



# **Introdução a redes complexas**

**Prof. Lilian Berton**

**São José dos Campos, 2018**

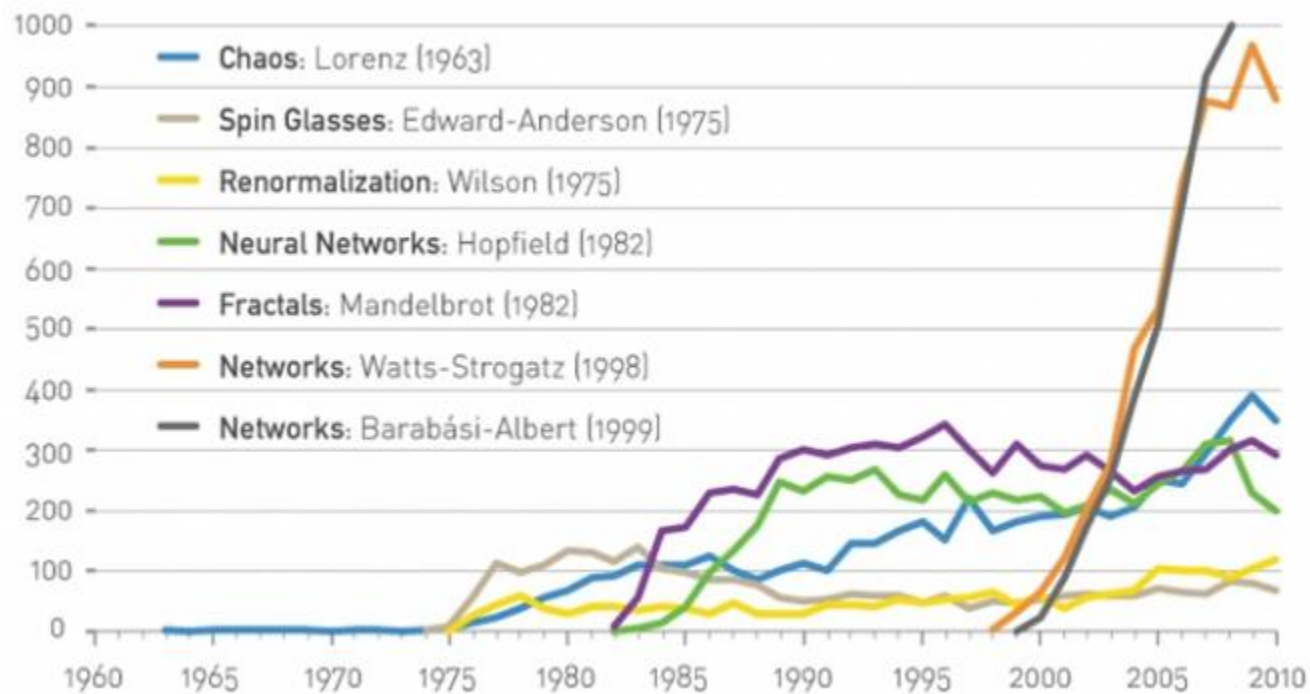
# Histórico

- Resolução do Problema das Sete **Pontes de Königsberg por Euler**, em 1736 é tido como o marco inicial da Teoria dos Grafos.
- Em 1926, Boruvka publica um algoritmo para encontrar a **árvore geradora mínima** em um grafo, almejando a construção de uma rede elétrica eficiente na República Checa.
- Em 1959, Erdos e Renyi criaram modelos para criação de **grafos aleatórios**.
- Em 1967, o psicólogo Stanley Milgram conduziu experimentos para demonstrar o “**Fenômeno do Mundo Pequeno**.”
- Em 1998, foi publicado na revista Nature o artigo “**Collective dynamics of small-world networks**”, de Watts e Strogatz, modificando o modelo de Erdos e Renyi.

# Histórico

- Em 1998, surge o Google, sistema de busca na internet que usa um algoritmo **PageRank**.
- 1999 - Albert-László Barabási e Réka Albert publicaram um artigo com a proposta de um modelo genérico de construção de redes, semelhante à estrutura encontrada em redes genéricas ou redes da Internet. Estas redes foram apelidadas de **Redes livres de escala**.
- Na década de 2000, surgem artigos na análise de **redes sociais aplicadas a redes de terroristas**. Em 2005, um comitê do governo americano considerou a Ciência de Redes uma área estratégica para a segurança nacional.
- Em 2004, surge **Orkut e Facebook**, gerando grandes conjuntos de dados para a análise de redes.

# Interesse em redes

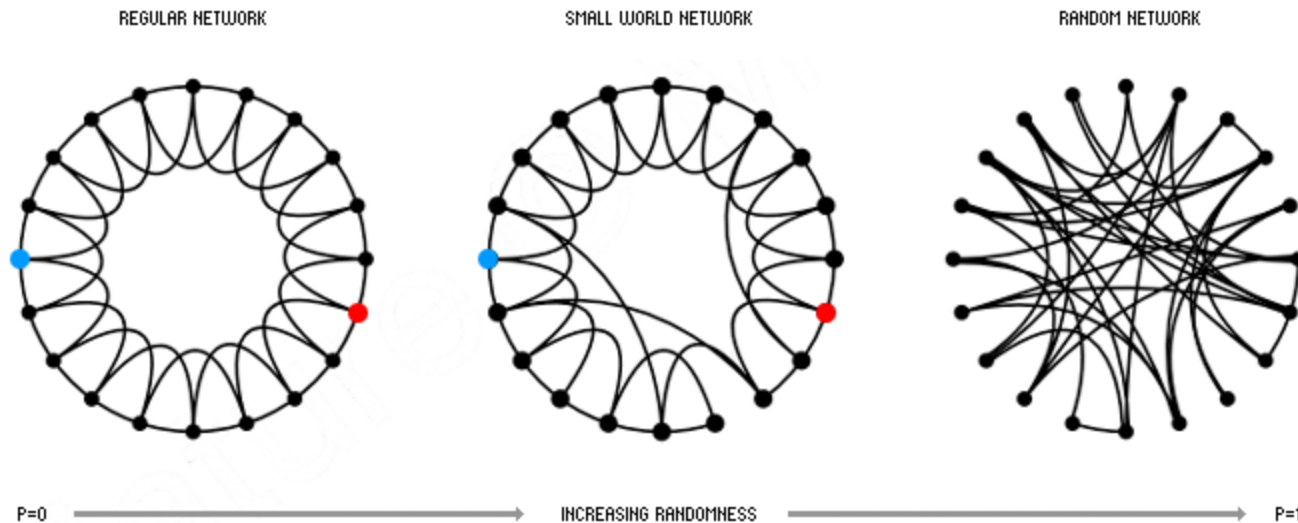


# Modelos de Redes

- **Grafos aleatórios:** inicia-se com  $N$  vértices desconectados, selecionam-se pares de vértices aleatoriamente, gerando uma conexão entre eles, até que haja  $K$  arestas. Esse grafo é denotado Grafo ER (Erdos Renyie) com  $N$  vértices e  $K$  arestas.
- Nesse modelo, dado um grafo com grau médio  $\langle k \rangle$ , para  $N$  muito grande, a geodésica média  $L$  tende a  $\log_{\langle k \rangle} N$ . O que confirma o fenômeno de mundo pequeno. Estimando a população mundial em  $N = 7 * 10^9$  e assumindo uma média de conexões por pessoa (família, amigos, conhecidos), de  $\langle k \rangle = 40$ , obtemos  $L = 6$ .
- Em outro modelo, inicia-se com  $N$  vértices desconectados e cada aresta tem uma probabilidade  $p$  de existir. A probabilidade desse grafo ter  $K$  arestas é:  
$$p^K (1 - p)^{\binom{N}{2} - K}$$
- Para  $N$  grande  $\langle k \rangle$  fixo a distribuição de grau é aproximadamente Poisson.

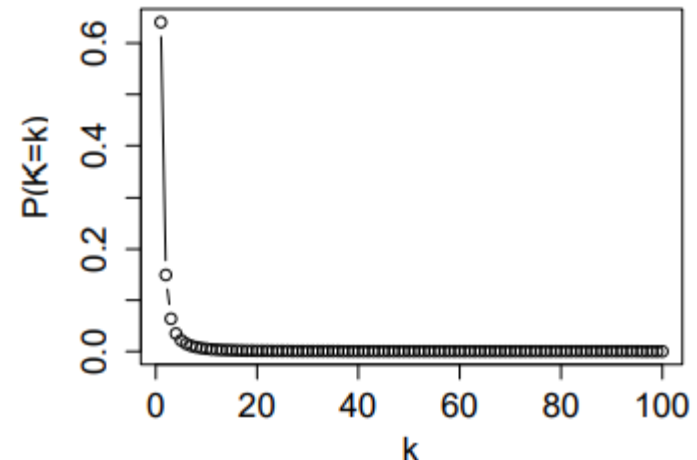
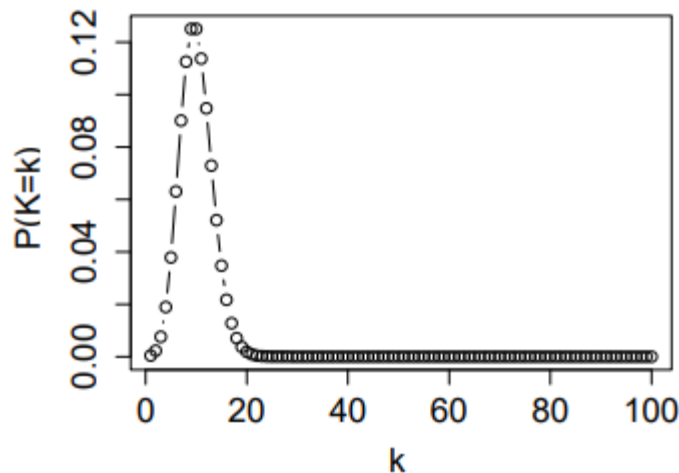
# Modelos de Redes

- **Redes mundo-pequeno:** cria-se um grafo regular e, para cada vértice, é escolhida uma aresta aleatória (que ainda não foi escolhida) e ela é apontada aleatoriamente pra outro vértice, ligando o vértice em questão a outro. Ao fazer isso, para valores baixos de  $p$ , obtém-se um modelo com agrupamento alto e geodésica baixo.



# Modelo de Redes

- **Redes com conexão preferencial:** ao estudar redes reais, descobriu-se que a maioria delas exibe uma distribuição de grau seguindo uma lei de potência:  $P(k) = \alpha k^{-\gamma}$ , com  $2 < \gamma < 3$
- Isso motivou a criação dos modelos de redes com conexão preferencial. A ideia é que quando se gera uma nova conexão, **o vértice não se conecta aleatoriamente, mas prefira se conectar a vértices com grau alto**. Ex: em uma rede de citação de artigos, é mais provável que um novo artigo cite artigos que já foram citados várias vezes, em vez de outros menos citados.



# Medidas de Centralidade

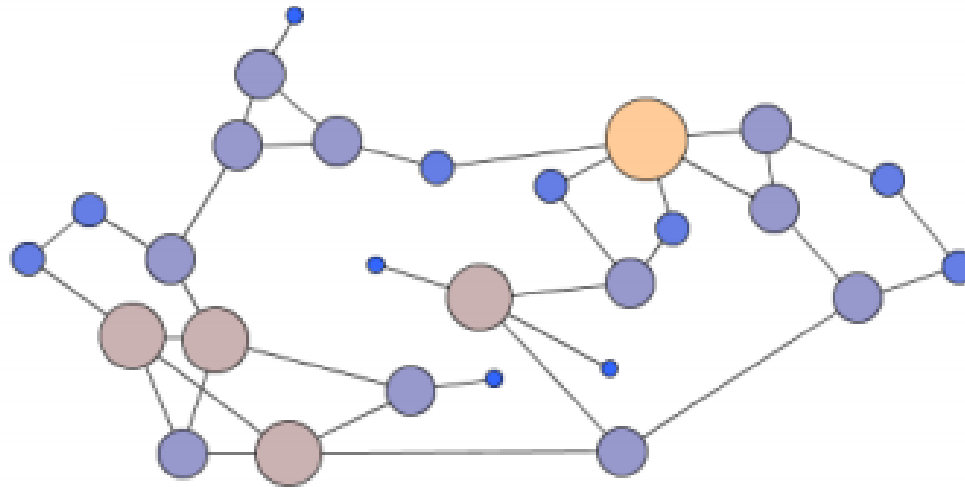
- Um aspecto importante na análise de redes é identificar **quais são os elementos mais importantes ou centrais de uma rede**. As medidas de centralidade são uma maneira de quantificar essa importância.
- Uma medida de centralidade é uma função  $f: V \rightarrow R$ , construída de modo que a relação de ordem entre  $f(v_i)$  e  $f(v_j)$  reflita na percepção de que  $v_i$  é mais central do que  $v_j$ , em algum sentido.



# Medidas de Centralidade

- **Centralidade de grau**, onde  $k_i$  é o grau do vértice  $v_i$ .

$$k_i = \sum_{j=1}^n a_{ij}$$

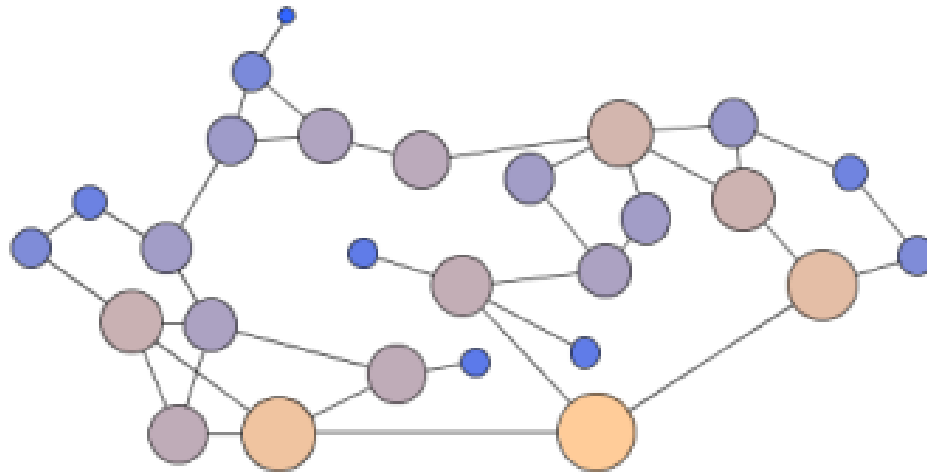


# Medidas de Centralidade

- **Centralidade de proximidade** : o recíproco da soma das distâncias de  $v$  aos demais vértices:

$$c_c(v) = \frac{1}{\sum_{a \in V} d(v, a)}.$$

onde  $d(v_i, a)$  é a distância entre os vértices  $v_i$  e  $a$ .

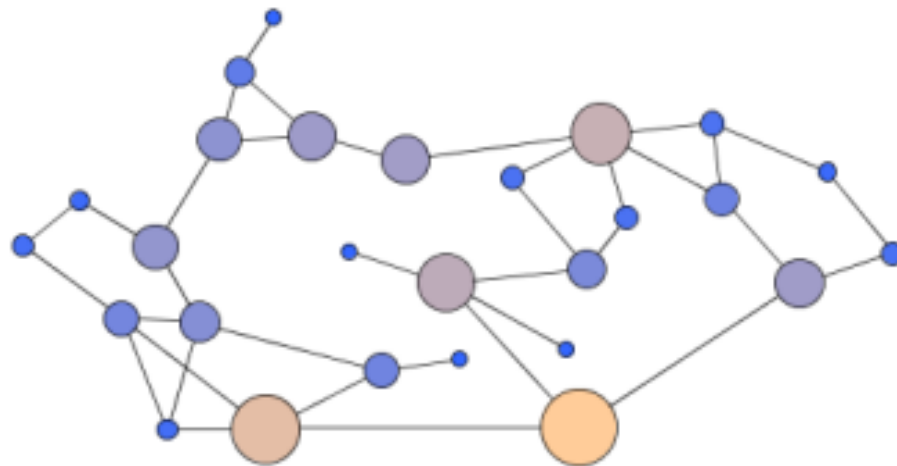


# Medidas de Centralidade

- **Centralidade de intermediação:** dada pela proporção das geodésicas (caminhos de tamanho mínimo) entre dois vértices quaisquer que passam pelo vértice em questão.

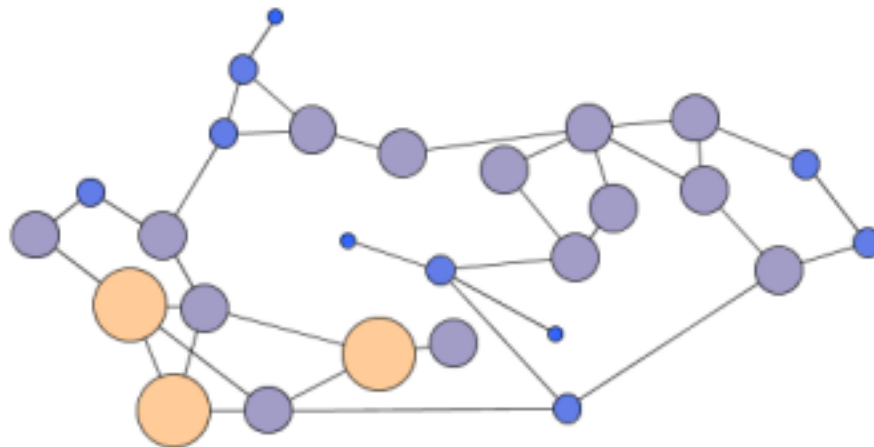
$$c_B(v) = \sum_{a,b \neq v} \frac{g_{avb}}{g_{ab}}.$$

onde  $g_{ab}$  é o número de geodésicas entre  $a$  e  $b$ , e  $g_{avb}$  é o número dessas que passam por  $v$ .



# Medidas de Centralidade

- Existem também **medidas espectrais**, sendo o maior autovalor em módulo da matriz de adjacência e o autovetor associado, a centralidade de autovetor do vértice  $v_i$  é definida como a  $i$ -ésima posição do seu autovetor.



# Teoria espectral dos grafos

- Trata da análise dos grafos a partir de suas matrizes associadas, com ênfase especial nos autovalores e autovetores dessas matrizes.

Uma matriz  $A = [a_{ij}]$  é dita *não-negativa* se  $a_{ij} \geq 0$  para qualquer  $i, j$ . Se  $A$  é quadrada de ordem  $n$ , é dita *irredutível* se não existe nenhuma decomposição da forma  $N_1 \cup N_2 = \{1, 2, \dots, n\}$ ,  $N_1 \neq \emptyset \neq N_2$ ,  $N_1 \cap N_2 = \emptyset$  tal que  $a_{ij} = 0$  para  $i \in N_1$  e  $j \in N_2$ .

**Teorema 2.1 (Perron-Frobenius):** Seja  $A_{n \times n}$  é não-negativa e irredutível, e  $\rho(A)$  o seu maior autovalor em módulo. Então,

- a)  $\rho(A) > 0$
- b) Existe um autovetor positivo  $\mathbf{v}_1$  associado a  $\rho(A)$ .
- c)  $\rho(A)$  é um autovalor algebricamente simples.

# Centralidade do autovetor

- Se considerarmos que a centralidade  $c_x(v_i)$  é proporcional (por um fator  $\alpha$ ) a soma das centralidades de seus vizinhos, temos que:

$$c_X(v_i) = \alpha \sum_{j=1}^n a_{ji} c_X(v_j) = \alpha A(:, i) \mathbf{c}_X$$

Em notação matricial,  $\mathbf{c}_X = \alpha A \mathbf{c}_X$ , ou seja,  $\alpha^{-1}$  é um autovalor de  $A$  e  $\mathbf{c}_X$  é o autovetor correspondente. Pelo Teorema de Perron-Frobenius, temos que um autovetor associado a  $\lambda_1$  pode ser tomado de modo a termos todas as entradas positivas. Bonacich [32] foi quem primeiro recomendou que o autovetor principal  $\mathbf{v}_1$  associado a  $\lambda_1$  fosse usado como uma medida de centralidade.

# PageRank

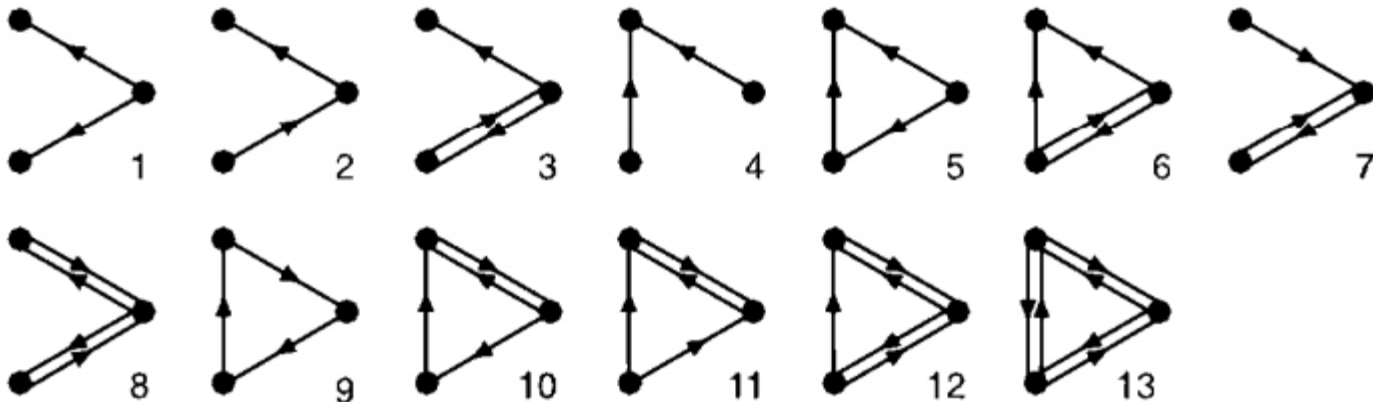
- Um problema da centralidade de autovetor é que se um vértice importante é vizinho de um grande número de vértices, todos eles ganham importância. Ex: se eu tenho uma página hospedada no Yahoo! Eu ganho importância em função disso, assim como milhões de outros sites. Uma página poderia inflar sua importância gerando muitos links.
- O algoritmo PageRank busca diluir a importância de um vértice entre seus links.

$$c_X(v_i) = \alpha \sum_{j=1}^n a_{ij} \frac{c_X(v_j)}{k_j^{out}} \quad \mathbf{c}_X = \alpha \mathbf{A} \mathbf{D}^{-1} \mathbf{c}_X ,$$

O PageRank é um autovetor da matriz  $\alpha \mathbf{A} \mathbf{D}^{-1}$ .

# Medidas de redes

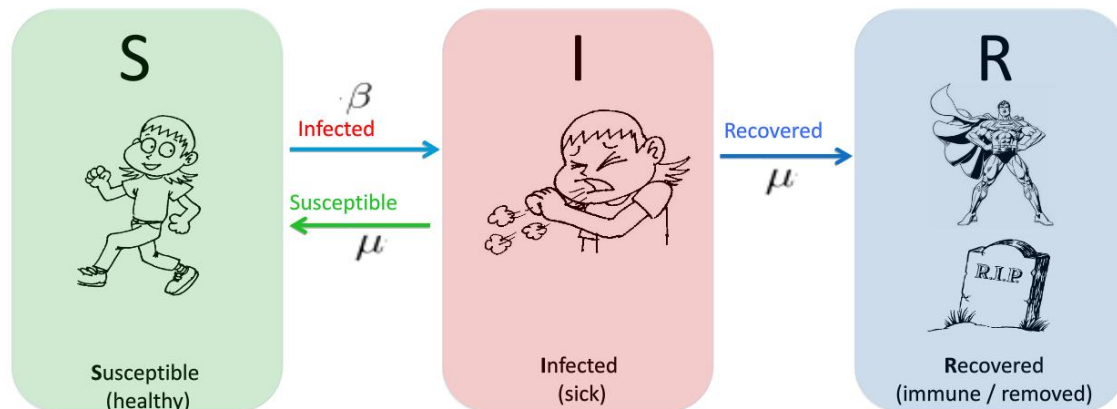
- Quando um subgrafo conexo aparece em um grafo com frequência significativamente maior que em um grafo aleatório equivalente, chamamos esse subgrafo de **motif**. Os possíveis *motifs* de  $n$  vértices são todos os possíveis subgrafos conexos de  $K_n$ . Para uma rede dirigida, há 13 possíveis *motifs* de 3 vértices.



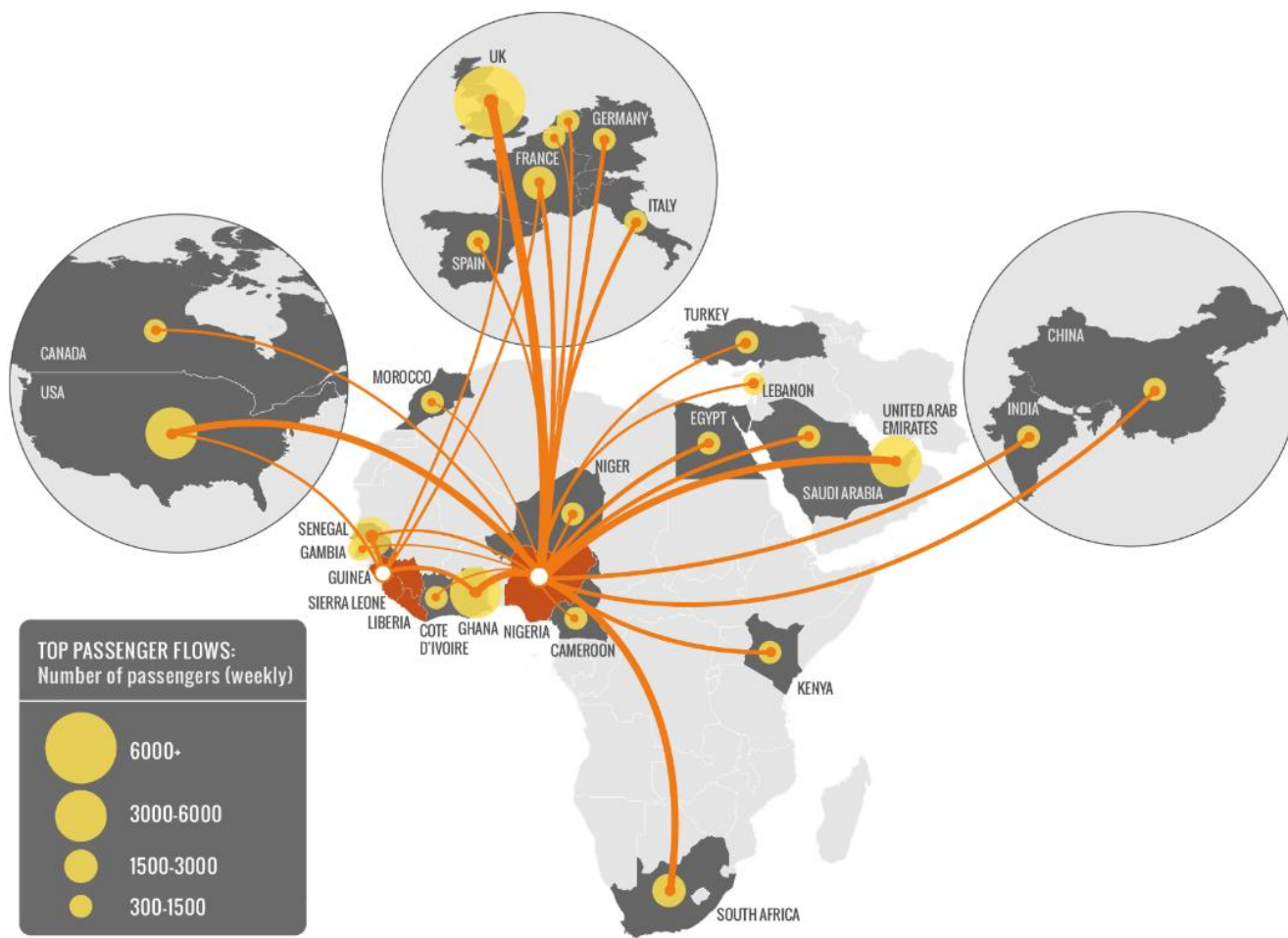


# Difusão de epidemias e rumores

- **Difusão:** pode-se estudar a evolução do contágio de uma doença através de um grafo onde os vértices são pessoas.
- No modelo de Bass, os elementos podem ter dois estados (0 ou 1) e  $F(t)$  representa a proporção de elementos no estado 1 no tempo  $t$ ,  $F' = p + qF/1 - F$ .
- Onde  $p$  é a taxa com que os elementos passam para o estado 1, e  $q$  é a taxa com os elementos “imitam” o estado de outro elemento no estado 1.
- Modelo **SIS (suscetível-infectado-suscetível)** e **SIR (suscetível-infectado-recuperado)**.



# Difusão Ebola 2014



# Difusão de rumores

BBC

Entrar

Menu

BRASIL

Notícias | Brasil | Internacional | Economia | Saúde | Ciência | Tecnologia | Aprenda Inglês

## Exclusivo: investigação revela exército de perfis falsos usados para influenciar eleições no Brasil

Juliana Gagnani  
Da BBC Brasil em Londres

8 dezembro 2017



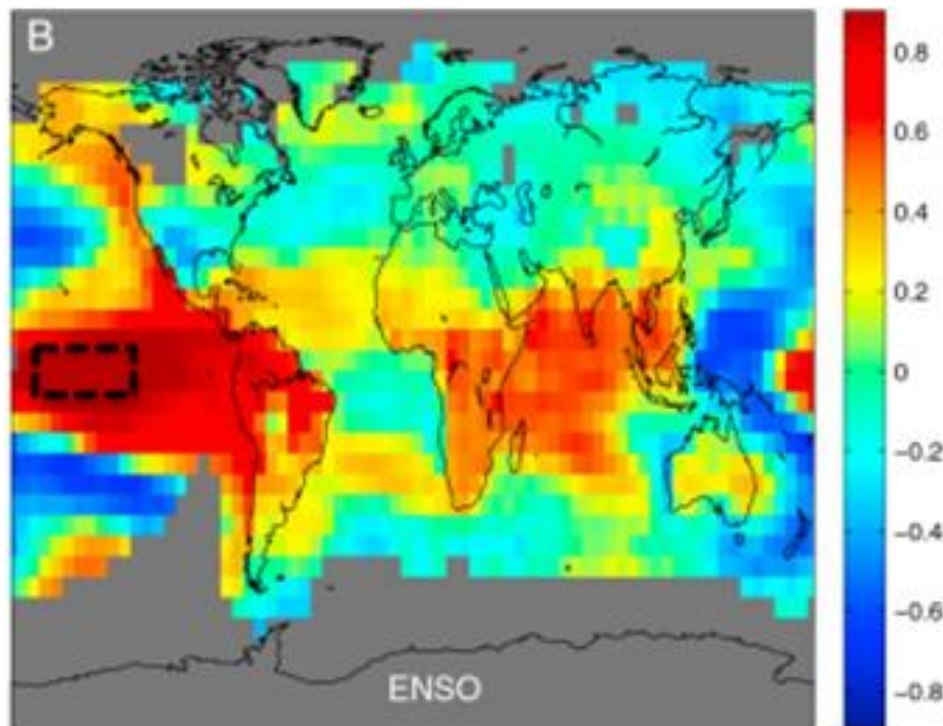
Compartilhar

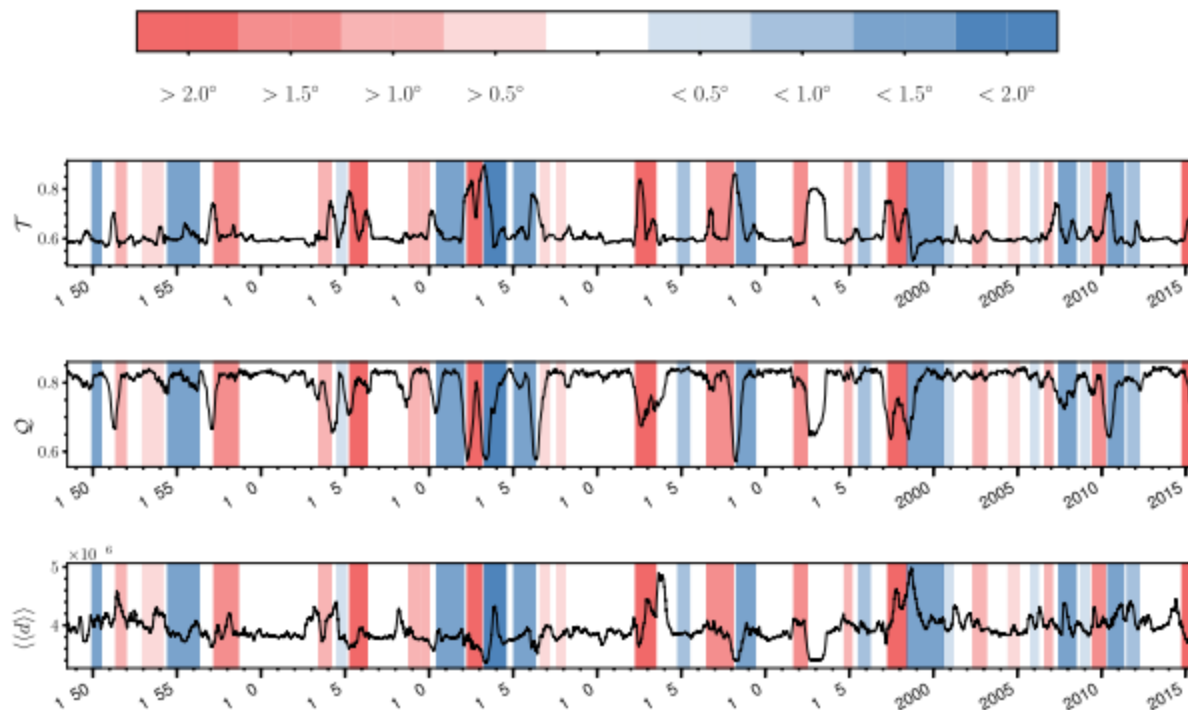


Cada funcionário seria responsável por controlar de 20 a 50 perfis falsos | Ilustração: Kako Abraham/BBC

# Análise de dados climáticos

- Cada subsistema do sistema climático, como atmosfera, oceano, biosfera, litosfera, etc, pode gerar uma resposta no tempo devido a perturbações nesse subsistema e levar algum tempo para voltar ao equilíbrio. Podemos estudar essas alterações por meio de variáveis coletadas por sensores distribuídos no globo terrestre.



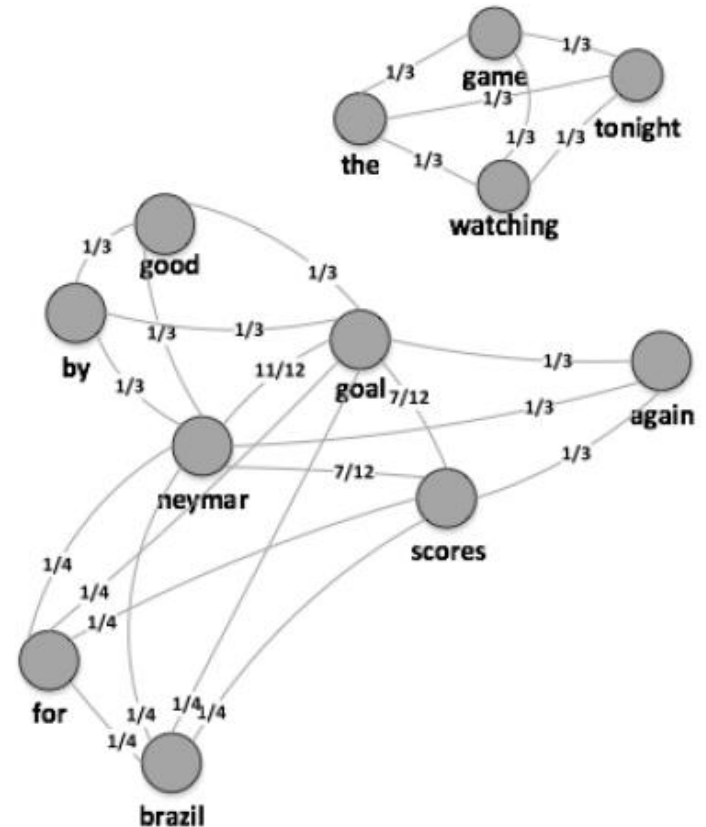


**Figure 2.** Time series of climate network (a) transitivity  $T$ , (b) modularity  $Q$  and (c) global average link distance  $\langle\langle d \rangle\rangle$ . Background colors highlight different ENSO phases (red: El Niño (EN), blue: La Niña (LN)) according to the Ocean Niño Index (ONI), with opacity representing to the corresponding index value. Ticks on the time axis indicate the 1st of January of a given year; all values are shown according to the midpoint dates of the respective time windows.

# Análise de Twitter usando grafos

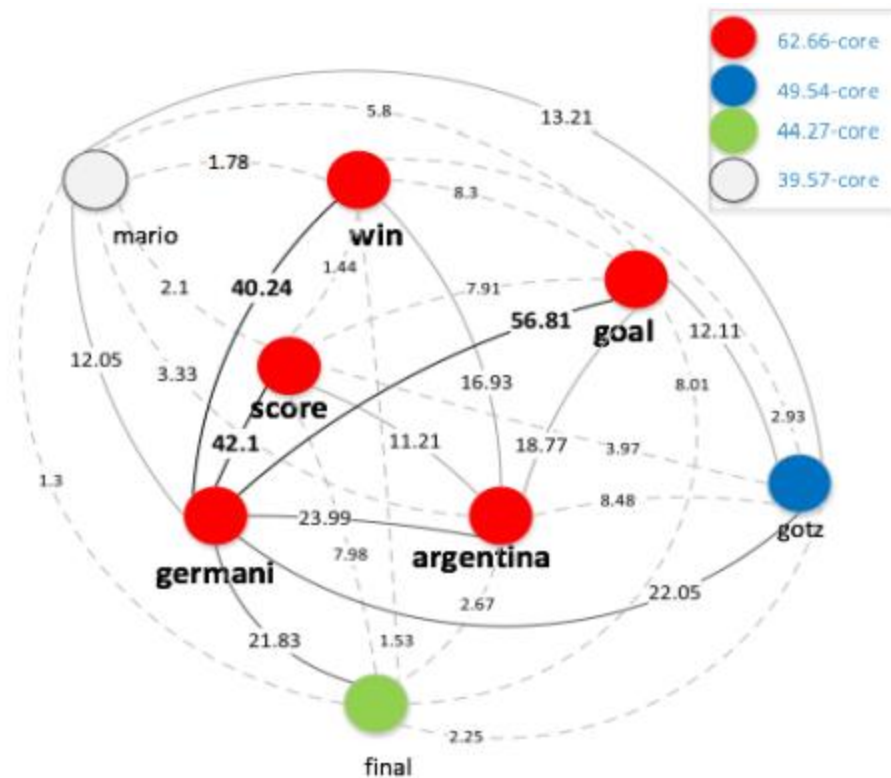
- Exemplo de um grafo a partir do Twitter
- Vértices = palavras
- Arestas = co-ocorrências em um twitter

- 1) Good goal by Neymar
- 2) Goal! Neymar scores for brazil
- 3) Goal!! Neymar scores again
- 4) Watching the game tonight



# Detecção de eventos

- The four highest cores of the graph generated after Germany's goal in the 2014 FIFA World Cup final



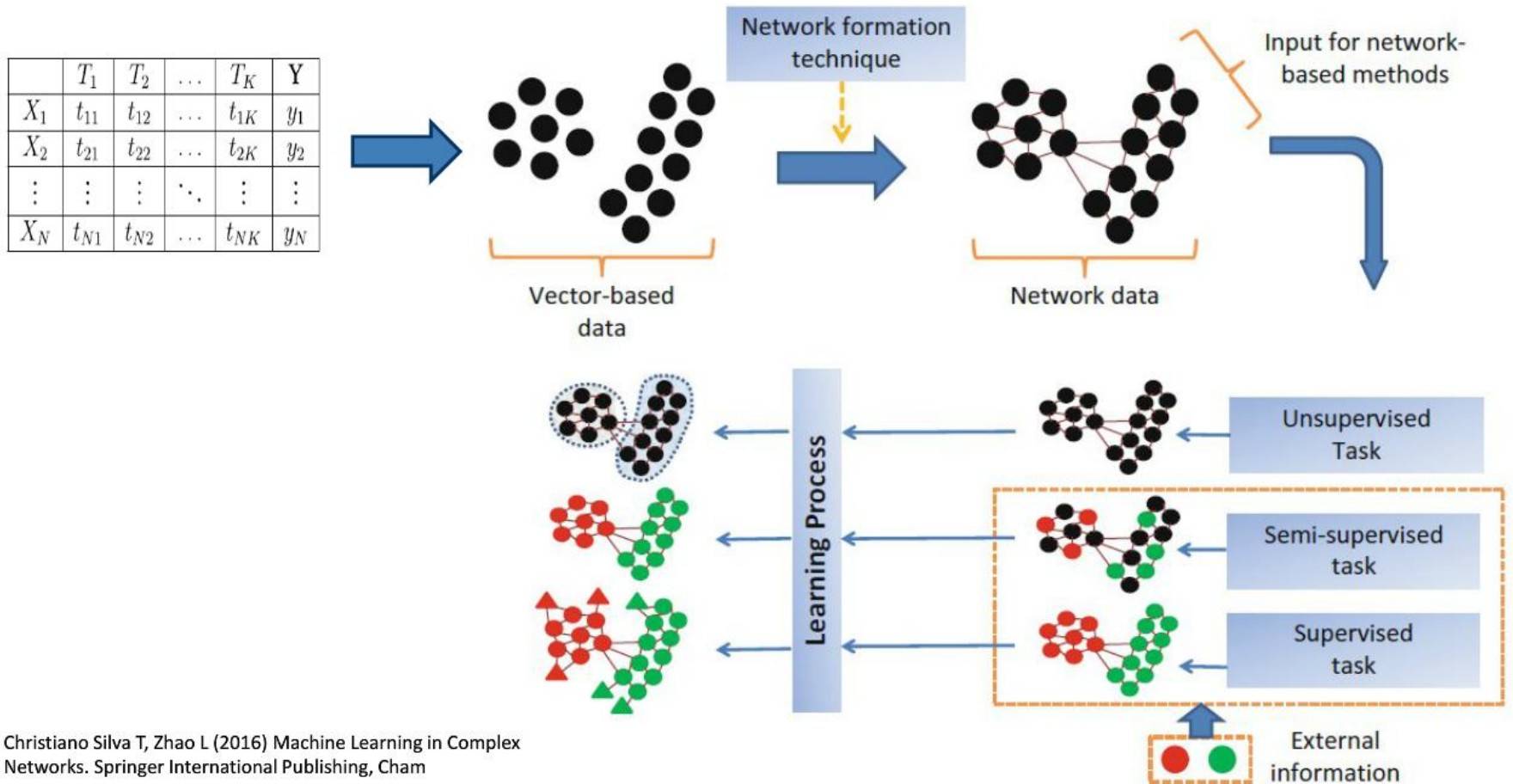
# Tweet summarization

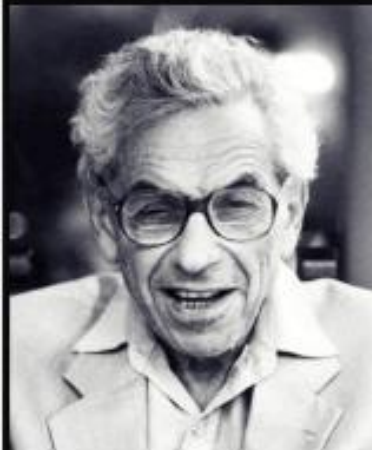
Time	Our Summary	ESPN FC
8'	Goal!!!!Argentina!! After eight minutes Argentina lead Belgium by 1-0 scored by Higuain	Goal! Argentina 1, Belgium 0. Gonzalo Higuain (Argentina) right footed shot from the centre of the box to the bottom left corner.
45'+2'	HT: Argentina 1-0 Belgium. Fantastic goal by Higuain gives Argentina the slight lead over the red devils.	First Half ends, Argentina 1, Belgium 0.
52'	52m - Belgium's Eden Hazard with the first yellow card of the game	Eden Hazard (Belgium) is shown the yellow card for a bad foul.
75'	Argentina 1 - 0 Belgium   Biglia booked a yellow card. Meanwhile, Chadli on for Eden Hazard.	Lucas Biglia (Argentina) is shown the yellow card for a bad foul.
90+5'	Well at least that goal makes them advance to the semi finals. Argentina gets the ticket to advance and Belgium goes home.	Match ends, Argentina 1, Belgium 0.

Summary of the Argentina vs. Belgium match generated automatically using Weight-Core and manually by ESPN.



# Aprendizado de máquina via redes complexas





Minha mente está aberta.  
(Paul Erdos)

**Nascimento:** 26 de março de 1913

**Morte:** 20 de setembro de 1996 (83 anos)

**Biografia:** Paul Erdős foi um matemático húngaro. Extremamente prolífico e de notável excentricidade, publicou 1475 artigos, alguns de extrema importância, o que é um número superior a qualquer outro matemático na história, trabalhando com centenas de colaboradores. Trabalhou em problemas de análise combinatória, teoria dos grafos e teoria dos números.