

# Aula 1 - Introdução à análise de redes sociais

José Devezas <[jld@fe.up.pt](mailto:jld@fe.up.pt)>

Faculdade de Engenharia da Universidade do Porto

Gestão da Informação em Redes Sociais

Mestrado em Ciência da Informação



# Conteúdos

- Introdução
- Perspetiva histórica
- Conceitos base
- Caraterísticas das redes sociais
- Métricas de centralidade
- Algoritmos clássicos
- Estrutura de comunidade
- Plano de avaliação



# Introdução

- Professores:
  - José Devezas
  - Jorge Sobrado
- Nesta componente (7 aulas), iremos aprender sobre a análise das redes sociais, do ponto de vista:
  - Da estrutura
    - Identificar pessoas (nós) relevantes na estrutura social (rede)
    - Identificar grupos coesos de pessoas (comunidades)
  - Do conteúdo
    - Perceber quais os tópicos discutidos na rede
  - Da estrutura e do conteúdo (abordagem combinada)
    - Perceber quais os tópicos discutidos na rede por diferentes pessoas e grupos de pessoas



# Introdução

- Aplicações da análise de redes (sociais e não só):
  - **Marketing e comunicação**
    - p.e. promoção e divulgação de produtos através de intervenções pontuais junto de indivíduos centrais e propensos à propagação da informação (“bem ligados”).
  - **Análise forense**
    - p.e. identificação de células terroristas com base em redes de comunicação.
  - **Análise de tráfego**
    - p.e. identificação de pontos com elevado potencial de congestionamento.
  - **Biologia**
    - p.e. caracterização de proteínas desconhecidas com base noutras proteínas conhecidas e que fazem parte do mesmo “grupo de interações”.

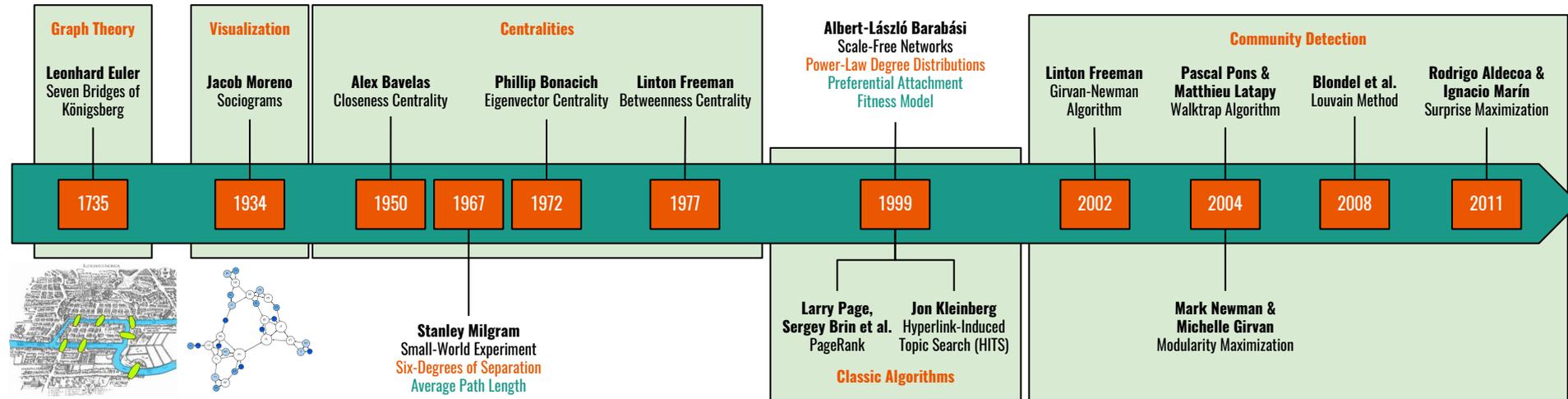


# Introdução

- Principais objetivos:
  - Identificação de indivíduos com elevado potencial na disseminação da informação pela rede.
  - Segmentação de mercado com base nas relações sociais.
  - Familiaridade com as ferramentas e formatos de dados suportados.
  - Capacidade de caracterizar uma rede social, obtendo informação de apoio à decisão na criação de planos de comunicação e gestão da informação.

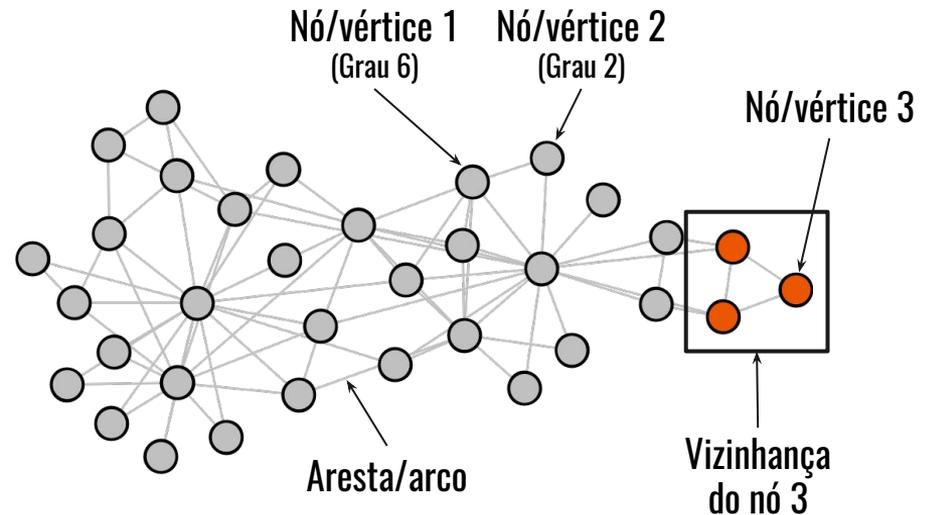
# Perspetiva histórica

- Contribuições relevantes da antropologia, sociologia, psicologia, matemática e física.



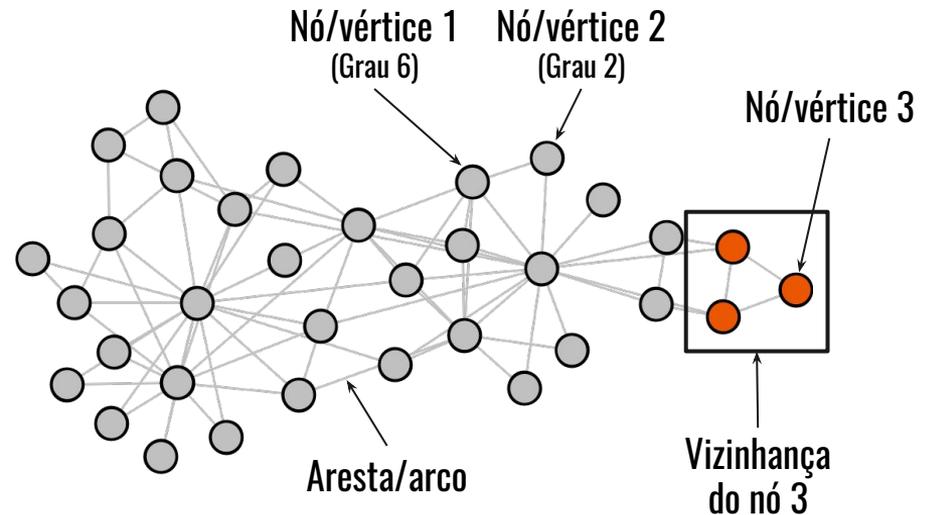
# Conceitos base

- **Rede ou grafo**  $\Rightarrow$  conjunto de nós/vértices e arestas/arcos.
  - Grafo dirigido  $\Rightarrow$  ligação entre os nós  $\mathcal{A}$  e  $\mathcal{B}$  é diferente de ligação entre os nós  $\mathcal{B}$  e  $\mathcal{A}$ .
  - Grafo não-dirigido  $\Rightarrow$  ligação entre os nós  $\mathcal{A}$  e  $\mathcal{B}$  é igual à ligação entre os nós  $\mathcal{B}$  e  $\mathcal{A}$ .
- **Nó/vértice**  $\Rightarrow$  geralmente uma pessoa na rede social.
- **Aresta/arco**  $\Rightarrow$  geralmente uma relação de amizade (ou interação) entre duas pessoas na rede social.
- **Vizinhança**  $\Rightarrow$  todos os nós ligados por uma aresta a um determinado nó (distância um).



# Conceitos base

- **Grau do nó**  $\Rightarrow$  número de ligações incidentes num nó (grafo não-dirigido).
  - Grau de entrada e grau de saída  $\Rightarrow$  o mesmo num grafo dirigido (número de ligações de entrada num nó e número de ligações de saída do nó).
  - O grau do nó pode ser calculado para um grafo dirigido, tomando o grafo como não-dirigido.
- **Distribuição do grau do nó**  $\Rightarrow$  contagem (frequência absoluta ou relativa) do número de nós (eixo dos  $\mathcal{N}$ ) com um determinado grau (eixo dos  $\mathcal{X}$ ).





# Caraterísticas das redes sociais

## Small-world experiment

- Experiência levada a cabo no EUA, pelo psicólogo social, Stanley Milgram.
- **Objetivo:** encontrar cadeias curtas de indivíduos conhecidos entre si, de forma a estabelecer uma ligação entre dois outros indivíduos predeterminados.
- **Método:** envio de cartas entre pessoas conhecidas, partindo de um indivíduo  $\mathcal{A}$ , com o objetivo de ligar a um indivíduo  $\mathcal{B}$ .
- **Resultado:** seis graus de separação (quatro no Facebook!).
- Em média, o caminho mais curtos entre quaisquer dois nós na rede era entre cinco e seis.



# Caraterísticas das redes sociais

## Small-world experiment

«The goal of the experiment was to **find short chains of acquaintances linking pairs of people** in the United States who did not know one another. In a typical instance of the experiment, a source person in Nebraska would be given a letter to deliver to a target person in Massachusetts. The source would initially be told basic information about the target, including his address and occupation; the source would then be instructed to send the letter to someone she knew on a first-name basis in an effort to transmit the letter to the target as efficaciously as possible. Anyone subsequently receiving the letter would be given the same instructions, and the chain of communication would continue until the target was reached. **Over many trials, the average number of intermediate steps in a successful chain was found to lie between five and six**, a quantity that has since entered popular culture as the “**six degrees of separation**” principle.»

Kleinberg, J. (2000, May). The Small-World Phenomenon: An Algorithmic Perspective. In Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (pp. 163-170). ACM.

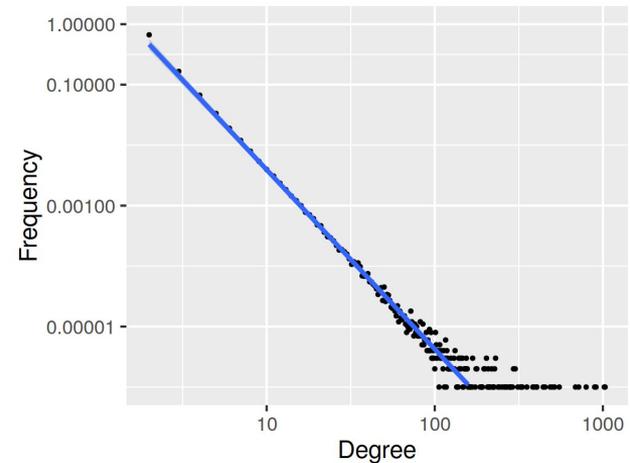
# Caraterísticas das redes sociais

## Scale-free networks

- Independentemente do número de nós da rede (*scale-free*), numa *scale-free network* existe uma grande fração de nós com um grau baixo e uma pequena fração com um grau elevado.
- A distribuição do grau segue uma *power law*:

$$P(k) \sim k^{-\gamma}$$

- A figura ilustra a distribuição do grau para um grafo gerado com base em *preferential attachment* (escala log-log).





# Métricas de centralidade

## (Normalized) degree centrality

- *Degree centrality*  $\Leftrightarrow$  grau do nó.
- A normalização é feita dividindo pelo máximo valor que o grau pode tomar:  $\mathcal{N}-1$ , em que  $\mathcal{N}$  é igual ao número de nós no grafo.
  - Assume-se que não podem existir *loops* (ligações para o mesmo nó, ou seja, de  $\mathcal{A}$  para  $\mathcal{A}$ ) ou múltiplas arestas (mais do que uma aresta entre o mesmo par de nós).
  - Um nó com 100% normalized degree centrality está ligado a todos os outros nós.
- Na rede social, mede a importância de um indivíduo quanto ao números de amigos ou interações.



# Métricas de centralidade

## Closeness centrality

- Calculado com base na distância de um nó a todos os outros nós, pelos caminhos mais curtos.
- Quanto menor a distância de um nó a todos os outros, maior a sua *closeness centrality*.
- *Closeness centrality* é o inverso da soma das distâncias.
- Na rede social, mede a importância de um indivíduo quanto à proximidade aos restantes indivíduos.



# Métricas de centralidade

## Eigenvector centrality

- Calculado recursivamente com base na centralidade relativa dos vizinhos.
- Isto significa que a *eigenvector centrality* é tanto maior quanto maior a centralidade dos vizinhos.
- Na rede social, mede a importância de um indivíduo quanto à importância dos seus amigos.
- **Intuição:** pessoas ligadas a pessoas importantes são mais importantes.



# Métricas de centralidade

## Betweenness centrality

- A *betweenness centrality* de um dado nó é calculada com base na fração de caminhos mais curtos que passam pelo nó, face ao número total de caminhos mais curtos (não incluindo o próprio nó).
- Isto significa que a *betweenness centrality* é tanto maior quanto mais essencial for às travessias entre pares de nós.
- Ou seja, quando removemos um nó com elevada *betweenness centrality*, é provável que as distâncias mínimas entre dois nós aumentem.
- **Na rede social, mede a importância de um indivíduo quanto ao seu papel como ponte.**
- **Intuição:** nós com elevada *betweenness centrality* geralmente fazem a ligação entre módulos pouco ligados externamente entre si (comunidades).



# Algoritmos clássicos

## PageRank

- Variante da *eigenvector centrality* — a intuição é a mesma.
- No entanto, o PageRank assume um *random surfer* (i.e., alguém a navegar na web — outra rede real com características similares às das redes sociais).
- O *random surfer* tem uma probabilidade (geralmente 85%) de seguir uma ligação do nó atual (página web ativa) e uma probabilidade (geralmente 15%) de saltar aleatoriamente para outro nó.
- PageRank é a probabilidade de um *random surfer* chegar a uma página web, clicando numa ligação.
- Na rede social, mede a importância de um indivíduo quanto à importância dos seus amigos, assumindo a probabilidade de um utilizador da rede social chegar ao indivíduo através da navegação pela lista de amigos (ou seguidores).



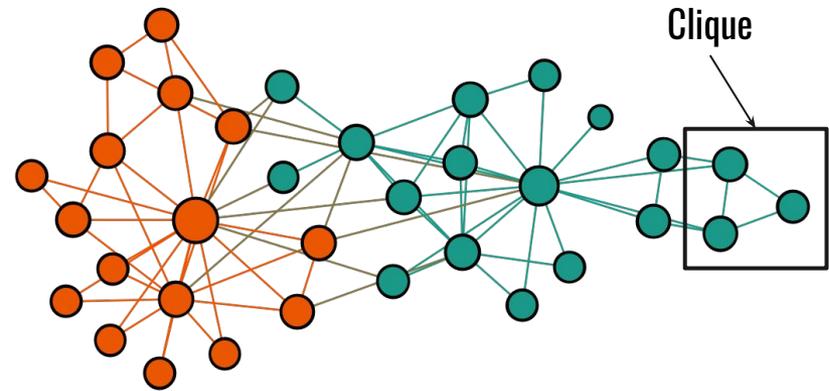
# Algoritmos clássicos

## Hyperlink-Induced Topic Search (HITS)

- Permite identificar *hubs* e autoridades, calculando duas pontuações por nó.
- Pontuação de *hub* é calculada com base na pontuação de autoridade e vice-versa (algoritmo recursivo).
- Num grafo não-dirigido, as pontuações de *hub* e autoridade são iguais.
  - E sensivelmente equivalentes às da *eigenvector centrality*.
- Isto acontece porque o algoritmo foi desenvolvido para a web, onde as ligações são dirigidas.
- Na rede social, em particular quando as ligações são assimétricas (p.e., *Twitter mention networks*), mede a importância de um indivíduo relativamente ao número de interações que gera (*hub*) ou ao número de interações que recebe (autoridade).

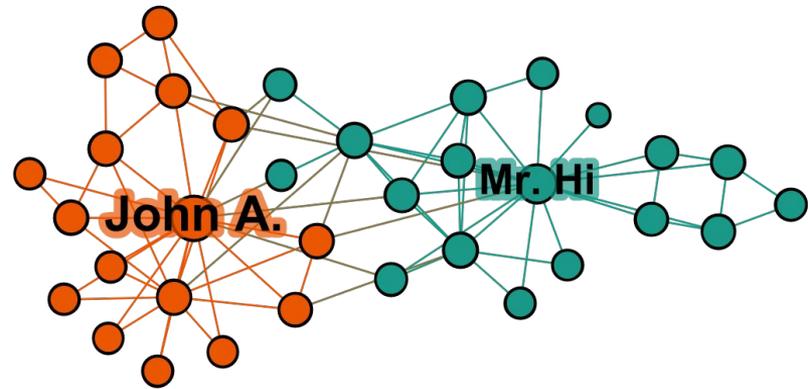
# Estrutura de comunidade

- Definição de comunidade num grafo social:
  - Comunidade é um subgrafo denso, isto é, um subgrafo em que existem mais ligações entre os seus nós do que com o resto da rede.
- Exemplo:
  - Um grupo de amigos no Facebook em que todos são amigos de todos (ou de quase todos).
    - Caso especial: **clique** (todos os nós estão ligados a todos os outros nós; i.e., todos são amigos de todos).



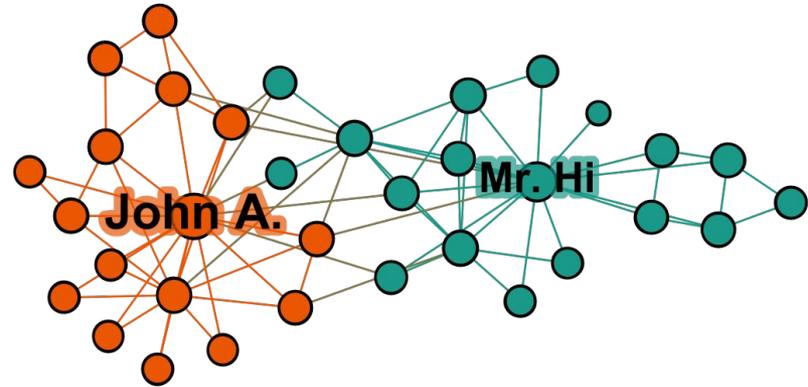
# Estrutura de comunidade

- *Zachary's Karate Club.*
- Laços de amizade entre os membros de um clube de karatê.
- Conflito sobre preços das lições entre o presidente do clube, **John A.**, e o instrutor de karatê, **Mr. Hi.**
- Fissão dos laços sociais — criaram-se dois grandes grupos (comunidades).
- PageRank (tamanho dos nós) é maior para o **John A.** e para o **Mr. Hi.**



# Estrutura de comunidade

- Detecção de comunidades:
  - *Louvain method*
  - *Walktrap algorithm*
  - *Label propagation*
  - *Surprise maximization*
- Vamos experimentar o *Louvain method* no Gephi, uma das ferramentas de análise de redes a explorar.





# Plano de avaliação

- Para a componente de análise de redes sociais:
  - **Dois trabalhos individuais** com entrega de um mini-relatório.
    - A desenvolver maioritariamente na aula (uma aula por trabalho).
  - **Um trabalho de grupo** com entrega de relatório e apresentação.
    - A desenvolver parcialmente na aula (duas aulas disponíveis).
  - Grupo de questões no **exame**.