

# DATA MINING & MACHINE LEARNING (I)

Thiago Marzagão



**Centro Universitário**

# análise de grafos



- ▶ Exemplos de grafos:

# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)

# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)

# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)

# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)
- ▶ ... licitações (empresa A é co-licitante da empresa B, etc)

# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)
- ▶ ... licitações (empresa A é co-licitante da empresa B, etc)
- ▶ ... sociedades (fulano é sócio de beltrano, etc)



# análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)
- ▶ ... licitações (empresa A é co-licitante da empresa B, etc)
- ▶ ... sociedades (fulano é sócio de beltrano, etc)
- ▶ ... filmes (ator A contracenou com ator B, etc)

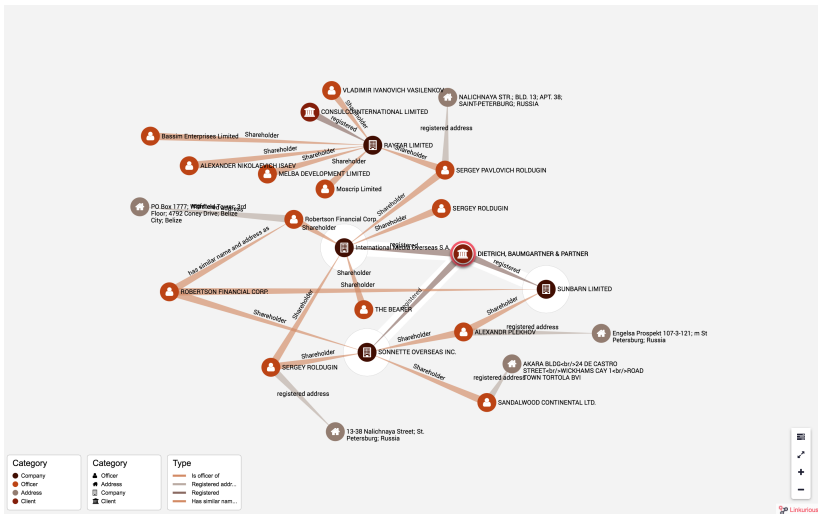
## análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)
- ▶ ... licitações (empresa A é co-licitante da empresa B, etc)
- ▶ ... sociedades (fulano é sócio de beltrano, etc)
- ▶ ... filmes (ator A contracenou com ator B, etc)
- ▶ ... relacionamentos diversos (cf. Bearman, Moody e Stovel 2004)

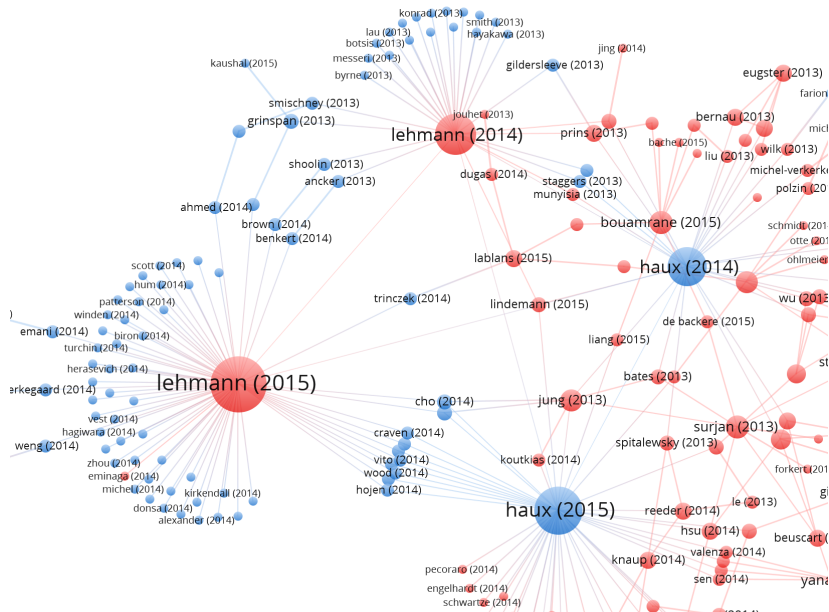
## análise de grafos

- ▶ Exemplos de grafos:
- ▶ ... redes sociais (fulano é amigo de beltrano, etc)
- ▶ ... crime organizado (fulano envia dinheiro p/ beltrano, etc)
- ▶ ... processos judiciais (processo A menciona processo B, etc)
- ▶ ... licitações (empresa A é co-licitante da empresa B, etc)
- ▶ ... sociedades (fulano é sócio de beltrano, etc)
- ▶ ... filmes (ator A contracenou com ator B, etc)
- ▶ ... relacionamentos diversos (cf. Bearman, Moody e Stovel 2004)
- ▶ ... Panamá Papers

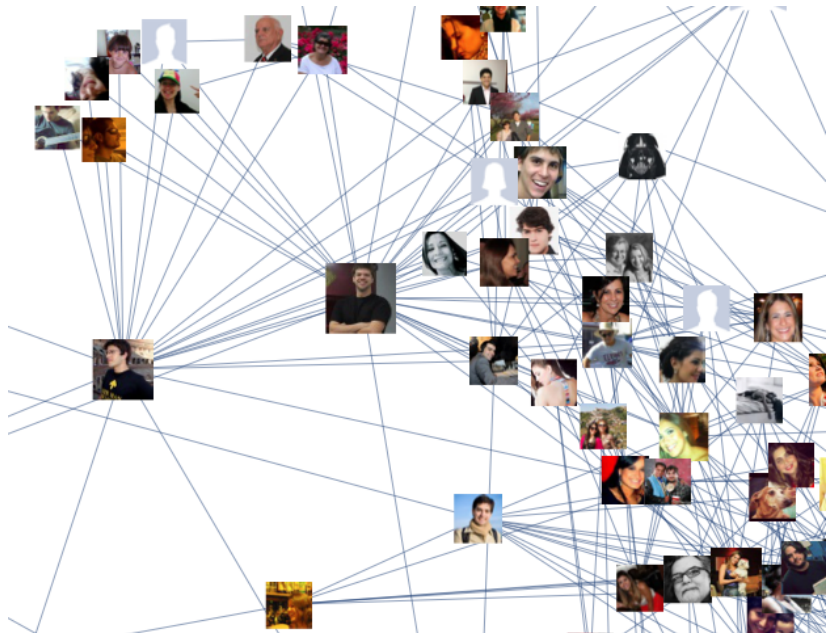
# Panamá Papers



# citações acadêmicas



# redes sociais





- ▶ Nós. São os “pontos” da rede. Pessoas, empresas, artigos acadêmicos, etc.



- ▶ Nós. São os “pontos” da rede. Pessoas, empresas, artigos acadêmicos, etc.
- ▶ (Às vezes são chamados de vértices.)

- ▶ Nós. São os “pontos” da rede. Pessoas, empresas, artigos acadêmicos, etc.
- ▶ (Às vezes são chamados de vértices.)
- ▶ Arestas. São as “ligações” entre os nós.

## nós e arestas

- ▶ Nós. São os “pontos” da rede. Pessoas, empresas, artigos acadêmicos, etc.
- ▶ (Às vezes são chamados de vértices.)
- ▶ Arestas. São as “ligações” entre os nós.
- ▶ Arestas podem ser direcionadas (ex.: Twitter) ou não-direcionadas (ex.: Facebook).

- ▶ Cada nó tem um grau.

## nós e arestas

- ▶ Cada nó tem um grau.
- ▶ Grau = qtde. de arestas do nó.

## nós e arestas

- ▶ Cada nó tem um grau.
- ▶ Grau = qtde. de arestas do nó.
- ▶ (Às vezes chamado de valência.)

## nós e arestas

- ▶ Cada nó tem um grau.
- ▶ Grau = qtde. de arestas do nó.
- ▶ (Às vezes chamado de valência.)
- ▶ Se o grafo é direcionado cada nó tem dois graus: o de saída e o de entrada.

- ▶ Caminho mínimo: é o caminho mais curto entre dois nós.



## como identificar os nós mais importantes?

- ▶ Métricas de centralidade:

## como identificar os nós mais importantes?

- ▶ Métricas de centralidade:
- ▶ centralidade de grau: é simplesmente o grau do nó

## como identificar os nós mais importantes?

- ▶ Métricas de centralidade:
- ▶ centralidade de grau: é simplesmente o grau do nó
- ▶ centralidade de proximidade:  $1 / (\text{soma dos caminhos mínimos até todos os outros nós})$

## como identificar os nós mais importantes?

- ▶ Métricas de centralidade:
- ▶ centralidade de grau: é simplesmente o grau do nó
- ▶ centralidade de proximidade:  $1 /$  (soma dos caminhos mínimos até todos os outros nós)
- ▶ centralidade de intermediação: qtde. de caminhos mínimos que passam pelo nó (às vezes dividida pelo total de caminhos mínimos existentes no grafo)

## como identificar os nós mais importantes?

- ▶ Métricas de centralidade:
- ▶ centralidade de grau: é simplesmente o grau do nó
- ▶ centralidade de proximidade:  $1 /$  (soma dos caminhos mínimos até todos os outros nós)
- ▶ centralidade de intermediação: qtd. de caminhos mínimos que passam pelo nó (às vezes dividida pelo total de caminhos mínimos existentes no grafo)
- ▶ ... diversas outras (exemplo: PageRank, que se baseia na centralidade dos nós adjacentes)

## coeficiente de clusterização

- ▶ Atenção! Aqui usamos o termo “clusterização num sentido totalmente distinto daquele que vimos nas aulas 2-3.

## coeficiente de clusterização

- ▶ Atenção! Aqui usamos o termo “clusterização num sentido totalmente distinto daquele que vimos nas aulas 2-3.
- ▶ Os nós adjacentes ao nó  $x$  formam uma “comunidade”?

## coeficiente de clusterização

- ▶ Atenção! Aqui usamos o termo “clusterização num sentido totalmente distinto daquele que vimos nas aulas 2-3.
- ▶ Os nós adjacentes ao nó  $x$  formam uma “comunidade”?
- ▶ Coeficiente de clusterização: qtde. de arestas entre os vizinhos imediatos de  $x$  / qtde. de arestas que *poderiam existir* entre os vizinhos imediatos de  $x$ .



## coeficiente de clusterização

- ▶ Atenção! Aqui usamos o termo “clusterização num sentido totalmente distinto daquele que vimos nas aulas 2-3.
- ▶ Os nós adjacentes ao nó  $x$  formam uma “comunidade”?
- ▶ Coeficiente de clusterização: qtde. de arestas entre os vizinhos imediatos de  $x$  / qtde. de arestas que *poderiam existir* entre os vizinhos imediatos de  $x$ .
- ▶ Em outras palavras: o quão “conectados” entre si são os vizinhos de  $x$ ?

## atributos do grafo

- ▶ Qtde. de nós.

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*.

$$\frac{N(N - 1)}{2}$$

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?
- ▶ Diâmetro: maior caminho mínimo.

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?
- ▶ Diâmetro: maior caminho mínimo.
- ▶ Grau médio

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?
- ▶ Diâmetro: maior caminho mínimo.
- ▶ Grau médio
  - ▶ grafo não-direcionado:  $(2L)/N$



## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?
- ▶ Diâmetro: maior caminho mínimo.
- ▶ Grau médio
  - ▶ grafo não-direcionado:  $(2L)/N$
  - ▶ grafo direcionado:  $L/N$

## atributos do grafo

- ▶ Qtde. de nós.
- ▶ Qtde. de arestas.
- ▶ Qtde. de arestas / qtde. de arestas que *poderiam existir*:  
$$\frac{N(N - 1)}{2}$$
- ▶ Direcionado ou não-direcionado?
- ▶ Diâmetro: maior caminho mínimo.
- ▶ Grau médio
  - ▶ grafo não-direcionado:  $(2L)/N$
  - ▶ grafo direcionado:  $L/N$
- ▶ Coeficiente de clusterização médio.

## o que não vamos ver

- ▶ Grafos aleatórios.

## o que não vamos ver

- ▶ Grafos aleatórios.
- ▶ Algoritmos de caminho mínimo.

## o que não vamos ver

- ▶ Grafos aleatórios.
- ▶ Algoritmos de caminho mínimo.
- ▶ ... muito mais!

# exercício - Panamá Papers

