Introdução às Redes de Interação – MO804 (MC908)
Centralidade de vértices e arestas

Prof. Dr. Ruben Interian

Instituto de Computação, UNICAMP

Resumo

Resumo

Objetivo

Responder a **pergunta**:

- O que é um **vértice importante em uma rede**?
- O que é uma **aresta importante em uma rede**?

Resumo

Centralidade

O que é um **vértice importante em uma rede**?

- **O que é importância?**

Centralidade

O que é um **vértice importante em uma rede**?

- **O que é importância?– Depende!** Não existe uma definição geral de centralidade. Há muitas formas de responder essa pergunta.

Centralidade

O que é um **vértice importante em uma rede**?

- **O que é importância?– Depende!** Não existe uma definição geral de centralidade. Há muitas formas de responder essa pergunta.
- O vértice ou aresta importante em uma aplicação pode não ser importante em outras. **Exemplo**: ator mais famoso **vs** ator que conhece diretores de cinema.

Centralidade

O que é um **vértice importante em uma rede**?

- **O que é importância?– Depende!** Não existe uma definição geral de centralidade. Há muitas formas de responder essa pergunta.
- O vértice ou aresta importante em uma aplicação pode não ser importante em outras. **Exemplo**: ator mais famoso **vs** ator que conhece diretores de cinema.
- Como não há consenso, foram propostas diferentes formas de avaliar a relevância de vértices e arestas na rede, chamados **índices** ou **medidas** de **centralidade**.

Centralidade

O que é um **vértice importante em uma rede**?

- **O que é importância?– Depende!** Não existe uma definição geral de centralidade. Há muitas formas de responder essa pergunta.
- O vértice ou aresta importante em uma aplicação pode não ser importante em outras. **Exemplo**: ator mais famoso **vs** ator que conhece diretores de cinema.
- Como não há consenso, foram propostas diferentes formas de avaliar a relevância de vértices e arestas na rede, chamados **índices** ou **medidas** de **centralidade**.
- Um **índice de centralidade** é uma função $f : V \to \mathcal{R}$ que atribui um valor real $f(v)$ a cada um dos vértices da rede.

Resumo

Centralidade de grau

A primeira e mais obvia medida de centralidade de um vértice é o seu próprio **grau**.

- **Centralidade de grau**, *degree centrality*: $f(v_i) = d(v_i) = k_i$.

Centralidade de grau

A primeira e mais obvia medida de centralidade de um vértice é o seu próprio **grau**.

- **Centralidade de grau**, *degree centrality*: $f(v_i) = d(v_i) = k_i$.
- O vértice de maior **grau** é o mais importante.
  Quanto maior o grau, maior a importância do vértice.

Centralidade de grau

A primeira e mais obvia medida de centralidade de um vértice é o seu próprio **grau**.

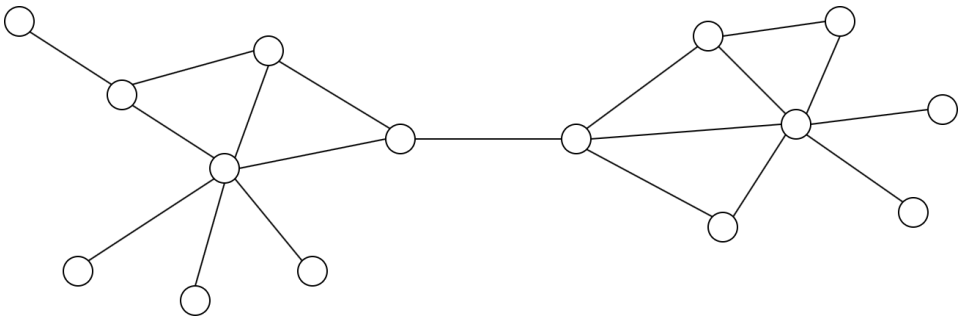
- **Centralidade de grau**, *degree centrality*: $f(v_i) = d(v_i) = k_i$.
- O vértice de maior **grau** é o mais importante.
  Quanto maior o grau, maior a importância do vértice.
- **Vantagens**: simplicidade.

Centralidade de grau

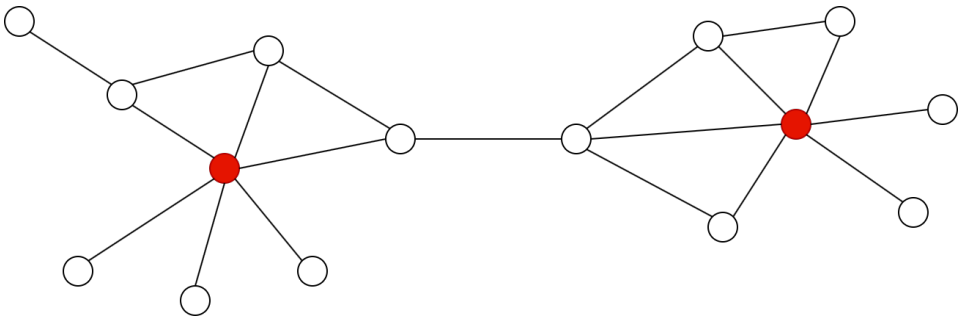
A primeira e mais obvia medida de centralidade de um vértice é o seu próprio **grau**.

- **Centralidade de grau**, *degree centrality*: $f(v_i) = d(v_i) = k_i$.
- O vértice de maior **grau** é o mais importante.
  Quanto maior o grau, maior a importância do vértice.
- **Vantagens**: simplicidade.
- **Limitações**: é uma medida **local**.
  O vértice pode estar muito longe do "centro" da rede.

Centralidade de grau



Centralidade de grau



Centralidade k-núcleo

k-núcleo, *k-core measure*:

- É um índice de centralidade baseado no número de conexões, e pode ser visto como uma **generalização** do conceito de **grau**.

Centralidade k-núcleo

k-núcleo, *k-core measure*:

- É um índice de centralidade baseado no número de conexões, e pode ser visto como uma **generalização** do conceito de **grau**.
- O **k-núcleo** de um grafo G é um **subgrafo conexo maximal** no qual o grau de todos os vértices é pelo menos k.

Centralidade k-núcleo

k-núcleo, *k-core measure*:

- É um índice de centralidade baseado no número de conexões, e pode ser visto como uma **generalização** do conceito de **grau**.
- O **k-núcleo** de um grafo G é um **subgrafo conexo maximal** no qual o grau de todos os vértices é pelo menos k.
- $f(v_i) = kc(v_i) = k$ se o vértice v_i **pertence** a um k-núcleo e **não pertence** a um (k+1)-núcleo. Vértices mais centrais possuem valores maiores de $kc(v_i)$.

Centralidade k-núcleo

k-núcleo, *k-core measure*:

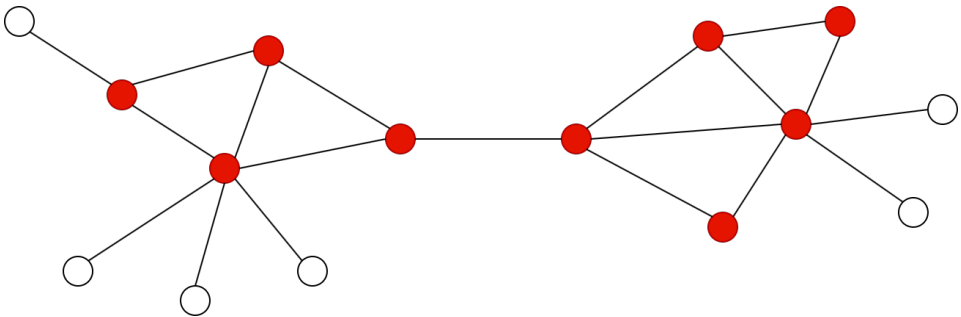
- É um índice de centralidade baseado no número de conexões, e pode ser visto como uma **generalização** do conceito de **grau**.
- O **k-núcleo** de um grafo G é um **subgrafo conexo maximal** no qual o grau de todos os vértices é pelo menos k.
- $f(v_i) = kc(v_i) = k$ se o vértice v_i **pertence** a um k-núcleo e **não pertence** a um (k+1)-núcleo. Vértices mais centrais possuem valores maiores de $kc(v_i)$.
- Como calcular o **k-núcleo** do grafo?

Centralidade k-núcleo

k-núcleo, *k-core measure*:

- É um índice de centralidade baseado no número de conexões, e pode ser visto como uma **generalização** do conceito de **grau**.
- O **k-núcleo** de um grafo G é um **subgrafo conexo maximal** no qual o grau de todos os vértices é pelo menos k.
- $f(v_i) = kc(v_i) = k$ se o vértice v_i **pertence** a um k-núcleo e **não pertence** a um (k+1)-núcleo. Vértices mais centrais possuem valores maiores de $kc(v_i)$.
- Como calcular o **k-núcleo** do grafo? – **Remover**, iterativamente, os vértices cujo grau é menor do que k, até que não seja possível remover nenhum outro vértice.

Centralidade, 2-núcleo



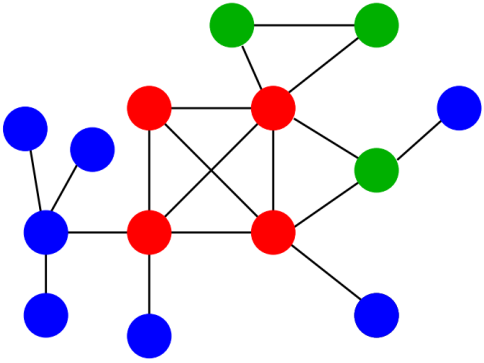
Centralidade, k-núcleo

Grafo onde cores representam:

- Azul: $kc(v_i) = 1$;
- Verde: $kc(v_i) = 2$;
- Vermelho: $kc(v_i) = 3$.

Limitações do k-núcleo:

- É local; muitos nós com o mesmo valor $kc(v_i)$; vários k-núcleos.



Resumo

Closeness centrality

Centralidade de proximidade, *closeness centrality*:

- **Ideia**: um vértice tem **centralidade alta** quando está **próximo** (em termos de distância) dos outros vértices da rede.

Closeness centrality

Centralidade de proximidade, *closeness centrality*:

- **Ideia**: um vértice tem **centralidade alta** quando está **próximo** (em termos de distância) dos outros vértices da rede. Media das **distâncias** até v_i: $\frac{\sum_{j \neq i} d(v_i, v_j)}{n-1}$.

Closeness centrality

Centralidade de proximidade, *closeness centrality*:

- **Ideia**: um vértice tem **centralidade alta** quando está **próximo** (em termos de distância) dos outros vértices da rede. Media das **distâncias** até v_i: $\frac{\sum_{j \neq i} d(v_i, v_j)}{n-1}$.
- Calculamos **o inverso**: quanto maior, melhor.

$$f(v_i) = C(v_i) = \frac{n-1}{\sum_{j \neq i} d(v_i, v_j)}.$$

Closeness centrality

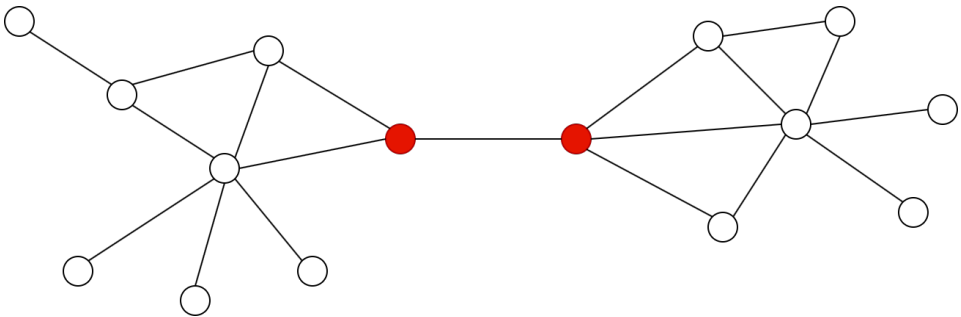
Centralidade de proximidade, *closeness centrality*:

- **Ideia**: um vértice tem **centralidade alta** quando está **próximo** (em termos de distância) dos outros vértices da rede. Media das **distâncias** até v_i: $\frac{\sum_{j \neq i} d(v_i, v_j)}{n-1}$.
- Calculamos **o inverso**: quanto maior, melhor.

$$f(v_i) = C(v_i) = \frac{n-1}{\sum_{j \neq i} d(v_i, v_j)}.$$

- **Vantagens**: ideia intuitiva, simples, de fácil interpretação.
- **Limitações**: Custo computacional.
  *Em diversas redes (aleatórias, *small world*), distâncias são tipicamente pequenas.

Closeness centrality



Closeness centrality

Usando *closeness centrality* **na prática**:

- Em **grafos não conexos**, precisamos adaptar a ideia, por exemplo, definindo a distância até um vértice inalcançável.

Closeness centrality

Usando *closeness centrality* **na prática**:

- Em **grafos não conexos**, precisamos adaptar a ideia, por exemplo, definindo a distância até um vértice inalcançável.
- Para evitar o custo computacional quadrático, em redes grandes e esparsas podemos **aproximar os valores** das distâncias usando uma heurística ou um algoritmo de aproximação.

Closeness centrality

Usando *closeness centrality* **na prática**:

- Em **grafos não conexos**, precisamos adaptar a ideia, por exemplo, definindo a distância até um vértice inalcançável.
- Para evitar o custo computacional quadrático, em redes grandes e esparsas podemos **aproximar os valores** das distâncias usando uma heurística ou um algoritmo de aproximação.
- Usando *closeness centrality*, podemos detectar vértices que estão mais próximos dos demais, e que têm maior **capacidade de espalhar informações** pela rede.

Resumo

Betweenness centrality

Centralidade de intermediação, *betweenness centrality*:

- **Ideia**: um vértice v_i tem **centralidade alta** quando há muitos caminhos mais curtos entre os vértices que passam por v_i.

Betweenness centrality

Centralidade de intermediação, *betweenness centrality*:

- **Ideia**: um vértice v_i tem **centralidade alta** quando há muitos caminhos mais curtos entre os vértices que passam por v_i.
- Sejam s e t dois vértices, diferentes de v_i. Pode haver **vários** caminhos mais curtos (equivalentes) entre s e t.

Betweenness centrality

Betweenness de um vértice v_i avalia a quantidade de vezes que v_i aparece nos caminhos mais curtos entre outros dois vértices do grafo:

$$f(v_i) = B(v_i) = \sum_{\substack{s \neq v_i \\ t \neq v_i}} \frac{\sigma(s, v_i, t)}{\sigma(s, t)},$$

onde $\sigma(s, t)$ é o **número de caminhos mínimos** entre s e t,
e $\sigma(s, v_i, t)$ é o **número desses caminhos** que passam por v_i.

Betweenness centrality

Betweenness de um vértice v_i avalia a quantidade de vezes que v_i aparece nos caminhos mais curtos entre outros dois vértices do grafo:

$$f(v_i) = B(v_i) = \sum_{\substack{s \neq v_i \\ t \neq v_i}} \frac{\sigma(s, v_i, t)}{\sigma(s, t)},$$

onde $\sigma(s, t)$ é o **número de caminhos mínimos** entre s e t,
e $\sigma(s, v_i, t)$ é o **número desses caminhos** que passam por v_i.

Observação: podemos dividir por $\frac{(n-1)(n-2)}{2}$, para sempre ter $B(v_i)$ entre 0 e 1.
(Irrelevante se queremos comparar os valores de $B(v_i)$ para vértices do mesmo grafo.)

Betweenness centrality

Betweenness de um vértice v_i avalia a quantidade de vezes que v_i aparece nos caminhos mais curtos entre outros dois vértices do grafo:

$$f(v_i) = B(v_i) = \sum_{\substack{s \neq v_i \\ t \neq v_i}} \frac{\sigma(s, v_i, t)}{\sigma(s, t)},$$

onde $\sigma(s, t)$ é o **número de caminhos mínimos** entre s e t,
e $\sigma(s, v_i, t)$ é o **número desses caminhos** que passam por v_i.

Observação: podemos dividir por $\frac{(n-1)(n-2)}{2}$, para sempre ter $B(v_i)$ entre 0 e 1.
(Irrelevante se queremos comparar os valores de $B(v_i)$ para vértices do mesmo grafo.)

Pergunta: Em que tipo de grafo podemos ter um vértice cujo valor de *betweenness* é igual ao máximo valor possível: $(n-1)(n-2)/2$, ou 1, no caso normalizado?

Betweenness centrality

Qual é a **diferença** entre *closeness* e *betweenness*?

Betweenness centrality

Qual é a **diferença** entre *closeness* e *betweenness*?

Se *closeness* é grande, *betweenness* pode ser pequena?

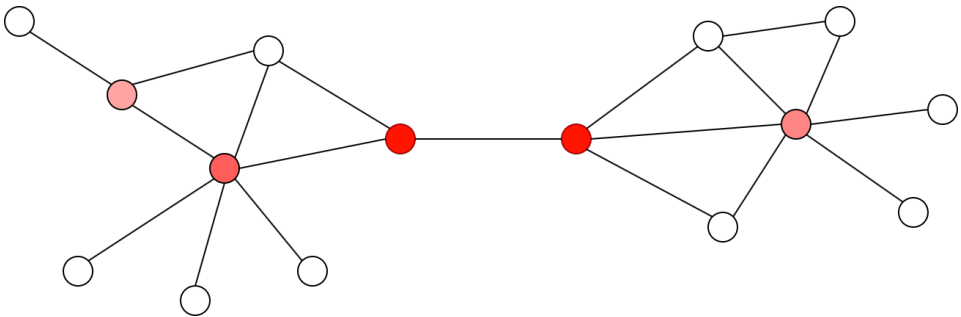
Betweenness centrality

Qual é a **diferença** entre *closeness* e *betweenness*?

Se *closeness* é grande, *betweenness* pode ser pequena?

Se *betweenness* é grande, *closeness* pode ser pequena?

Betweenness centrality



Betweenness centrality

Vantagens: Ideia simples e intuitiva. Permite avaliar até que ponto um vértice pode **controlar** a rede, pois mais **informação** vai passar através dele. Vértices com maior *betweenness* estão mais presentes nos fluxos de informação rápidos/eficientes da rede.

Betweenness centrality

Vantagens: Ideia simples e intuitiva. Permite avaliar até que ponto um vértice pode **controlar** a rede, pois mais **informação** vai passar através dele. Vértices com maior *betweenness* estão mais presentes nos fluxos de informação rápidos/eficientes da rede.

Limitações: custo computacional em grafos grandes. Para diminuir o custo, podemos adaptar a ideia usando **caminhos aleatórios**: quantos caminhos aleatórios contêm v_i?

Betweenness centrality

Centralidade de intermediação para as arestas, *edge betweenness centrality*:

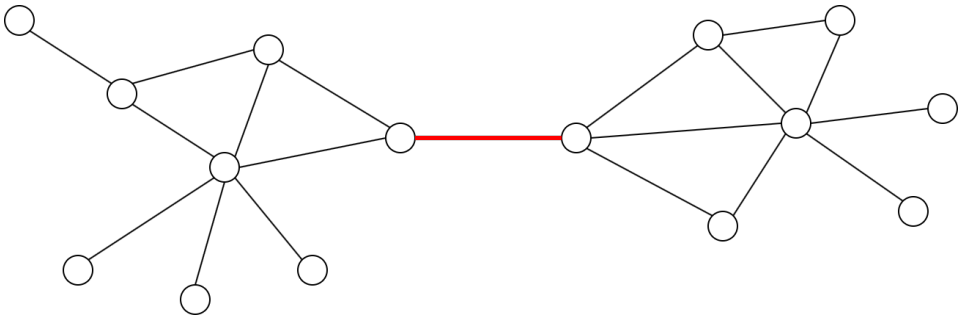
- **Ideia**: uma aresta e tem **centralidade alta** quando há muitos caminhos entre os diferentes vértices que passam através de e.

Betweenness centrality

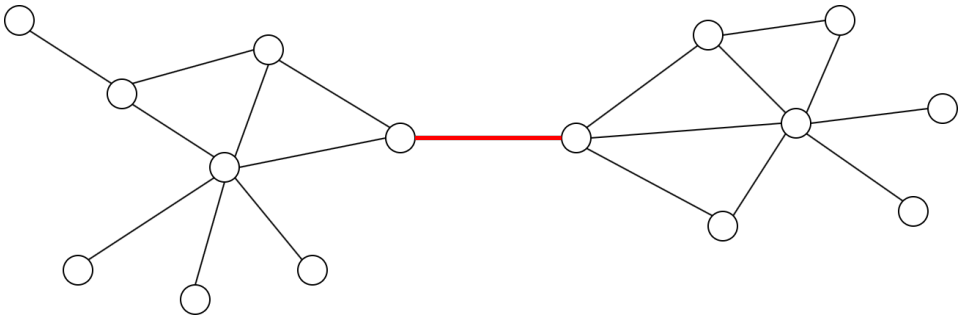
Centralidade de intermediação para as arestas, *edge betweenness centrality*:

- **Ideia**: uma aresta e tem **centralidade alta** quando há muitos caminhos entre os diferentes vértices que passam através de e.
- *Edge betweenness* pode ser calculado dividindo o número de caminhos mínimos entre vértices no grafo **que passam pela aresta** entre o número **total de caminhos mínimos**.

Betweenness centrality



Betweenness centrality



Betweenness de arestas pode ser usado para **resolver algum problema conhecido**?...

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

1. **Calcular** os valores *betweenness* para todas as arestas na rede.

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

1. **Calcular** os valores *betweenness* para todas as arestas na rede.
2. **Remover** a aresta com o maior valor calculado.

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

1. **Calcular** os valores *betweenness* para todas as arestas na rede.
2. **Remover** a aresta com o maior valor calculado.
3. **Recalcular** os valores *betweenness* das arestas.

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

1. **Calcular** os valores *betweenness* para todas as arestas na rede.
2. **Remover** a aresta com o maior valor calculado.
3. **Recalcular** os valores *betweenness* das arestas.
4. **Repetir** passos 2 e 3 até não sobrar nenhuma aresta.

Betweenness centrality

A ideia de *edge betweenness* foi usada no famoso **algoritmo de Girvan-Newman**, um dos principais algoritmos de detecção de comunidades:

1. **Calcular** os valores *betweenness* para todas as arestas na rede.
2. **Remover** a aresta com o maior valor calculado.
3. **Recalcular** os valores *betweenness* das arestas.
4. **Repetir** passos 2 e 3 até não sobrar nenhuma aresta.

Comunidades: são formadas cada vez que removemos uma aresta e aumenta o número de componentes.

Betweenness centrality

Limitações: custo computacional em grafos grandes.

Betweenness de todas as arestas do grafo pode ser calculada em tempo $O(nm)$.

Complexidade do algoritmo: $O(m^2 n)$, $O(n^3)$ para grafos esparsos.

Betweenness centrality

Limitações: custo computacional em grafos grandes.
Betweenness de todas as arestas do grafo pode ser calculada em tempo $O(nm)$.
Complexidade do algoritmo: $O(m^2 n)$, $O(n^3)$ para grafos esparsos.

Melhoria simples: A etapa de **recalcular** *betweenness* é realizada somente para arestas da componente conexa que continha a aresta removida, ou nas duas novas componentes, se a remoção da aresta aumentou o número de componentes.

- Ou seja, recalculamos apenas os valores **afetados pela remoção da aresta**.
- O valor de *betweenness* de todas as outras arestas **permanece a mesmo**.
- Em grafos com comunidades bem definidas, essa melhoria já **acelera bastante o algoritmo**.

Betweenness centrality

Limitações: custo computacional em grafos grandes.
Betweenness de todas as arestas do grafo pode ser calculada em tempo $O(nm)$.
Complexidade do algoritmo: $O(m^2 n)$, $O(n^3)$ para grafos esparsos.

Melhoria simples: A etapa de **recalcular** *betweenness* é realizada somente para arestas da componente conexa que continha a aresta removida, ou nas duas novas componentes, se a remoção da aresta aumentou o número de componentes.

- Ou seja, recalculamos apenas os valores **afetados pela remoção da aresta**.
- O valor de *betweenness* de todas as outras arestas **permanece a mesmo**.
- Em grafos com comunidades bem definidas, essa melhoria já **acelera bastante o algoritmo**.

Melhoria caso há pesos nas arestas: Remover as arestas de menor peso. Por quê?

Betweenness centrality

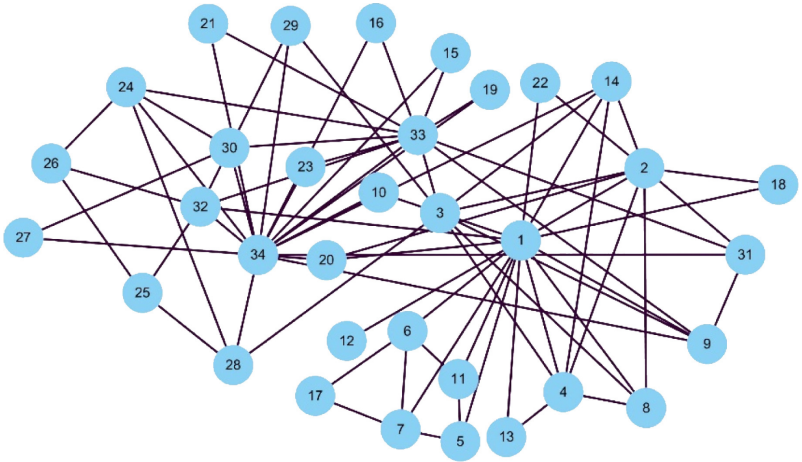
Limitações: custo computacional em grafos grandes.
Betweenness de todas as arestas do grafo pode ser calculada em tempo $O(nm)$.
Complexidade do algoritmo: $O(m^2 n)$, $O(n^3)$ para grafos esparsos.

Melhoria simples: A etapa de **recalcular** *betweenness* é realizada somente para arestas da componente conexa que continha a aresta removida, ou nas duas novas componentes, se a remoção da aresta aumentou o número de componentes.

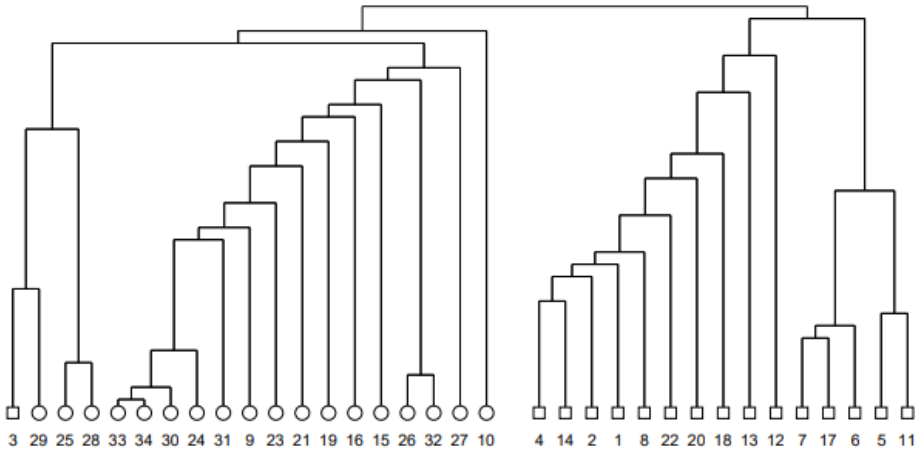
- Ou seja, recalculamos apenas os valores **afetados pela remoção da aresta**.
- O valor de *betweenness* de todas as outras arestas **permanece a mesmo**.
- Em grafos com comunidades bem definidas, essa melhoria já **acelera bastante o algoritmo**.

Melhoria caso há pesos nas arestas: Remover as arestas de menor peso. Por quê?
– Vínculos fracos costumam ser pontes ou pontes locais! (mas não temos garantias...)

Betweenness centrality: clube de karatê do Zachary



Betweenness centrality



Betweenness centrality

O algoritmo gera diversas partições. Como escolher **a melhor delas**?

Betweenness centrality

O algoritmo gera diversas partições. Como escolher **a melhor delas**?

– O critério mais usado é escolher a partição com a maior **modularidade**!

$$Q = \frac{1}{2m} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(s_i, s_j)$$

Betweenness centrality

O algoritmo gera diversas partições. Como escolher **a melhor delas**?

– O critério mais usado é escolher a partição com a maior **modularidade**!

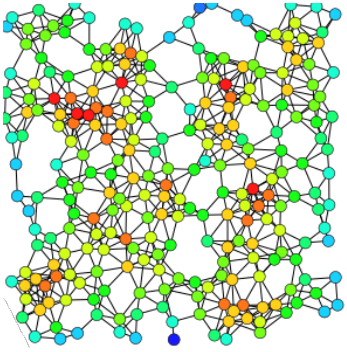
$$Q = \frac{1}{2m} \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2m} \right) \delta(s_i, s_j)$$

Há muitas **variantes** do algoritmo de Girvan-Newman. Porém, em aplicações práticas, o algoritmo original fornece melhores resultados do que variantes nas quais trocamos *betweenness* por medidas mais simples de calcular. Estudos mostram que o **passo 3**, a atualização do valor do *betweenness*, é essencial para os bons resultados.

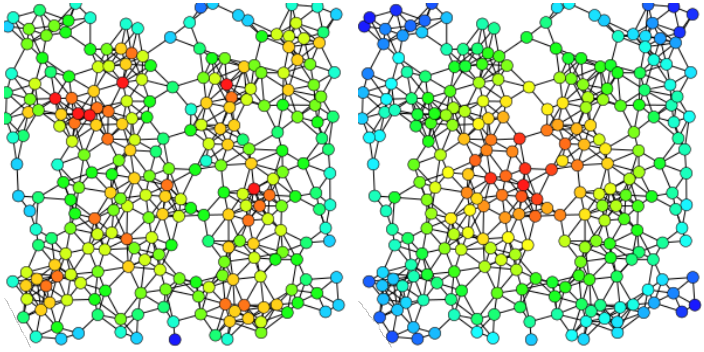
O algoritmo está implementado em diversos **módulos** e **bibliotecas**.

M. Girvan & M. Newman: "Community structure in social and biological networks"
+12000 citações no Scopus, +19000 no Google Scholar (desde 2002).

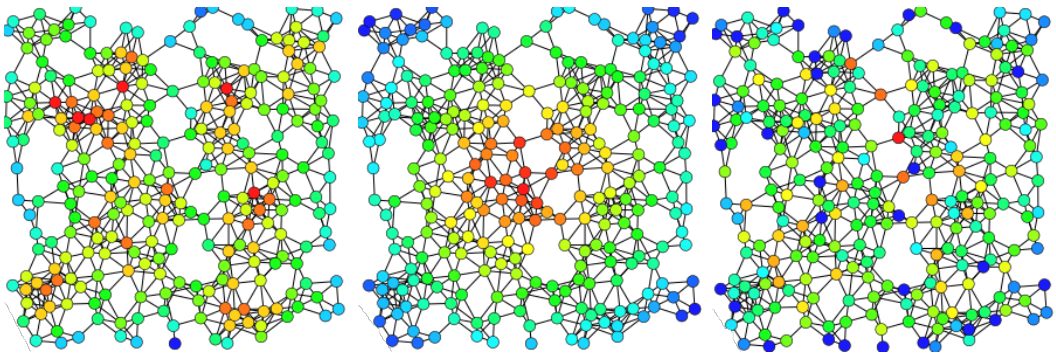
Centralidade – grau, *closeness*, *betweenness*



Centralidade – grau, *closeness*, *betweenness*



Centralidade – grau, *closeness*, *betweenness*



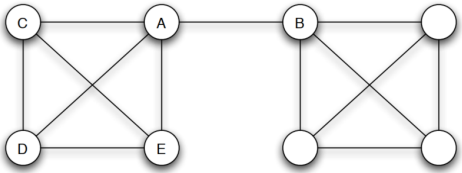
Resumo

Outros indicadores de centralidade

Há ideias e conceitos, alguns **já vistos em aula**, que podem ser considerados indicadores de centralidade:

- **Ponte** (aresta-de-corte),
- **Ponte local** (aresta-de-corte local),
- **Ponto de articulação** (vértice de corte).

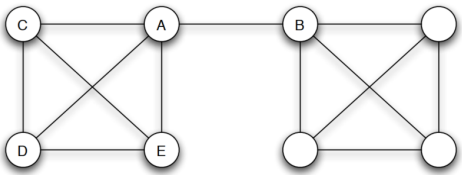
Outros indicadores de centralidade



Definição

A aresta e é uma aresta-de-corte (ou ponte) do grafo G, se $c(G - e) = c(G) + 1$.

Outros indicadores de centralidade

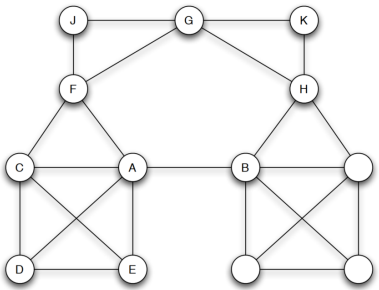


Definição

A aresta e é uma aresta-de-corte (ou ponte) do grafo G, se $c(G - e) = c(G) + 1$.

Se uma aresta é **ponte**, ela pode ser considerada **importante**: alguns caminhos mínimos com certeza passarão pela aresta (*betweenness* da aresta > 0). Quantos?

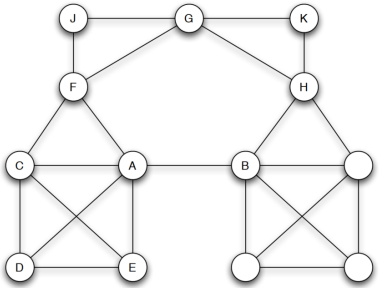
Outros indicadores de centralidade



Definição

A aresta $(a, b) \in E$ é uma ponte local do grafo G, se $dist_{G-(a,b)}(a, b) > 2$.

Outros indicadores de centralidade

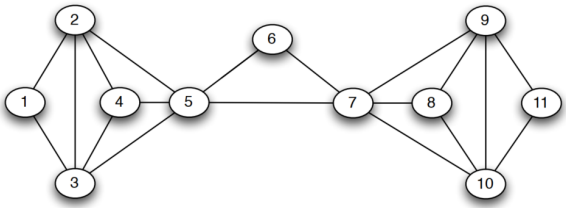


Definição

A aresta $(a, b) \in E$ é uma ponte local do grafo G, se $dist_{G-(a,b)}(a, b) > 2$.

Se uma aresta é **ponte local**, ela pode ser considerada **importante**: alguns caminhos mínimos com certeza passarão pela aresta (*betweenness* da aresta > 0). Quantos?

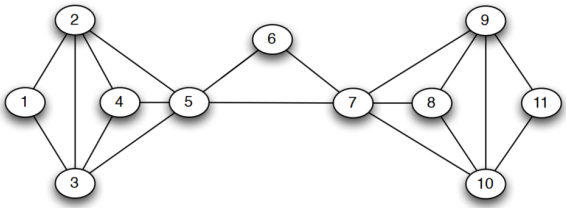
Outros indicadores de centralidade



Definição

Um vértice v é um ponto de articulação (ou vértice de corte) do grafo G, se $c(G - v) > c(G)$.

Outros indicadores de centralidade



Definição

Um vértice v é um ponto de articulação (ou vértice de corte) do grafo G, se $c(G - v) > c(G)$.

Se um vértice é **ponto de articulação**, ele pode ser considerado **importante**: alguns caminhos mínimos passarão por ele (*betweenness* do vértice > 0). Quantos?

Outros indicadores de centralidade

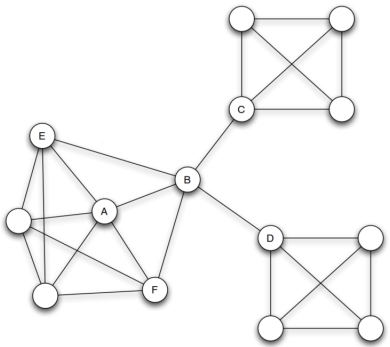
Relação entre ponte e ponto de articulação:

Suponha que sabemos que a aresta (v, w) é uma ponte no grafo G.
O grafo possui pontos de articulação?

Outros indicadores de centralidade

Relação entre ponte e ponto de articulação:

Suponha que sabemos que a aresta (v, w) é uma ponte no grafo G.
O grafo possui pontos de articulação? – **Sim** (desde que haja pelo menos 3 vértices).



Outros indicadores de centralidade

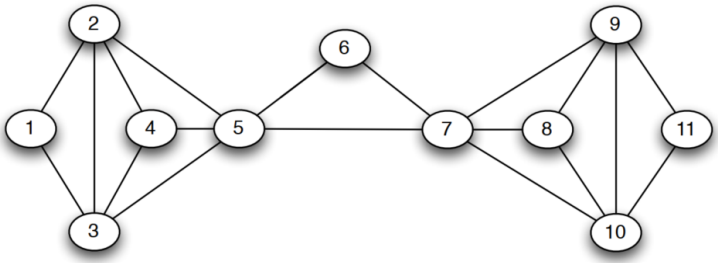
Relação entre ponte e ponto de articulação:

Suponha que sabemos que o vértice v é um ponto de articulação no grafo G.
O grafo possui pontes ou pontes locais?

Outros indicadores de centralidade

Relação entre ponte e ponto de articulação:

Suponha que sabemos que o vértice v é um ponto de articulação no grafo G.
O grafo possui pontes ou pontes locais? – **Não necessariamente.**



Resumo

Aplicações

- **Identificação de proteínas que são vitais**.

  Proteínas com alta centralidade são mais essenciais do que outras. Em redes genéticas, o gene p53, supressor de tumores, possui alta centralidade segundo vários critérios. Problemas no p53 tem consequências severas, levando a dificuldades em manter as células sob controle.

  H. Jeong et al.: "Lethality and centrality in protein networks". Nature (2001).

Aplicações

- **Vértices centrais das redes neurais**

  Desempenham um papel fundamental no funcionamento do **cérebro**. O cérebro humano possui um núcleo neocortical de hubs altamente conectados. Esses hubs garantem a robustez da rede, e representam estruturas relacionadas ao processo de cognição.

  S. Achard et al.: "A resilient, low-frequency, small-world human brain functional network with highly connected cortical hubs." Journal of Neuroscience (2006).

- **Detecção e diagnóstico da esquizofrenia**

  Das 54 métricas extraídas das redes corticais (um tipo de redes neurais), 4 permitem diferenciar pessoas com esquizofrenia de indivíduos saudáveis, sendo 3 delas relacionadas à centralidade.

  De Arruda et al.: "Structure and dynamics of functional networks in child-onset schizophrenia". Clinical Neurophysiology (2014).

Material bibliográfico

F. A. Rodrigues: "Network centrality: an introduction" (2019).

M. Newman & M. Girvan: "Finding and evaluating community structure in networks" (2004).

M. Girvan & M. Newman: "Community structure in social and biological networks" (2002).

Dúvidas

Dúvidas?