

Padrões de Segregação

Comunidades

Comunidades

Como vimos anteriormente, em redes sociais é possível verificar grupos fechados de pessoas (pessoas que se conhecem mutuamente) e, em alguns nós, pessoas que interligam esses grupos.

Isoladamente cada grupo não pode contribuir com muita informação nova, pois a informação dentro daquele grupo já é de conhecimento de todos que estão ali.

Porém, ao interagirem com outros grupos, novos conhecimentos surgem e levam a uma interação complexa.



Particionamento de Redes

Particionamento é o procedimento de divisão da rede em várias subredes de tal forma a identificar as comunidades existentes.

Exemplos de utilidades para o particionamento:

- ❑ Redes sociais para descobrir comunidades com interesses semelhantes
- ❑ Redes biológicas para descobrir funcionalidades de genes para combater doenças
- ❑ **Estudar como funciona o processo de comunicação e difusão de informação na rede.**



Dinâmica da Comunicação (prévia)

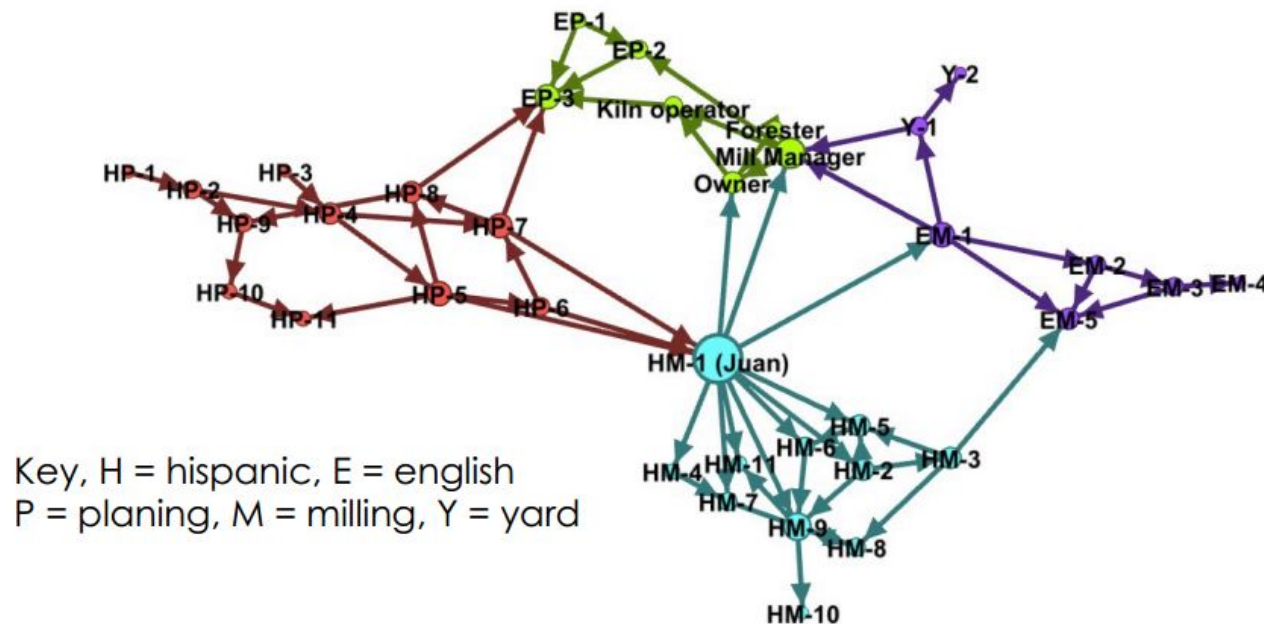
É importante conhecer como ocorre o processo de comunicação em uma rede pois podemos:

- ❑ Melhorar a forma como a comunicação é feita
- ❑ Controlar a informação e alterá-la em nosso favor
- ❑ Proteger ou preservar o processo de comunicação



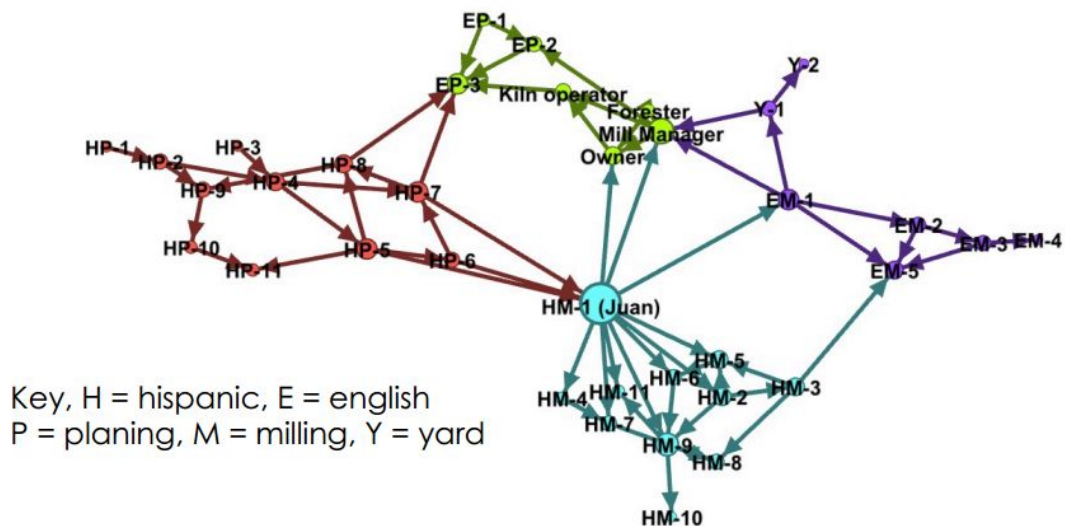
Dinâmica da Comunicação (prévia)

Uma serralheria americana estava com problemas para convencer os funcionários a adotar novas práticas de trabalho.



Dinâmica da Comunicação (prévia)

Em particular, os hispânicos rejeitaram em massa o novo método



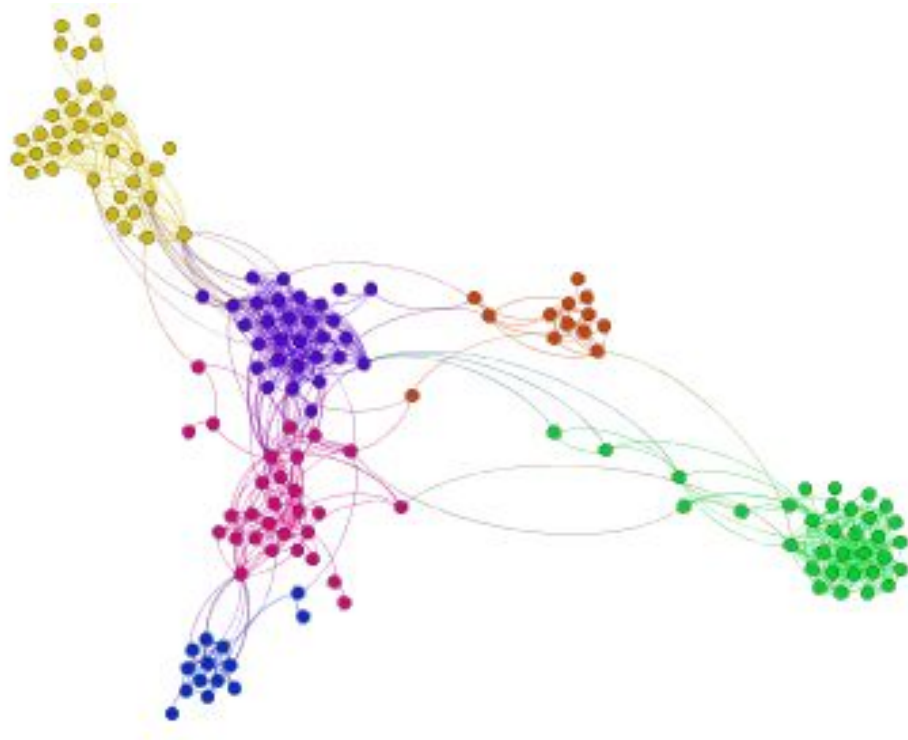
Sawmill network: source Exploratory Social Network Analysis with Pajek

Ao analisar o nó mais central seguindo critérios de betweenness e identificar as partições da rede, percebeu-se que o problema se resolveria apenas convencendo o Juan.



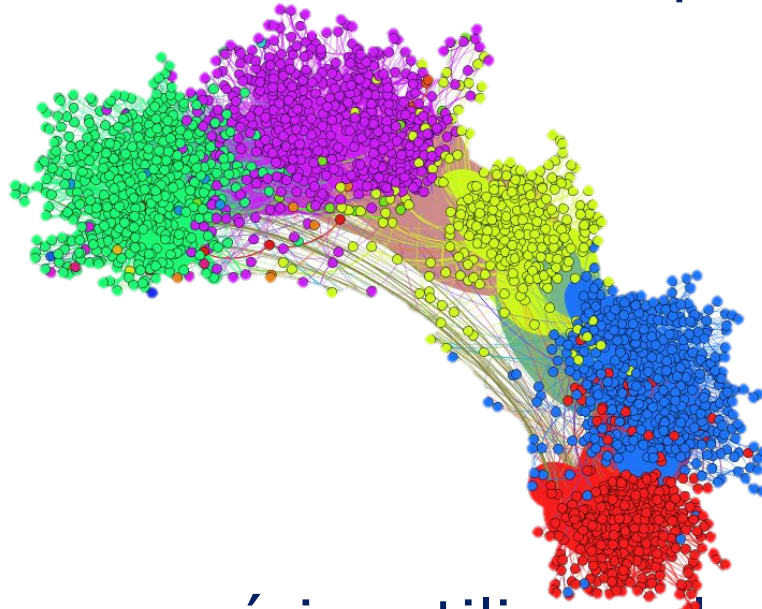
Particionamento Visual

Em algumas redes reais as partições de nós são facilmente percebidas visualmente.



Particionamento Visual

Mas muitas vezes não é possível ver onde começa e onde termina uma determinada partição.



Por isso é necessário utilizar algum método para determinar tais partições e indicar qual partição é melhor.



Métodos de Particionamento

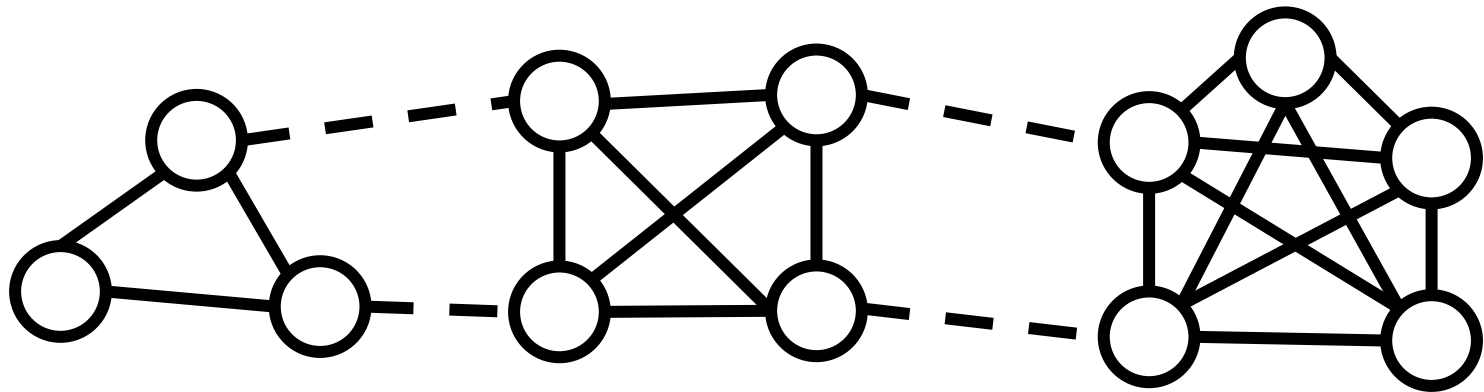
Antes de criar métodos para particionar a rede precisamos definir o que define uma partição:

- ❑ **Ligações mútuas:** todos conhecem todos em um grupo
- ❑ **Ligações frequentes:** todos conhecem a maioria do grupo
- ❑ **Proximidade:** todos estão separados por no máximo n nós



Particionamento por Cliques

Um particionamento ideal é aquele em que todas as partições formem uma **REDE COMPLETA**, ou seja, todos conhecem todos:

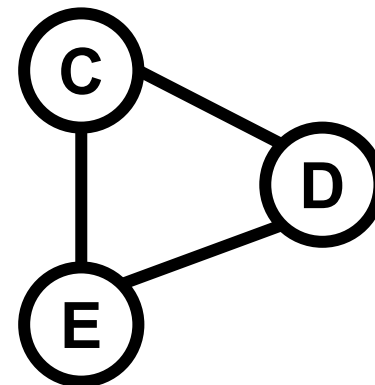
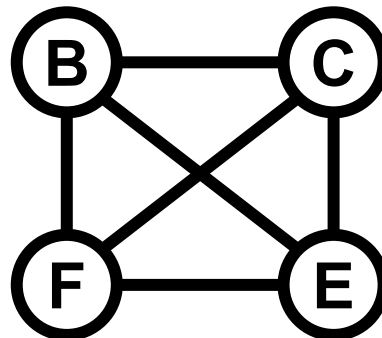
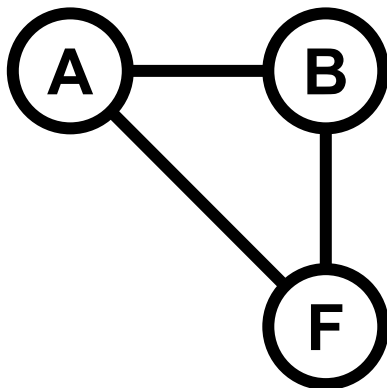
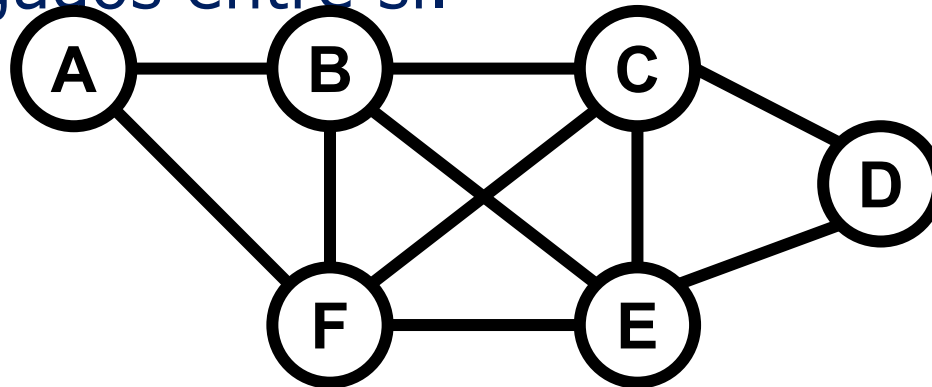


O objetivo é cortar um número mínimo de arestas de tal forma que obtemos N **COMPONENTES CONEXOS** que formem **REDES COMPLETA**.



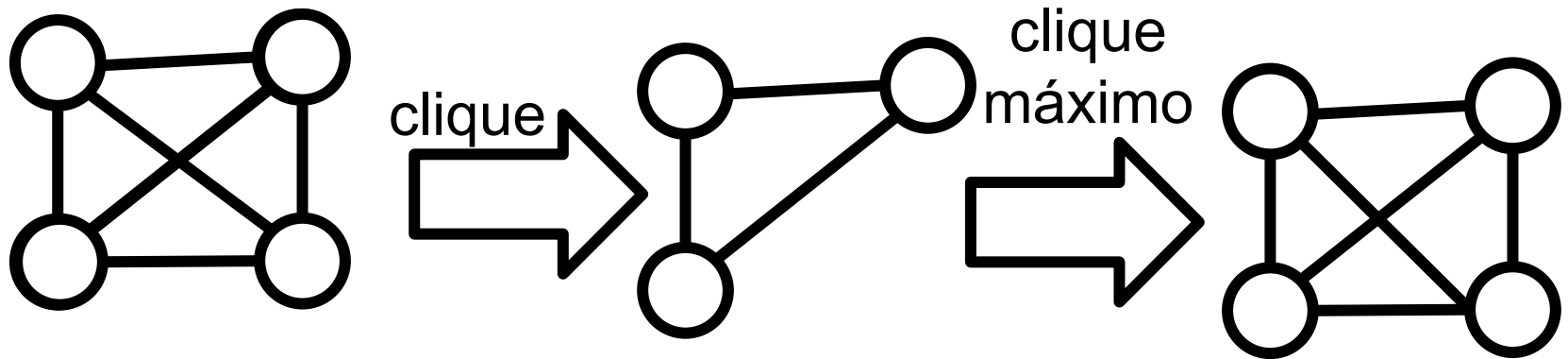
Clique de um Grafo

O **CLIQUE** de um grafo é um grupo de nós que formam uma rede completa, ou seja, com todos os nós interligados entre si.



Clique de um Grafo

O **CLIQUE MÁXIMO** de uma rede é um clique que não é um subconjunto de nenhum outro clique.



Em outras palavras é um clique em que não existe nenhum outro nó não pertencente a ele que, se adicionado, forme também um clique



Clique de um Grafo

Achar os cliques de um grafo pode ser pensado como enumerar todos os subgrafos e todas as possibilidades de arestas nesse subgrafo.

Se temos n nós, temos n subgrafos de tamanho 1, $n(n-1)/2$ subgrafos de tamanho 2, $n(n-1)(n-2)/6$ subgrafos de tamanho 3,...

$$n(n-1)\dots(n-k+1) / (1.2\dots k)$$



Clique de um Grafo

•

$$\frac{n(n-1) \dots (n-k+1)}{1.2.3 \dots k} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k}$$

Devemos verificar para $k=1..n$:

$$\sum_{1 \leq k \leq n} \binom{n}{k} = 2^n - 1$$



Clique de um Grafo

$n = 1$, 1 possibilidade diferente

$n = 2$, 3 possibilidades

$n = 10$, 1.023 possibilidades

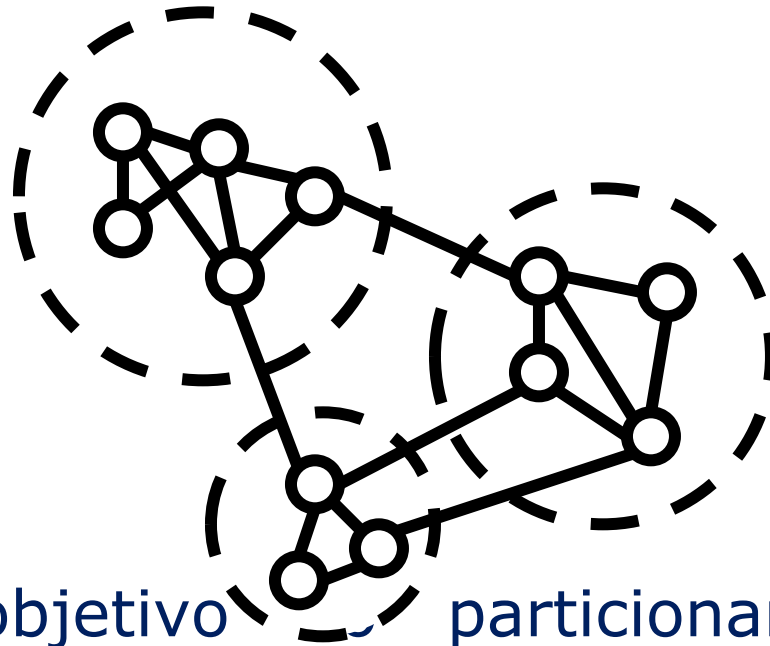
$n = 100$,

1.267.650.600.228.229.401.496.703.205.375
possibilidades diferentes



Particionamento por Cliques

Na prática é impossível determinar grupos tão bem definidos que formam cliques

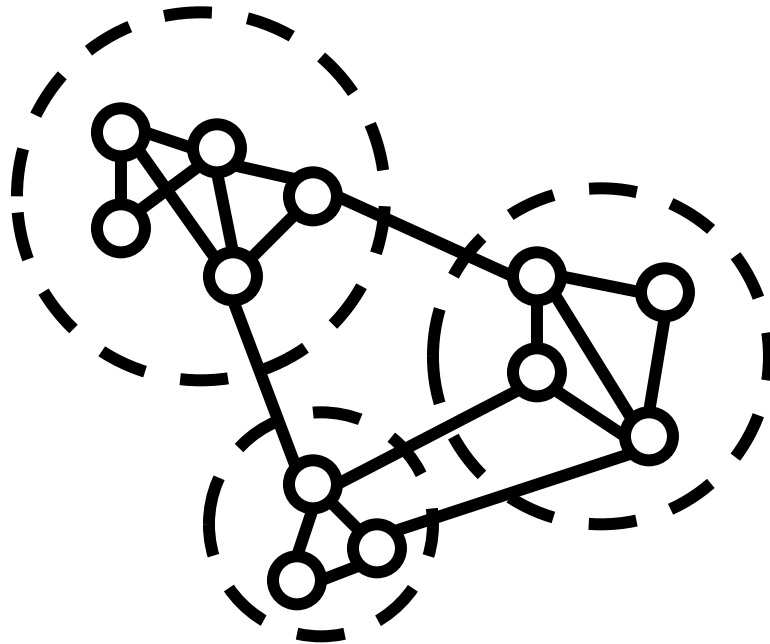


Logo, o objetivo de encontrar os conjuntos de nós que mais se aproximam de cliques torna-se



Particionamento por Cliques

Por exemplo, grupos de nós que tenham ligações com pelo menos k nós do mesmo grupo.



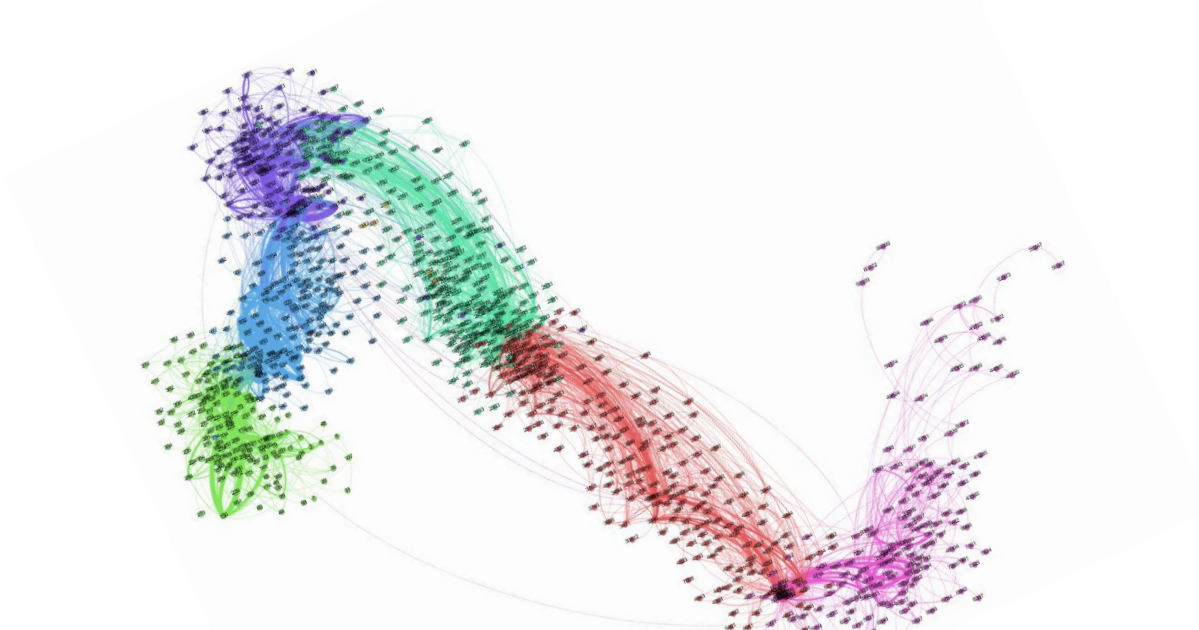
Procedimento conhecido como **K-CORE**.

Qual o valor de k para as partições da figura acima?



E se os cliques não forem suficientes?

Em algumas redes as estruturas formadas por cliques não indicam a sua funcionalidade.



E se as comunidades não são formadas por cliques, mas por alguma particularidade da estrutura?



Métodos Baseados em Centralidade

Vimos anteriormente a centralidade betweenness de um nó, que mede quantos caminhos mais curtos passam por aquele nó.

Essa medida indica nós que servem como pontes entre dois conjuntos distintos de nós.

Se modificarmos essa medida para calcularmos a centralidade das **ARESTAS**, podemos encontrar as arestas que dividem dois grupos de nós.



Métodos Baseados em Centralidade

O método de Girvan-Newman é baseado na centralidade betweenness **DE ARESTA**.

Esse método é chamado **DIVISIVO** pois, a cada passo divide a rede em duas formando uma hierarquia de partições.



Métodos Girvan-Newman

Centralidades:

$$(A,B) = 0,167$$

$$(A,C) = 0,167$$

$$(B,C) = 0,133$$

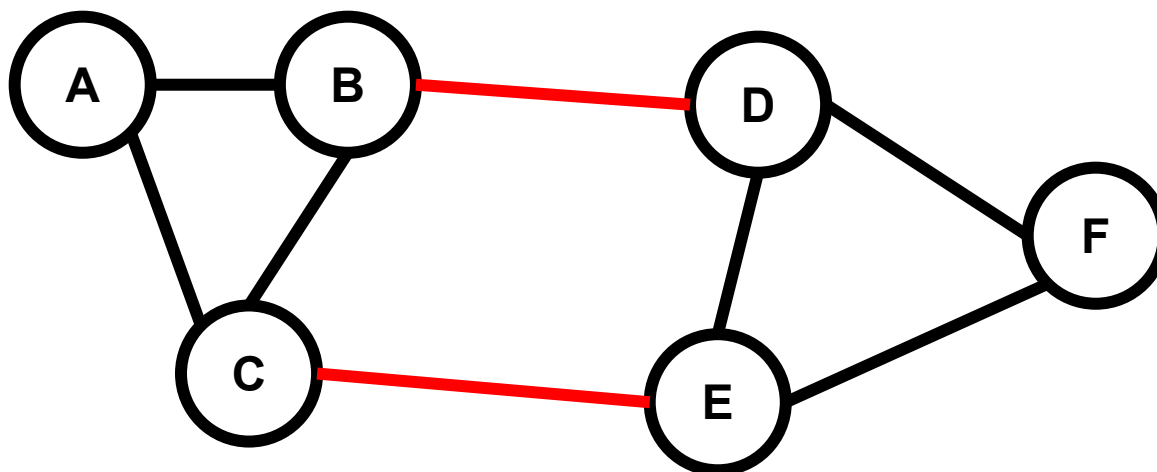
$$(B,D) = \mathbf{0,300}$$

$$(C,E) = \mathbf{0,300}$$

$$(D,E) = 0,133$$

$$(D,F) = 0,167$$

$$(E,F) = 0,167$$



Métodos Girvan-Newman

Centralidades:

$$(A,B) = \mathbf{0,167}$$

$$(A,C) = \mathbf{0,167}$$

$$(B,C) = 0,133$$

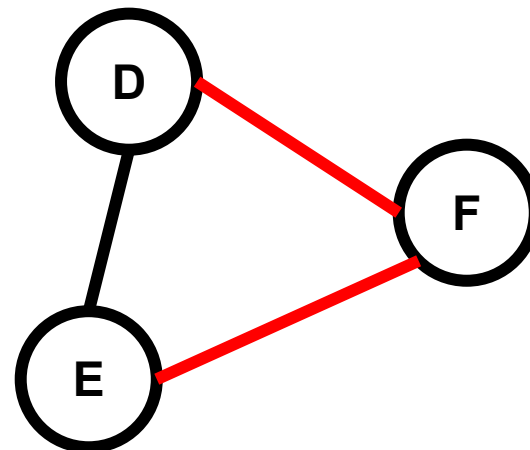
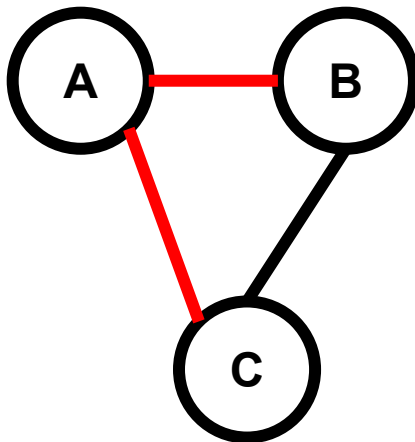
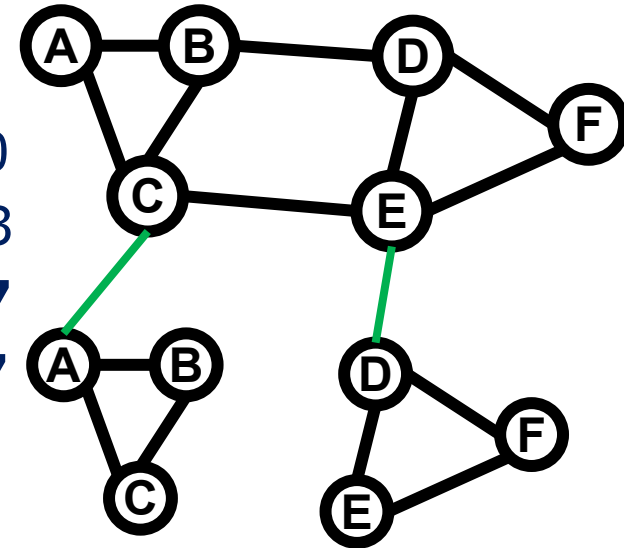
$$(B,D) = 0,000$$

$$(C,E) = 0,000$$

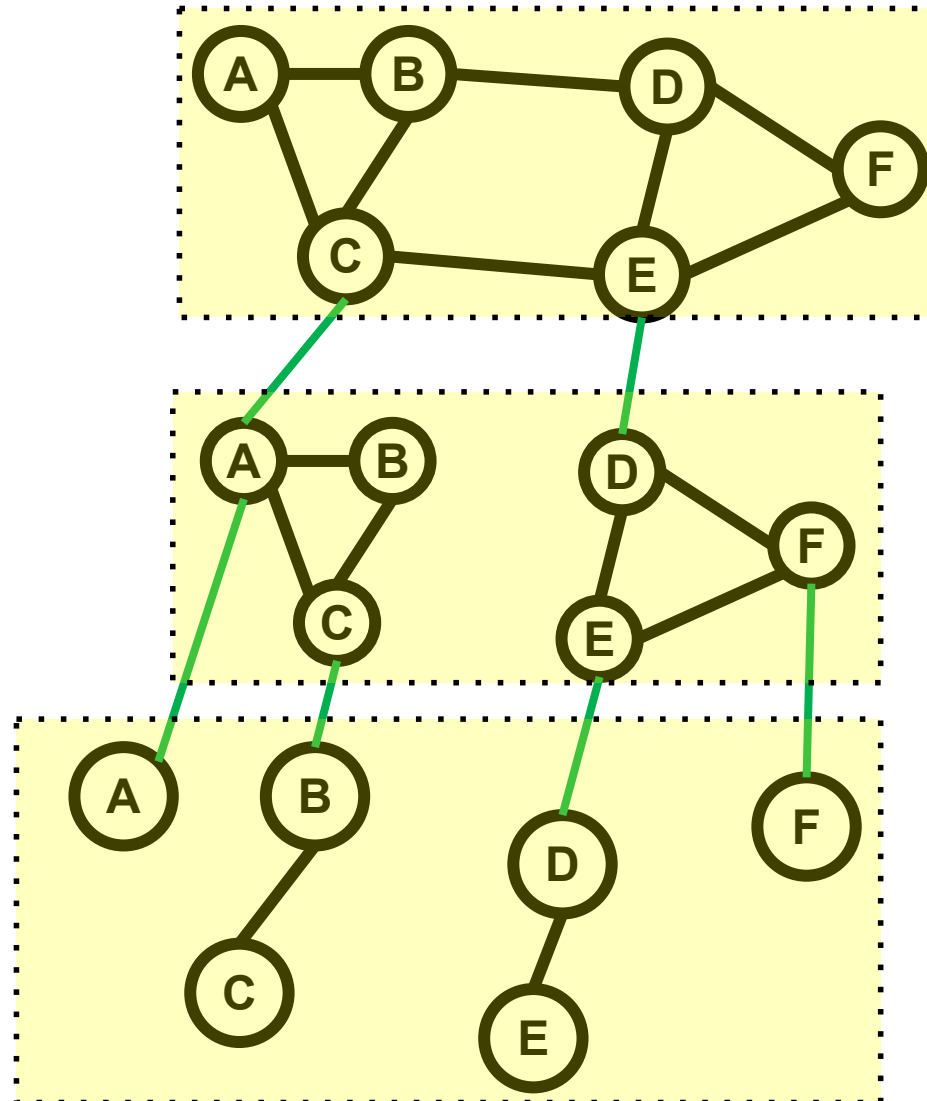
$$(D,E) = 0,133$$

$$(D,F) = \mathbf{0,167}$$

$$(E,F) = \mathbf{0,167}$$



Métodos Girvan-Newman



Métodos Girvan-Newman no Facebook



- <http://www.touchgraph.net/>
- <http://www1.maths.leeds.ac.uk/statistics/workshop/lasr2006/proceedings/pinney-talk.pdf>



Particionamento Espectral

Algumas propriedades interessantes de uma rede aparecem ao calcularmos os autovalores e autovetores de uma matriz de adjacência modificada.

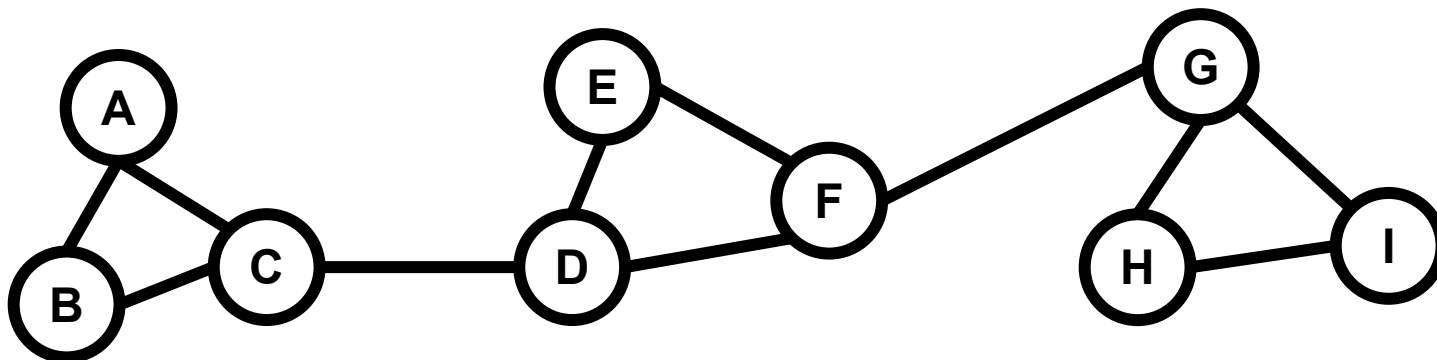
Vamos utilizar a matriz Laplaciana que é formada pela matriz de graus dos nós do grafo subtraída da matriz de adjacência:

$$L = G - A$$

Agora vamos fazer alguns experimentos em uma rede simples.



Particionamento Espectral



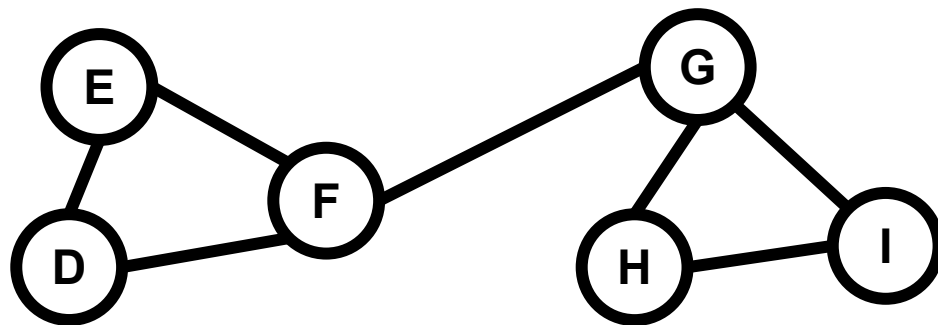
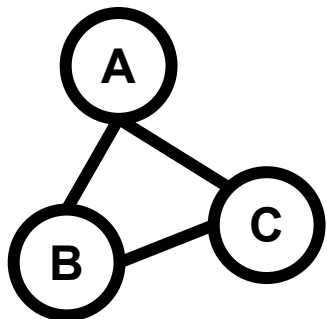
	A	B	C	D	E	F	G	H	I
A	2	-1	-1	0	0	0	0	0	0
B	-1	2	-1	0	0	0	0	0	0
C	-1	-1	3	-1	0	0	0	0	0
D	0	0	-1	3	-1	-1	0	0	0
E	0	0	0	-1	2	-1	0	0	0
F	0	0	0	-1	-1	3	-1	0	0
G	0	0	0	0	0	-1	3	-1	-1
H	0	0	0	0	0	0	-1	2	-1
I	0	0	0	0	0	0	-1	-1	2

← Matriz Laplaciana

Autovalores

0,0 0,2 0,7 3,0 3,0 3,0 3,0 4,3 4,8

Particionamento Espectral



	A	B	C	D	E	F	G	H	I
A	2	-1	-1	0	0	0	0	0	0
B	-1	2	-1	0	0	0	0	0	0
C	-1	-1	2	0	0	0	0	0	0
D	0	0	0	2	-1	-1	0	0	0
E	0	0	0	-1	2	-1	0	0	0
F	0	0	0	-1	-1	3	-1	0	0
G	0	0	0	0	0	-1	3	-1	-1
H	0	0	0	0	0	0	-1	2	-1
I	0	0	0	0	0	0	-1	-1	2

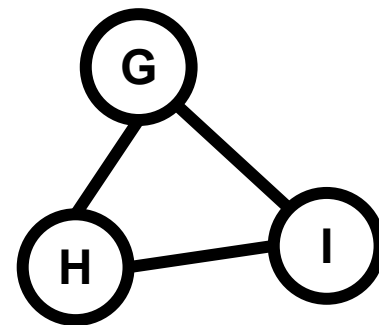
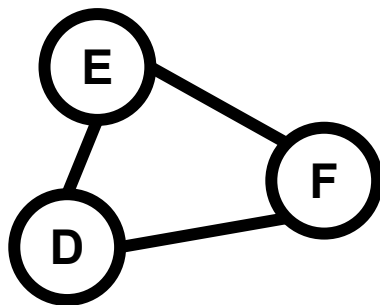
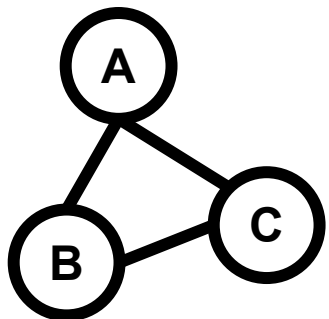
← Matriz Laplaciana

Autovalores

0,0 | 0,0 | 0,4 | 3,0 | 3,0 | 3,0 | 3,0 | 3,0 | 4,6



Particionamento Espectral



	A	B	C	D	E	F	G	H	I
A	2	-1	-1	0	0	0	0	0	0
B	-1	2	-1	0	0	0	0	0	0
C	-1	-1	2	0	0	0	0	0	0
D	0	0	0	2	-1	-1	0	0	0
E	0	0	0	-1	2	-1	0	0	0
F	0	0	0	-1	-1	2	0	0	0
G	0	0	0	0	0	0	2	-1	-1
H	0	0	0	0	0	0	-1	2	-1
I	0	0	0	0	0	0	-1	-1	2

← Matriz Laplaciana

Autovalores

0,0 | 0,0 | 0,0 | 3,0 | 3,0 | 3,0 | 3,0 | 3,0 | 3,0



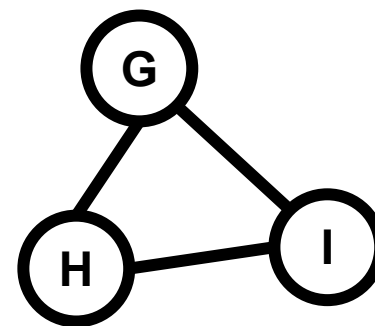
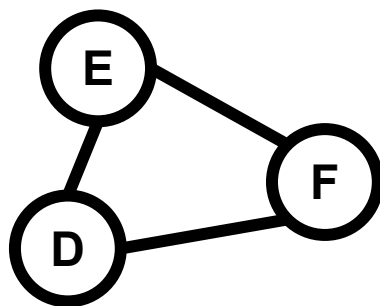
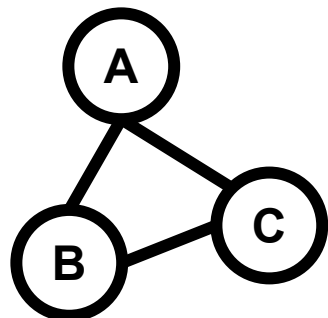
Particionamento Espectral

Reparem que o número de **AUTOVALORES** iguais a **ZERO** representa o número de **COMPONENTES CONEXOS** da rede.

E como se comportam os autovetores correspondentes?



Particionamento Espectral



Autovalores

0,0	0,0	0,0	3,0	3,0	3,0	3,0	3,0	3,0
-----	-----	-----	-----	-----	-----	-----	-----	-----

Autovetores

0,6	0,6	0,6	0,0	0,0	0,0	0,0	0,0	0,0
0,0	0,0	0,0	0,6	0,6	0,6	0,0	0,0	0,0
0,0	0,0	0,0	0,0	0,0	0,0	0,6	0,6	0,6



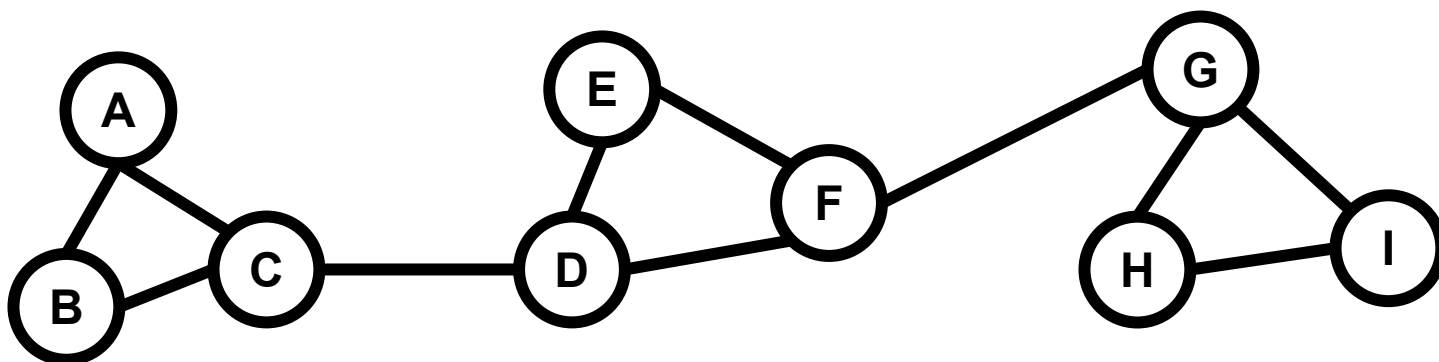
Particionamento Espectral

Dá para perceber claramente que os autovetores identificam os nós de cada componente conexo!

E no primeiro caso, em que temos uma rede totalmente conectada, como se comportam o segundo e terceiro autovetor?



Particionamento Espectral



Autovalores

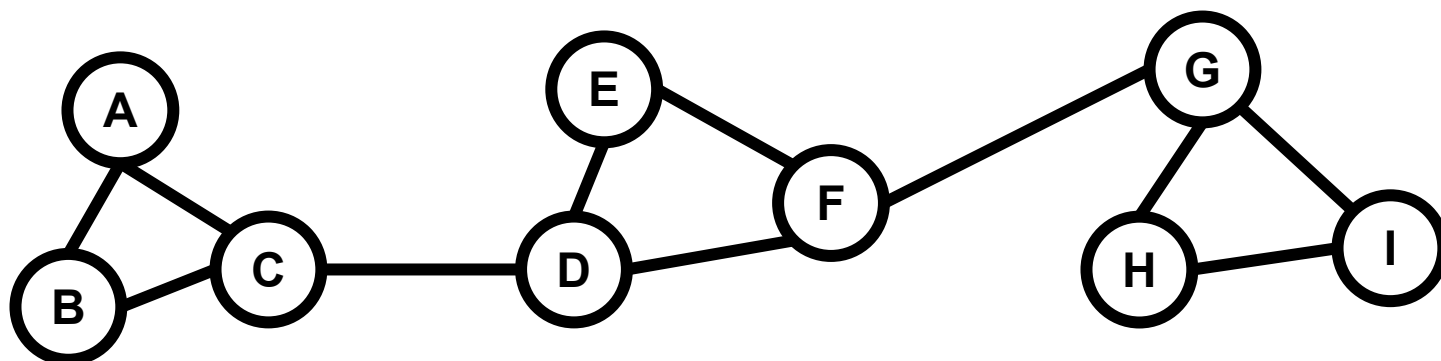
0,0	0,2	0,7	3,0	3,0	3,0	3,0	4,3	4,8
-----	-----	-----	-----	-----	-----	-----	-----	-----

Autovetores

0,4	0,4	0,3	0,1	0,0	-0,1	-0,3	-0,4	-0,4
0,3	0,3	0,1	-0,4	-0,6	-0,4	0,1	0,3	0,3



Particionamento Espectral



Autovalores

0,0	0,2	0,7	3,0	3,0	3,0	3,0	4,3	4,8
-----	-----	-----	-----	-----	-----	-----	-----	-----

Autovetores

0,4	0,4	0,3	0,1	0,0	-0,1	-0,3	-0,4	-0,4
0,3	0,3	0,1	-0,4	-0,6	-0,4	0,1	0,3	0,3



Particionamento Espectral

O segundo autovetor mostra claramente a separação entre dois grupos distintos.

No terceiro autovetor podemos perceber a formação de 3 grupos, mas isso porque os nós estão ordenados corretamente!



Particionamento Espectral

O segundo autovetor, conhecido como vetor Fiedler, tem as seguintes propriedades:

- ❑ Tem autovalor correspondente maior que **ZERO** se e somente se a rede for conectada e;
- ❑ Particiona a rede em duas redes distintas repartindo em torno de um nó que divide a rede de forma simétrica.



Particionamento Espectral

O autovalor correspondente ao vetor Fiedler é conhecido como conectividade algébrica e indica o quão conectada é a rede.

Esse valor varia entre $4/nD \leq \sigma \leq 1$,

D é o diâmetro da rede e n é o número de nós.



Particionamento Espectral

As partições são definidas como:

- Nós em que os valores correspondentes do autovetor são maiores que 0 representam o **GRUPO 1**.
- Nós em que os valores correspondentes do autovetor são menores que 0 representam o **GRUPO 2**.



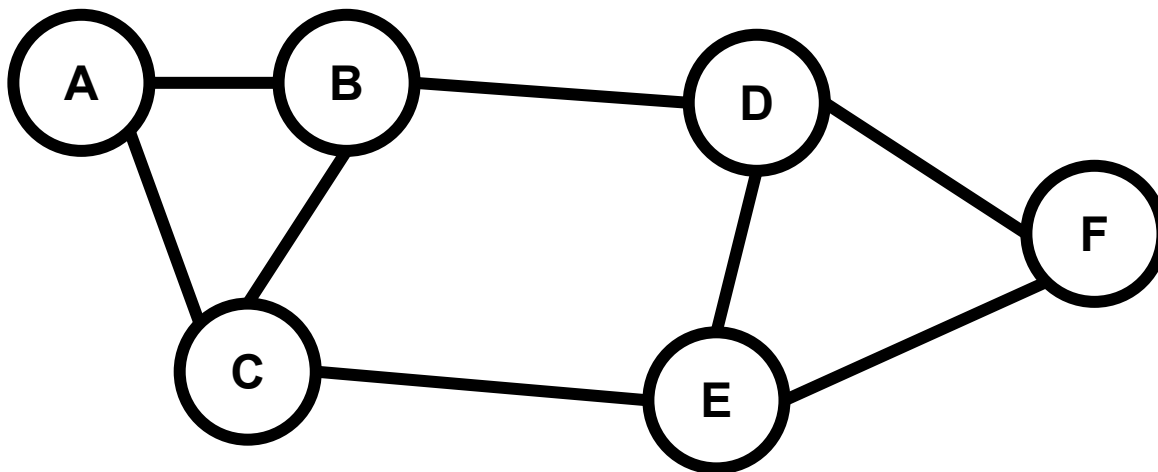
Particionamento Espectral

G

2					
	3				
		3			
			3		
				3	
					2

A

0	1	1	0	0	0
1	0	1	1	0	0
1	1	0	0	1	0
0	1	0	0	1	1
0	0	1	1	0	1
0	0	0	1	1	0



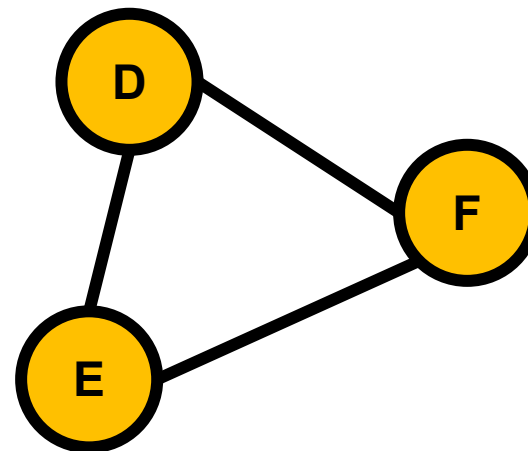
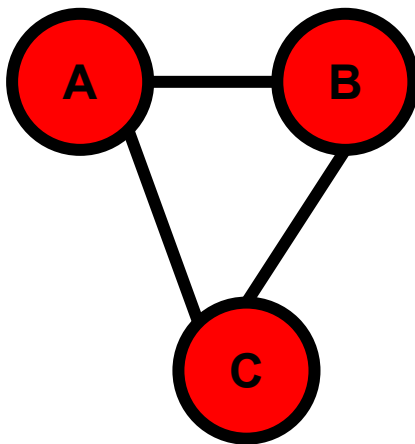
Particionamento Espectral

L

2	-1	-1	0	0	0
-1	3	-1	-1	0	0
-1	-1	3	0	-1	0
0	-1	0	3	-1	-1
0	0	-1	-1	3	-1
0	0	0	-1	-1	2

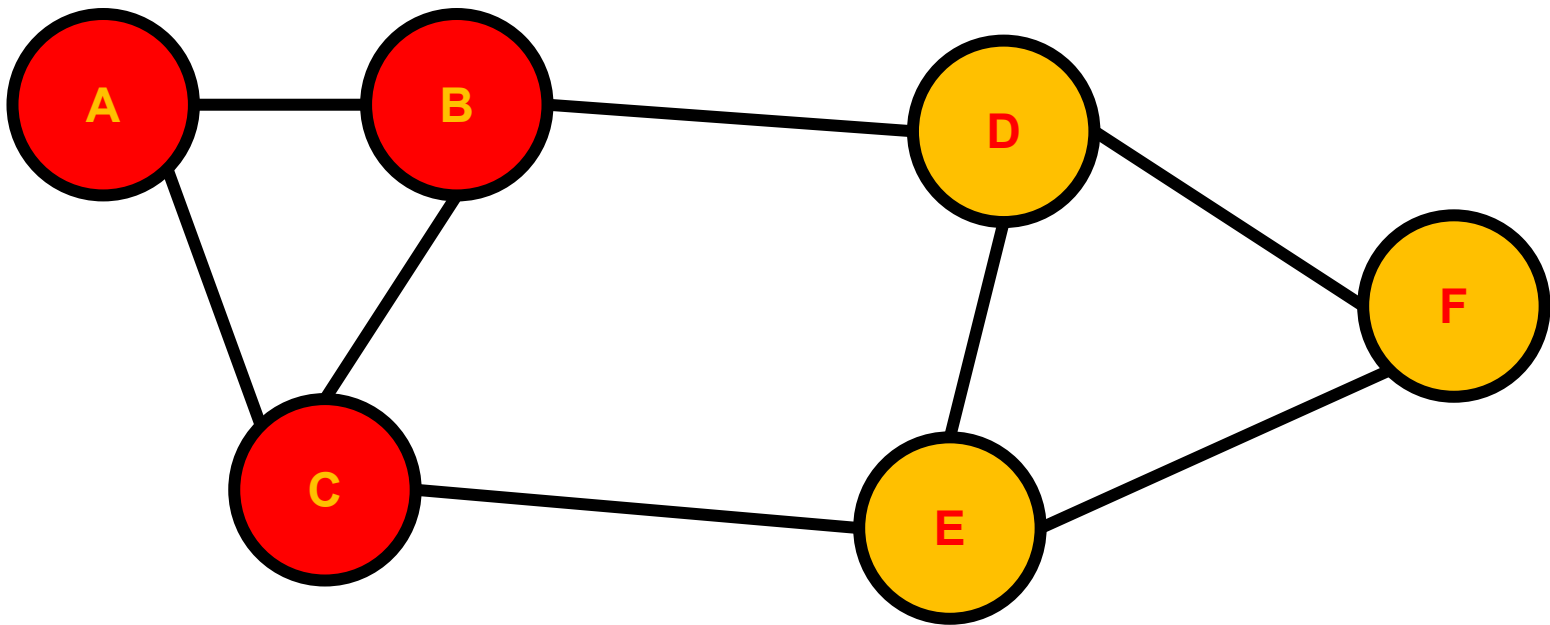
Autovetor Fiedler

0,5773	Grupo 1
0,2887	Grupo 1
0,2887	Grupo 1
-0,2887	Grupo 2
-0,2887	Grupo 2
-0,5773	Grupo 2



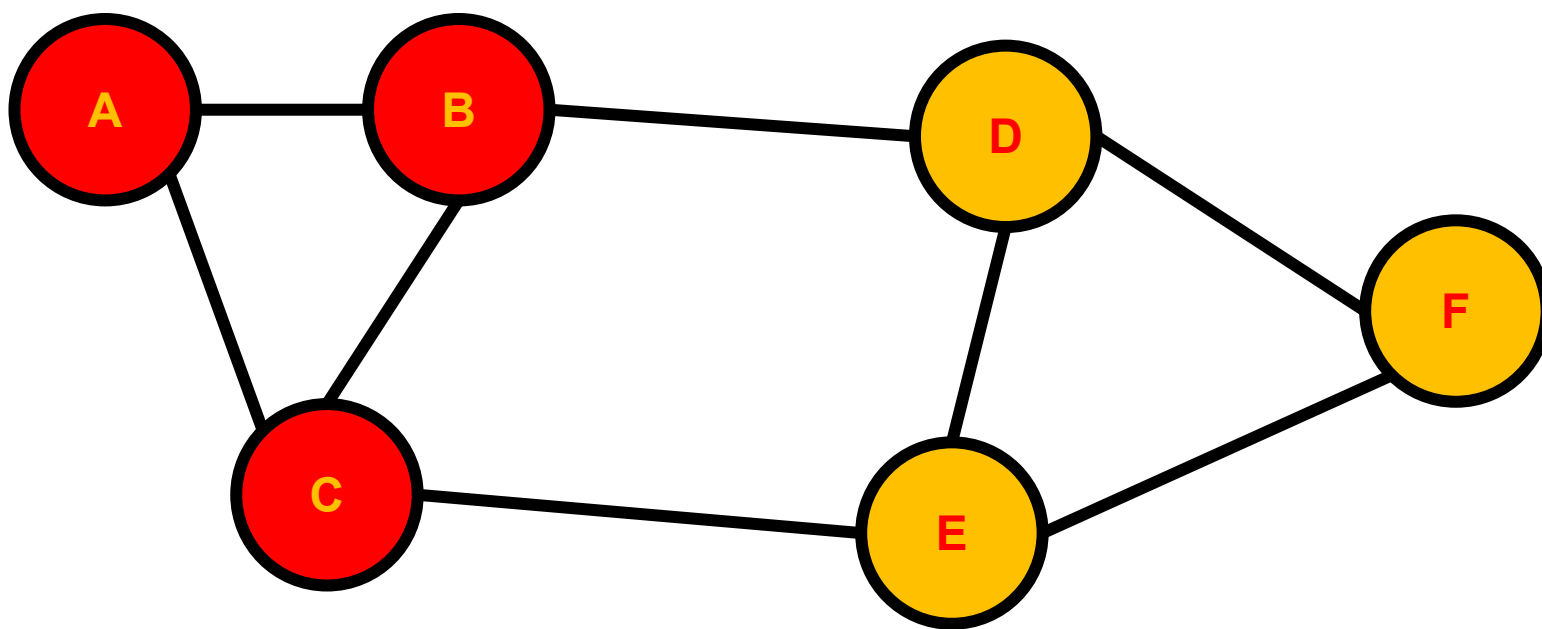
Particionamento Espectral

Os valores dos elementos do autovetor indica a conectividade de cada nó de acordo com sua vizinhança.



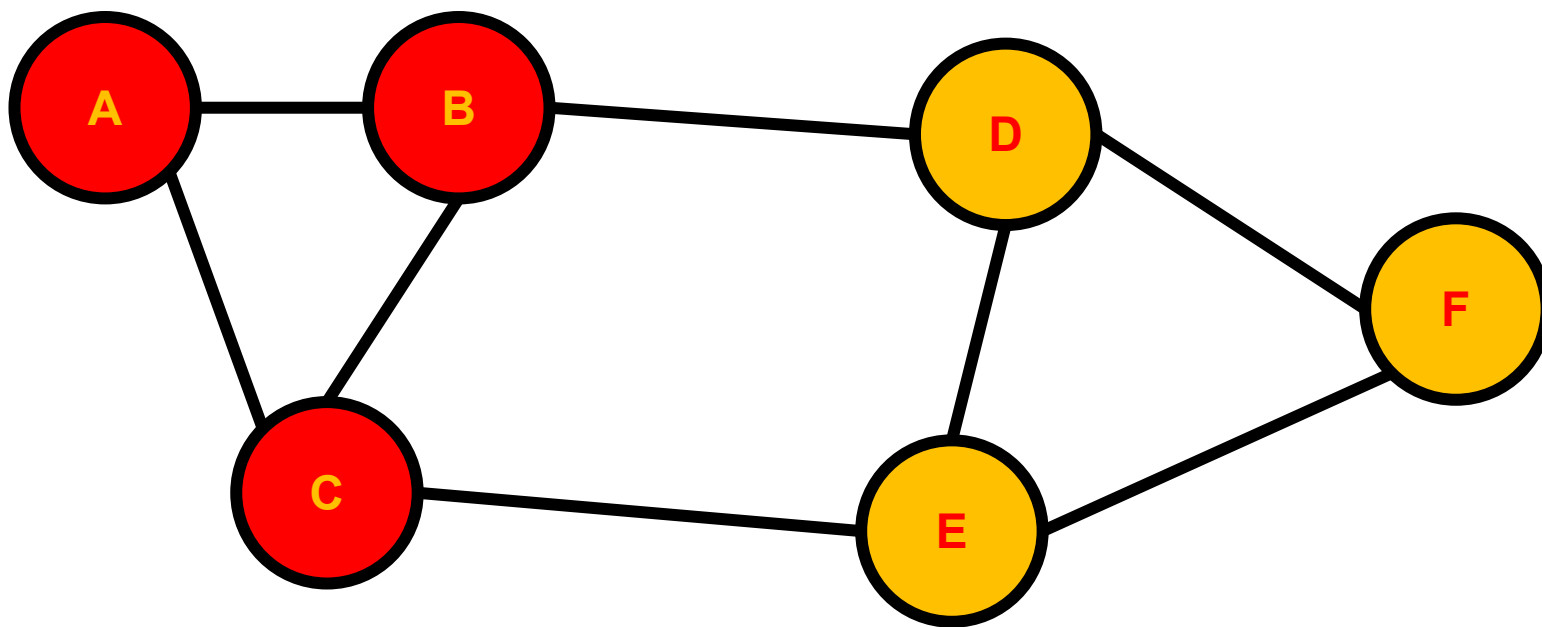
Particionamento Espectral

O segundo autovetor indica a conectividade se a rede fosse particionada em dois, o terceiro se ela fosse particionada em três, e assim por diante.



Particionamento Espectral

Eles tem relação com a frequência de onda em um fio em vibração.



Particionamento Espectral

A. Pothen, H. Simon, K.-P. Liou, "Partitioning sparse matrices with eigenvectors of graphs", SIAM J. Mat. Anal. Appl. 11:430-452 (1990)

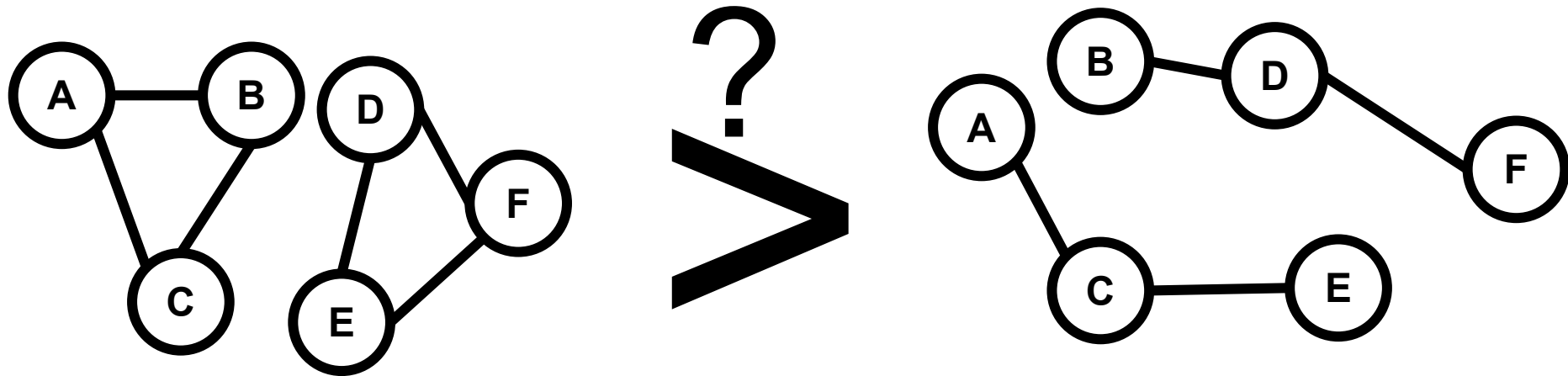
M. Fiedler, "Algebraic Connectivity of Graphs", Czech. Math. J., 23:298-305 (1973)

M. Fiedler, Czech. Math. J., 25:619-637 (1975)



Modularidade

Muitos problemas podem ser resolvidos matematicamente se existir uma função-objetivo que mede a qualidade de determinada solução.



Para medir a qualidade de particionamento de redes uma das medidas mais populares é a **MODULARIDADE.**



Modularidade

- Dada uma divisão do grafo em g grupos, e uma matriz \mathbf{E} de tamanho $g \times g$ onde e_{ij} representa a fração de arestas na rede não-particionada que liga o grupo i ao grupo j (*diferente de i*):

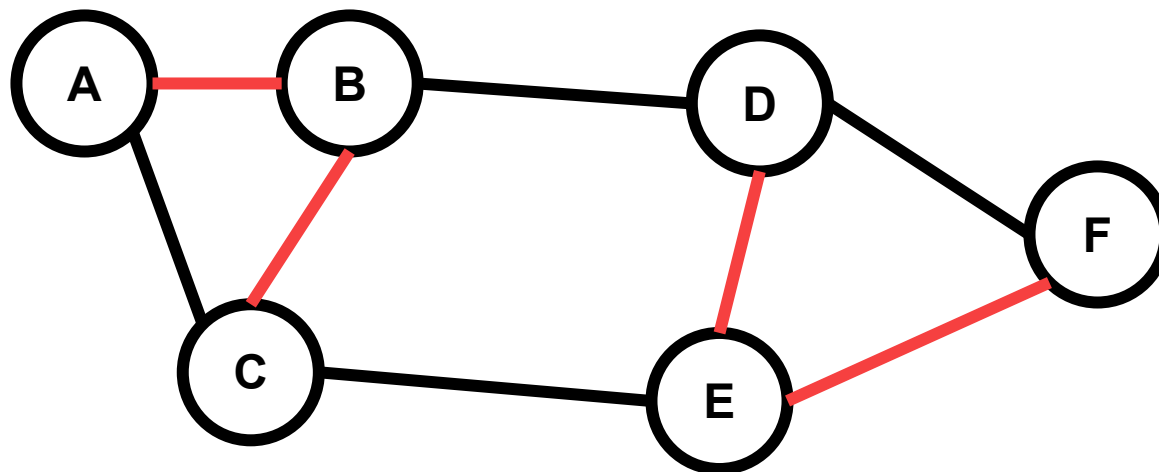
$$Q = \sum_i e_{ii} - \left[\sum_j e_{i,j} \right]^2$$



Maximizar modularidade

Vamos gerar 2 grupos aleatórios. Ex.: $\{A,C,E\}$ e $\{B,D,F\}$ e calcular a modularidade.

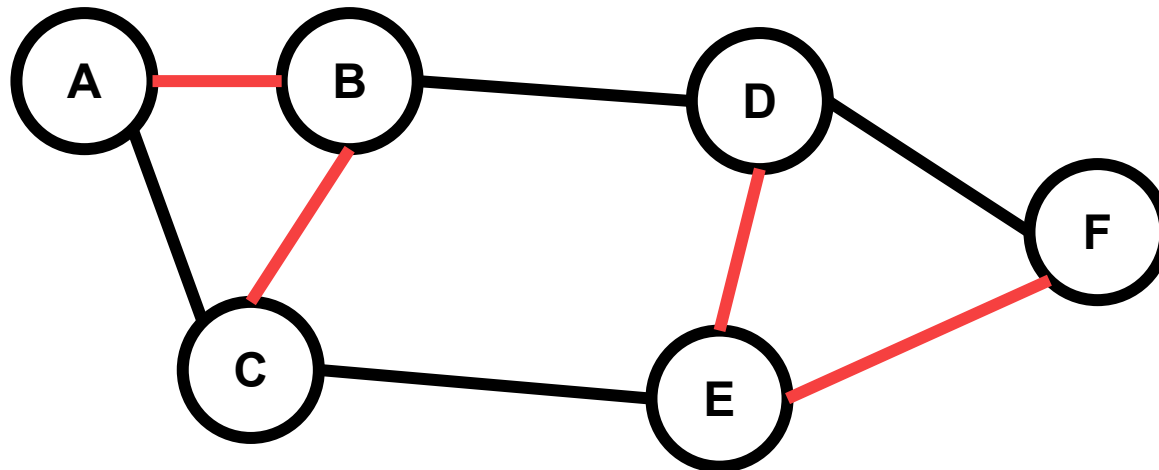
As arestas vermelhas são as arestas que seriam removidas para formar tal partição.



Maximizar modularidade

Existem 4 arestas ao todo ligando cada grupo com ele mesmo em um total de 16 arestas, contando ida e volta.

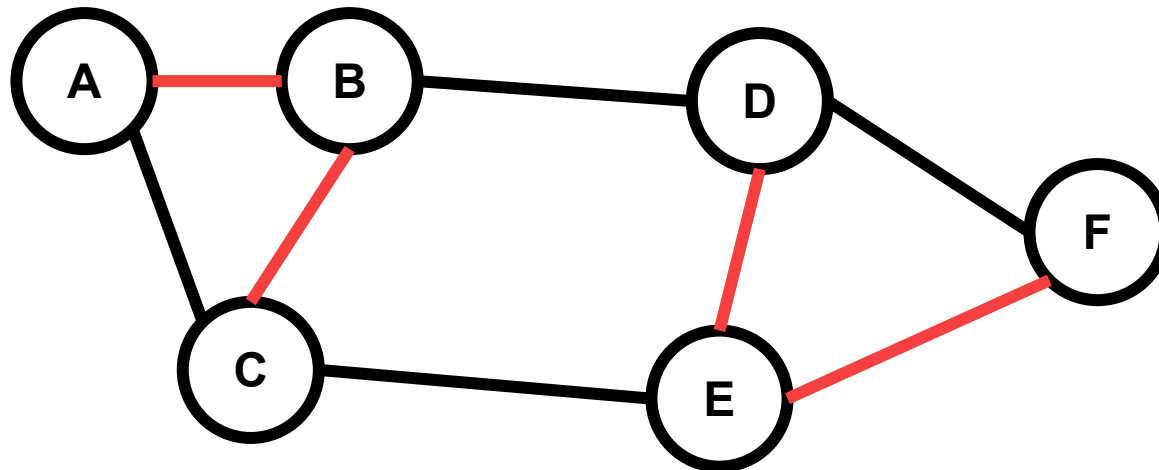
$$e_{11} = e_{22} = 2/8$$



Maximizar modularidade

Existem 4 arestas ao todo ligando cada grupo com o outro mesmo em um total de 16 arestas, contando ida e volta.

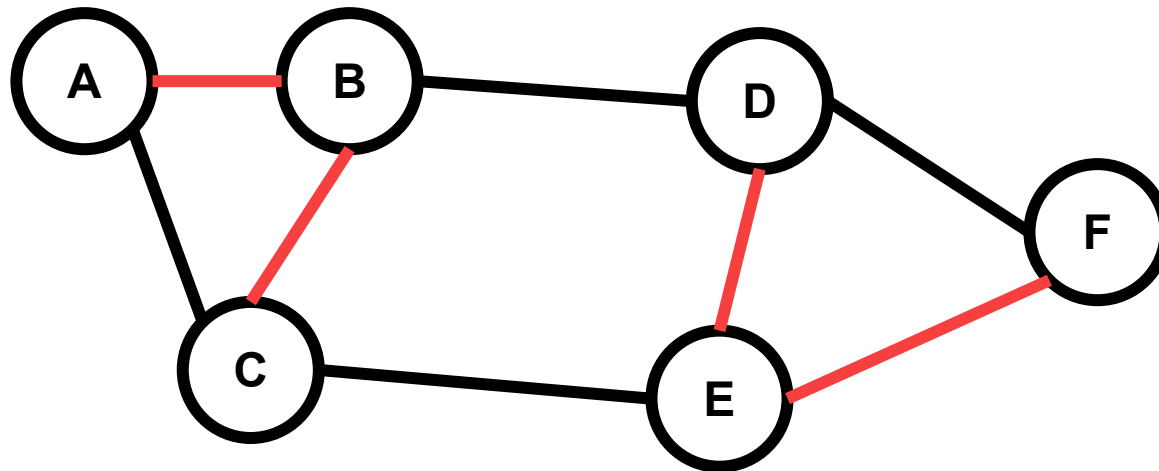
$$e_{12} = e_{21} = 2/8$$



Maximizar modularidade

$$\dot{E} = \begin{bmatrix} 2/8 & 2/8 \\ 2/8 & 2/8 \end{bmatrix}$$

$$G = [e_{11} - (e_{11}+e_{12})^2] + [e_{22} - (e_{21}+e_{22})^2]$$

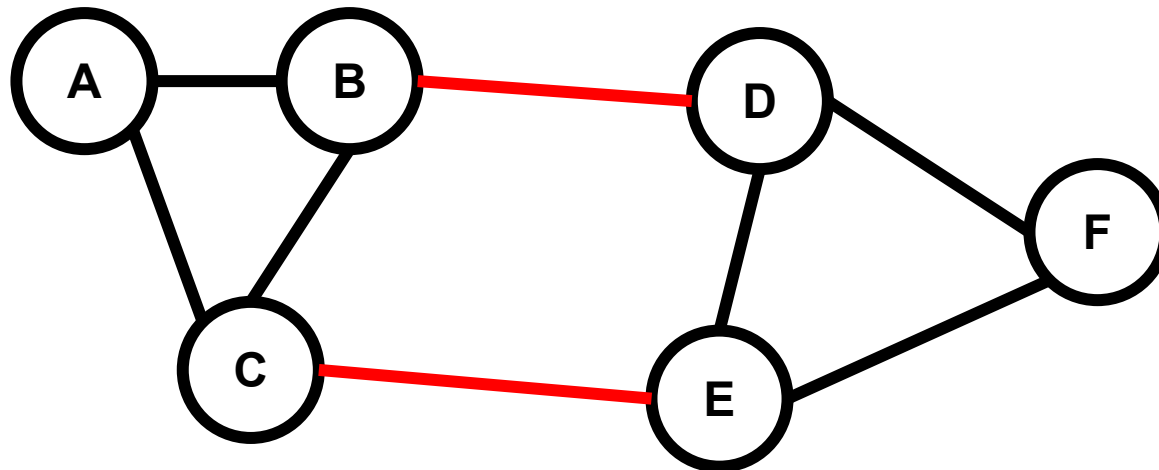


$$G = (2/8 - (4/8)^2) + (2/8 - (4/8)^2) = 0,00$$



Maximizar modularidade

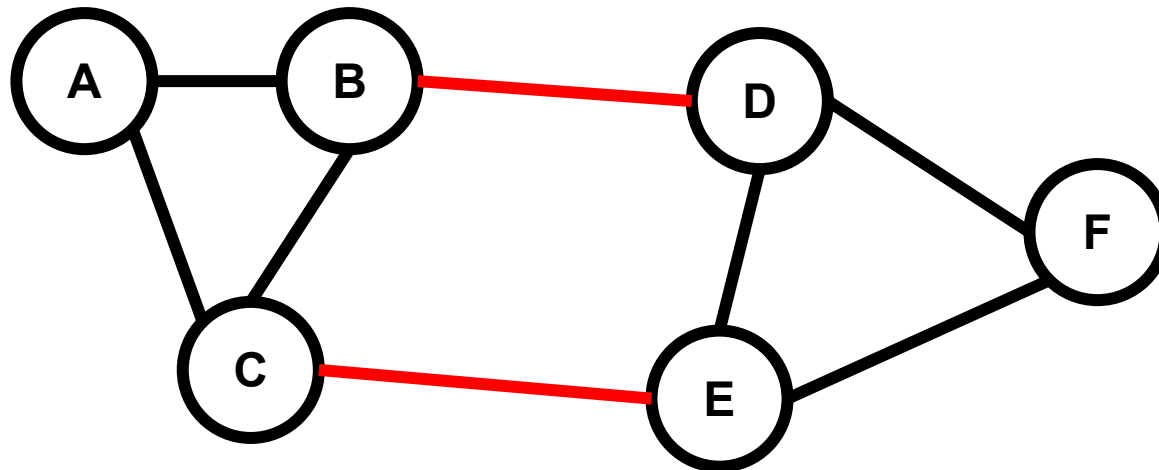
Se trocarmos o nó E pelo nó B, ficaremos com os grupos $\{A,B,C\}$ e $\{D,E,F\}$.



Maximizar modularidade

Existem 6 arestas ao todo ligando cada grupo com ele mesmo em um total de 16 arestas, contando ida e volta.

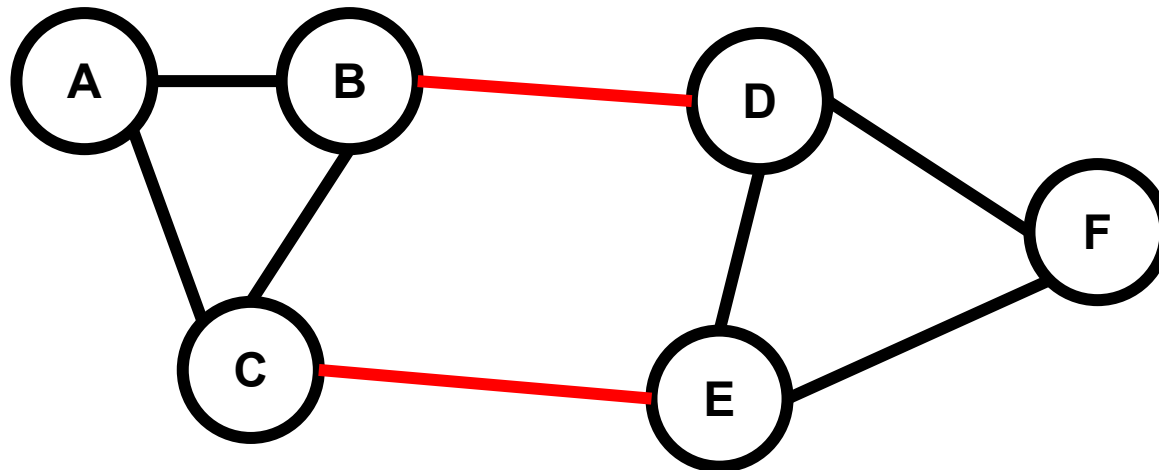
$$e_{11} = e_{22} = 3/8$$



Maximizar modularidade

Existem 2 arestas ao todo ligando cada grupo com o outro em um total de 16 arestas, contando ida e volta.

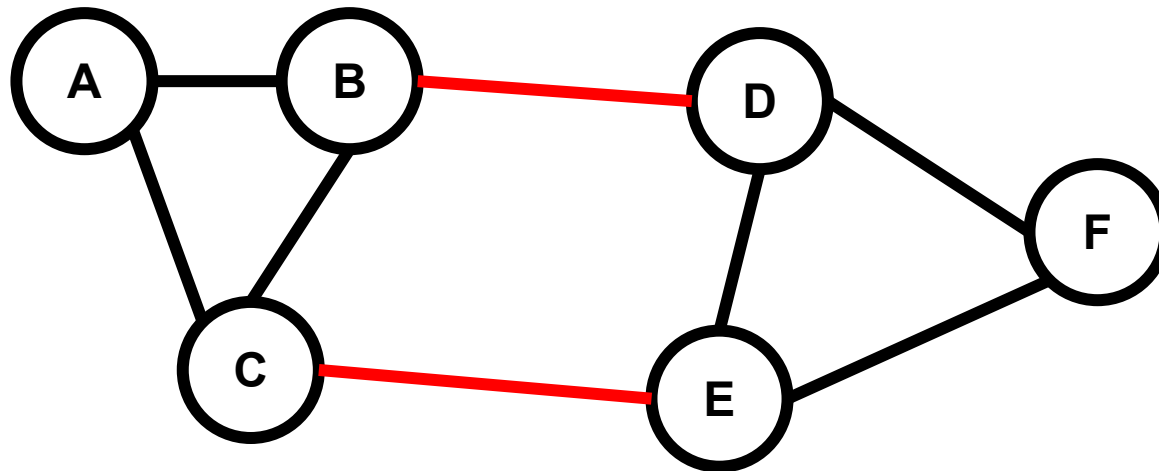
$$e_{12} = e_{21} = 1/8$$



Maximizar modularidade

$$\dot{E} = \begin{bmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{bmatrix}$$

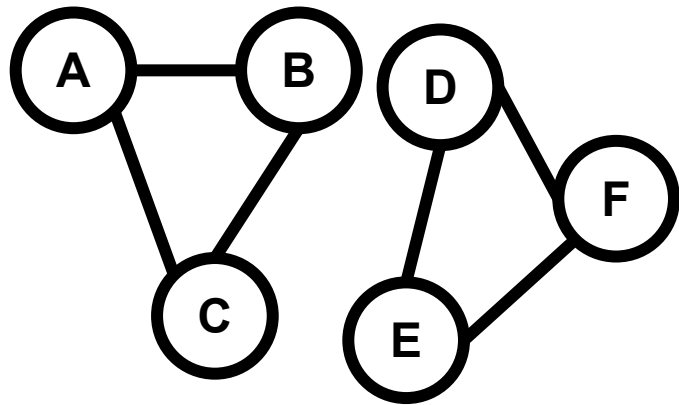
$$G = [e_{11} - (e_{11}+e_{12})^2] + [e_{22} - (e_{21}+e_{22})^2]$$



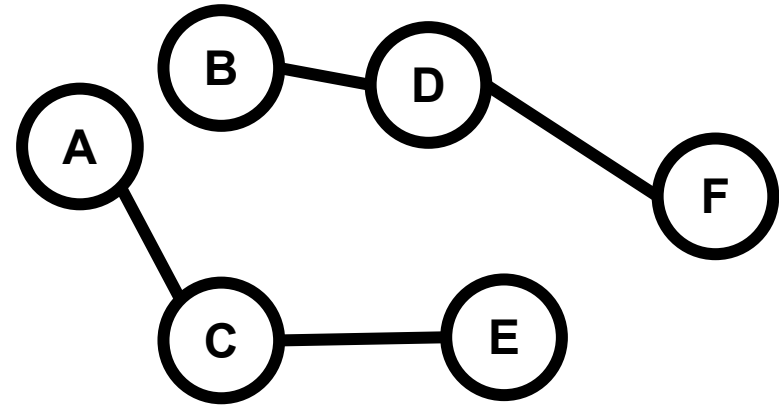
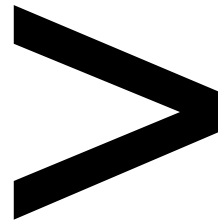
$$G = (3/8 - (4/8)^2) + (3/8 - (4/8)^2) = 0,25$$



Modularidade



0,25



0,00



Algoritmo de Maximização de Modularidade

Um algoritmo simples para encontrar uma partição com alta modularidade:

- ❑ Inicie com uma partição inicial e calcule G ;
- ❑ Calcule G para todas as possíveis trocas de nós entre grupos (inclusive simplesmente mover um nó de um grupo ao outro);
- ❑ Se alguma troca for favorável, efetue a troca;
- ❑ Repita até não existir nenhuma outra troca favorável



Algoritmo de Maximização de Modularidade

A modularidade permite uso de algoritmos mais simples conhecidos como métodos heurísticos.

Eles são capazes de lidar com grandes problemas utilizando de forma eficiente o processamento computacional MAS não garante encontrar a melhor solução possível.

É utilizado pela Amazon para determinar “quem comprou X também comprou Y”.

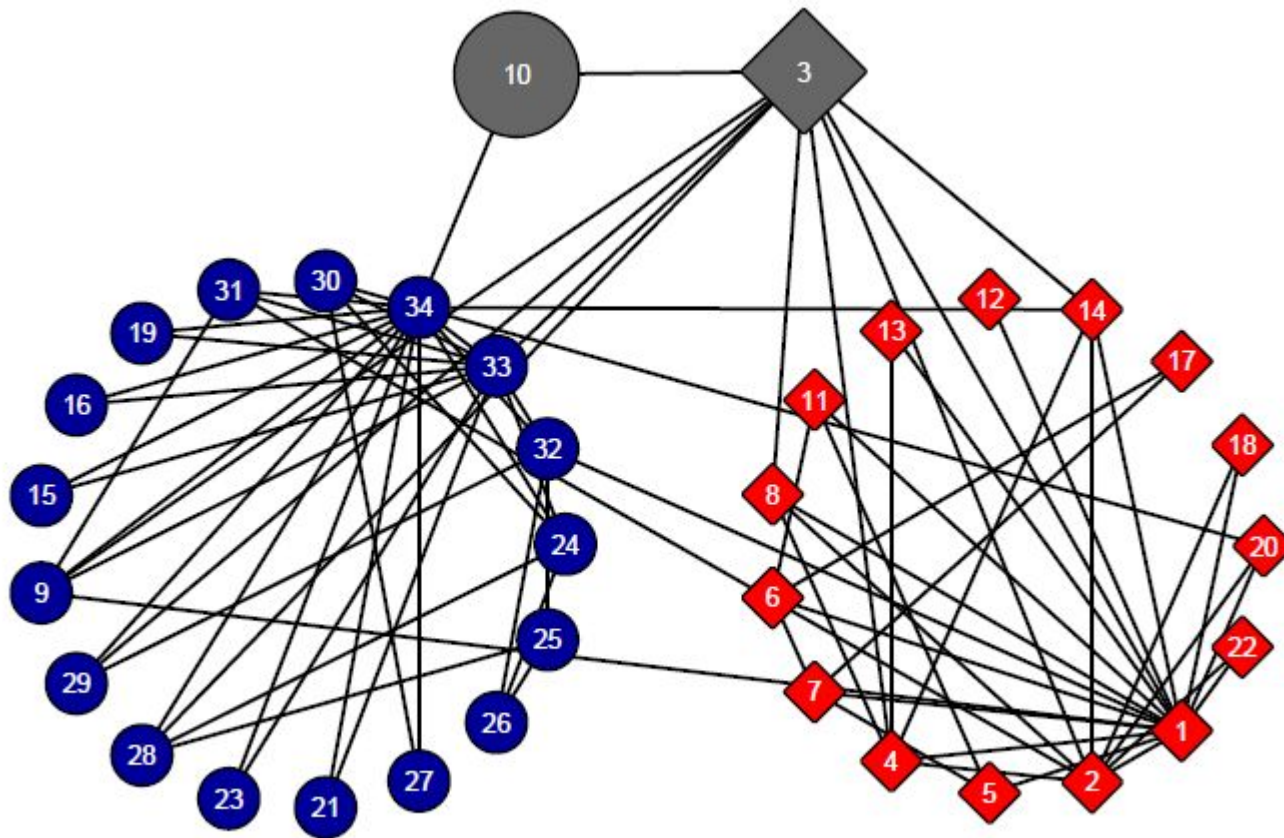




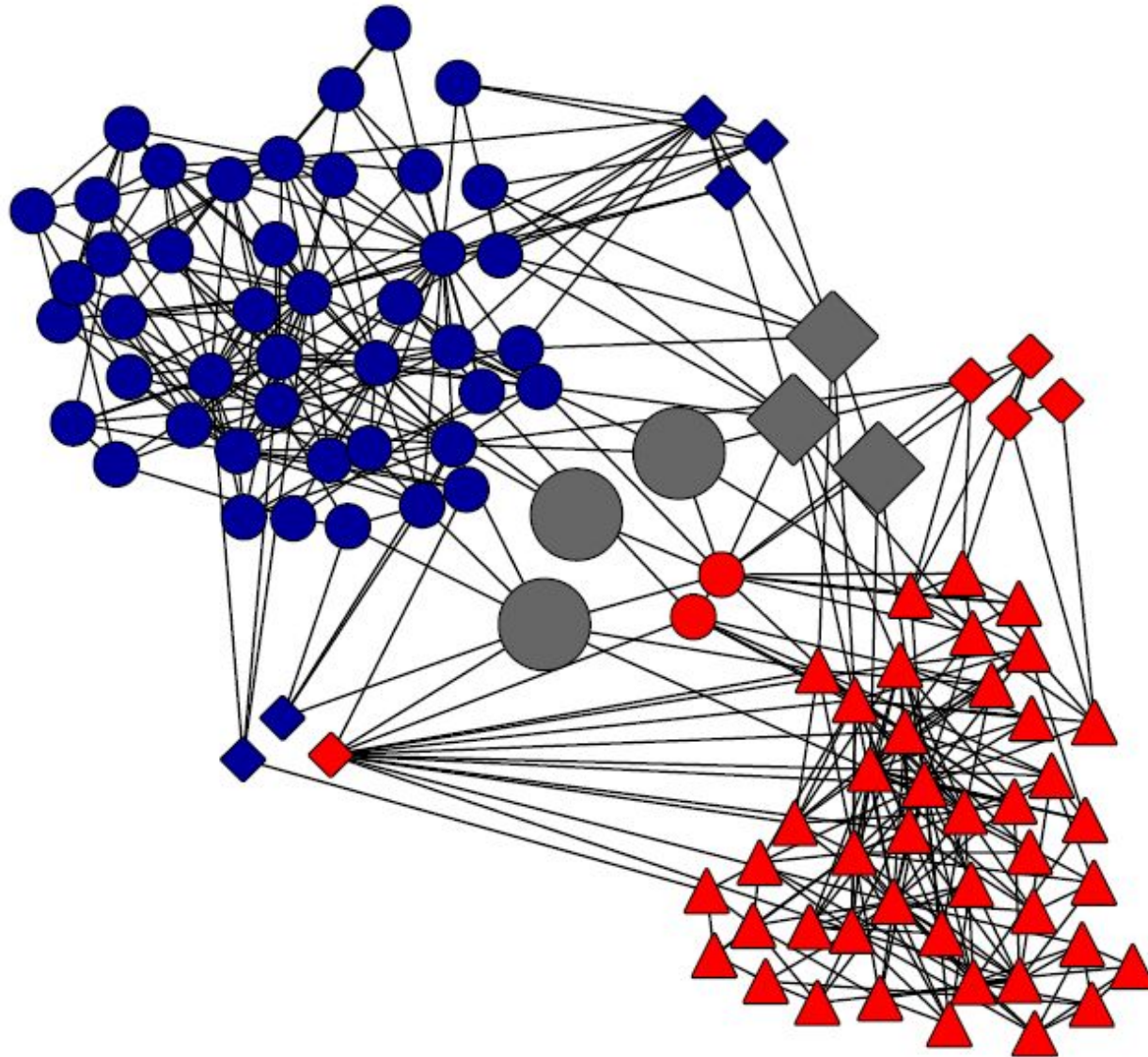
Universidade Federal do ABC

COMUNIDADES EM REDES REAIS

Karate Club



Krebs Political Books



American Football

