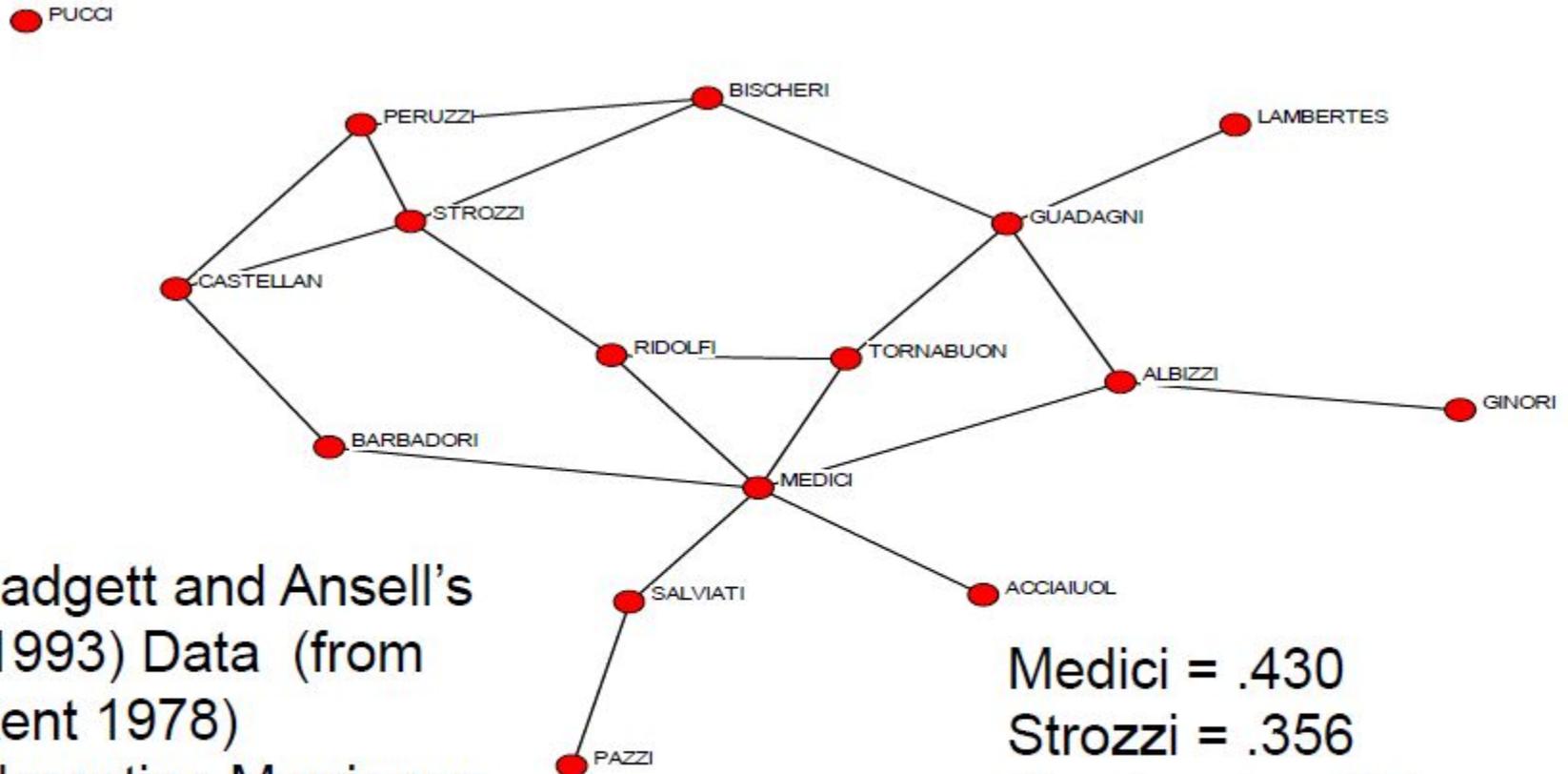
● PUCCI



Padgett and Ansell's
(1993) Data (from
Kent 1978)
Florentine Marriages,
1430's

Medici = .522
Strozzi = .103
Guadagni = .255

Florentina – Cent. Autovetor

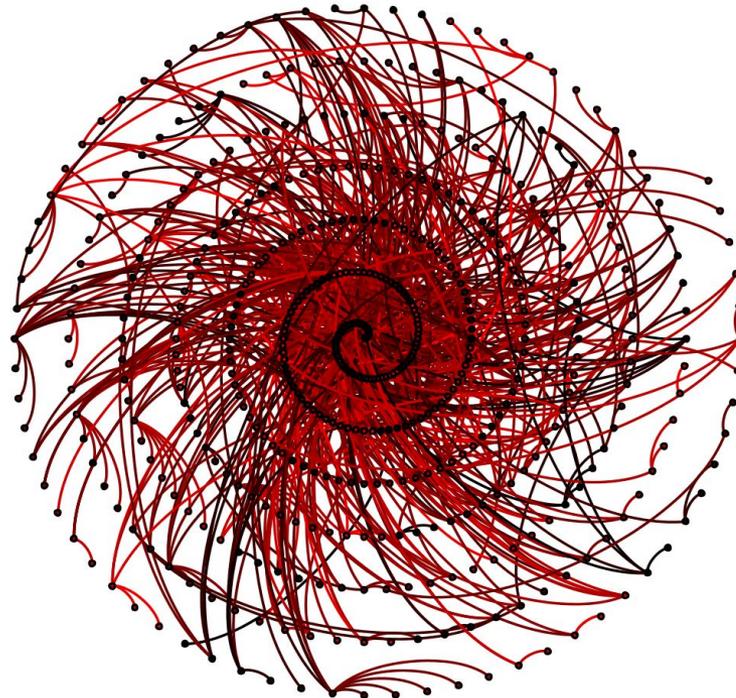


Padgett and Ansell's
(1993) Data (from
Kent 1978)
Florentine Marriages,
1430's

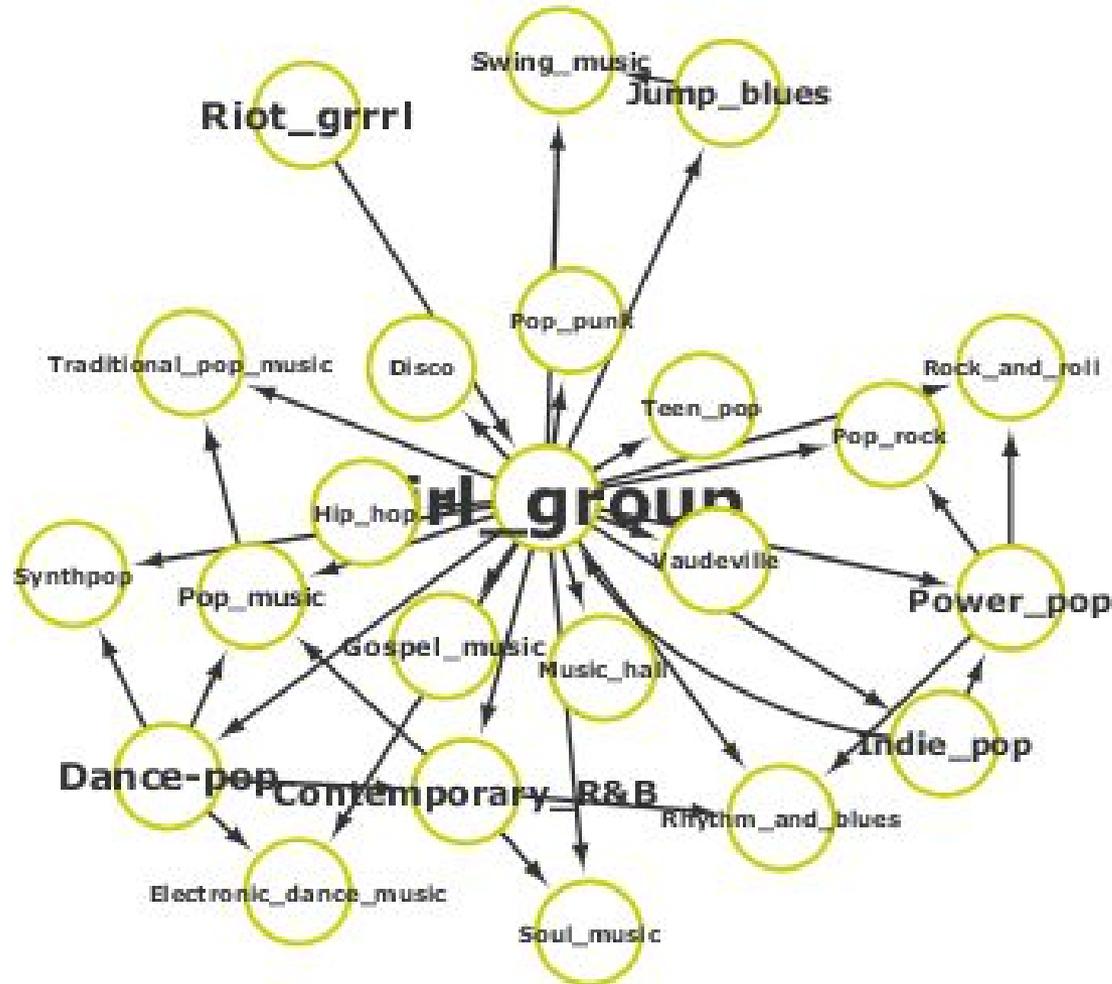
Medici = .430
Strozzi = .356
Guadagni = .289
Ridolfi = .341
Tornabuoni = .326

Influências Musicais

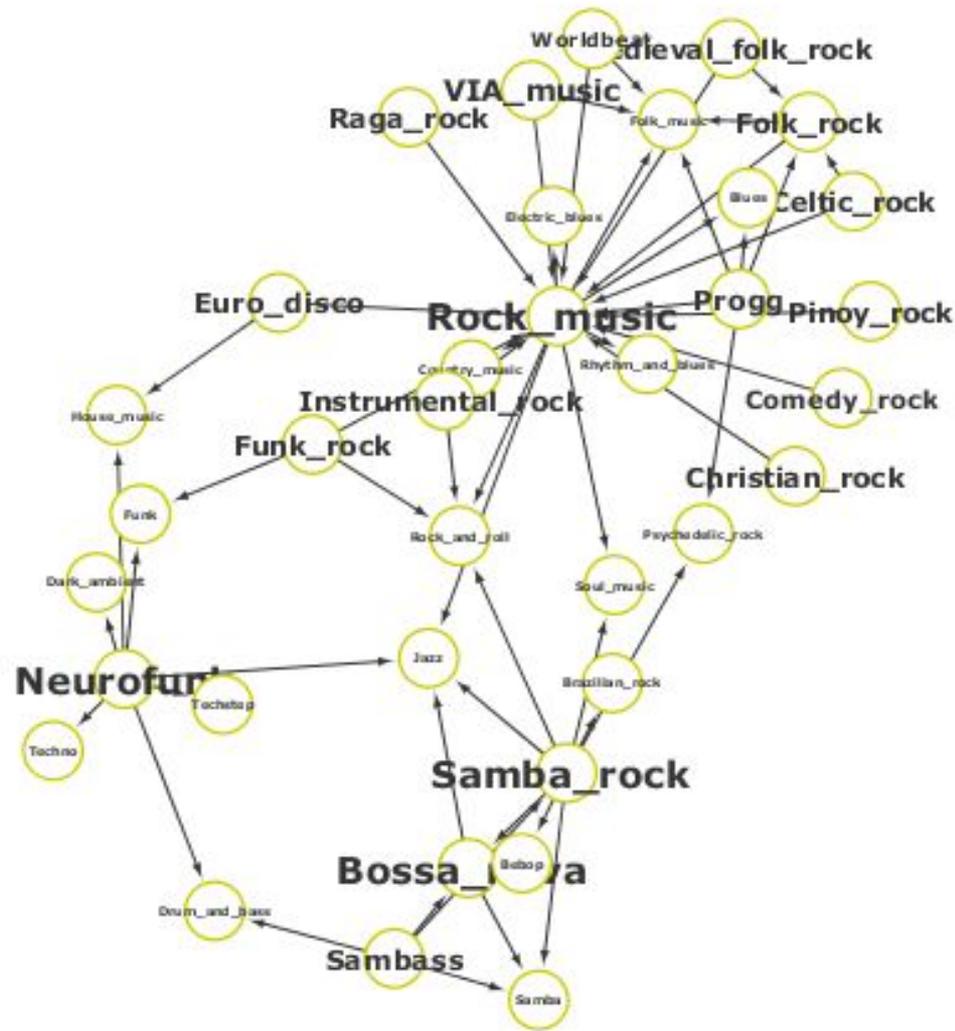
Em uma rede de estilos musicais em que a relação $A \rightarrow B$ indica que o estilo A recebeu influência de B podemos perceber alguns padrões interessantes nas métricas de centralidade.



Influências Musicais – Grau Saída



Influências Musicais – Proximidade





Universidade Federal do ABC

RESUMO

Centralidade

A centralidade mede a importância dos nós.

Grau: quão bem conectado o nó está.

Proximidade: facilidade em alcançar outros nós.

Betweenness: identifica intermediários.

Autovetores e Pagerank: influência e prestígio.

