

Estatística de Redes Sociais

Antonio Galves

Módulo 1

Motivação, quadro conceitual e perguntas básicas

Maiores empresas do mundo em 2019

1. Apple. Valor: US\$ 205.5 bilhões
2. Google. Valor: US\$ 167.7 bilhões
3. Microsoft. Valor: US\$ 125,3 bilhões
4. Amazon. Valor: US\$ 97 bilhões
5. Facebook. Valor: US\$ 88.9 bilhões

Número de usuários do Facebook até abril de 2020

1. Índia: 280 milhões
2. EUA: 190 milhões
3. Indonésia: 130 milhões
4. **Brasil**: 120 milhões

fonte: www.statista.com

Número de usuários do WhatsApp até 2019

1. Índia: 340 milhões
2. **Brasil**: 99 milhões
3. EUA: 68,1 milhões

fonte: www.statista.com

Número de usuários do Twitter até abril 2020

1. EUA: 64,2 milhões
2. Japão: 48,45 milhões
3. Rússia: 23,55 milhões
4. Reino Unido: 17,75 milhões
5. Arábia Saudita: 15 milhões
6. **Brasil**: 14,35 milhões

fonte: www.statista.com

Um problema interdisciplinar

O estudo de redes sociais

- ▶ exige a construção de novos modelos probabilísticos;
- ▶ apresenta grandes desafios para os estatísticos e os cientistas de dados, dado o caráter parcial das amostras disponíveis.
- ▶ É preciso discutir desde as ferramentas computacionais que podem ser usadas para o levantamento de dados nas redes sociais,
- ▶ até as questões de fundo que a análise de redes sociais coloca para a Teoria Estatística e para a Ciência de Dados.

Curso Estatística de Redes Sociais

Objetivos: Fornecer um quadro estatístico para analisar

- ▶ a transmissão da informação
- ▶ a evolução do conjunto de opiniões
- ▶ a **formação rápida de consenso** em redes sociais
- ▶ a identificação de comunidades homogêneas de opinião

Formação rápida de consenso em redes sociais

Estudo de caso: as eleições para governador do Rio de Janeiro em 2018.

Resumo:

- ▶ As sondagens de intenção de votos indicavam uma disputa acirrada entre Romário e Eduardo Paes.
- ▶ Apurados os votos, apareceu em primeiro lugar um obscuro candidato que nas sondagens aparecia numa longínqua terceira posição, ficando Romário fora da disputa.

Estudo de caso: eleições para senador em 2018

- ▶ Na disputa para o Senado no Paraná, Minas Gerais e S.Paulo, Roberto Requião, Dilma Roussef e Eduardo Suplicy, respectivamente, apareciam largamente na frente em todas as sondagens.
- ▶ Os tres foram surpreendentemente derrotados nas urnas.

Estudo de caso: eleições presidenciais de 2018

- ▶ Nas avaliações das intenções de voto para presidente, nenhuma sondagem previu que o Cabo Daciolo ficaria na frente de Henrique Meirelles e Marina Silva, nem que Alckmin só teria a metade dos votos estimados pelas pesquisas prévias.
- ▶ Finalmente e mais importante, as sondagens não previram a quase vitória no primeiro turno do candidato Bolsonaro.

O que aprendemos com as pesquisas de boca de urna?

- ▶ A pesquisa de boca de urna feita pelo IBOPE no dia mesmo da eleição apontou com muita precisão o resultado que sairia efetivamente das urnas.
- ▶ O acerto da pesquisa de boca de urna sugere que a escolha de cidades e grupos sociais utilizada pelos institutos reflete corretamente a diversidade entre regiões e grupos sociais brasileiros.
- ▶ Sugere também que o número de eleitores entrevistados nas pesquisas dos dias anteriores era suficientemente grande para apontar com precisão o resultado da eleição.

Como explicar o erro nas pesquisas de intenção de voto?!

- ▶ No dia 18 de outubro de 2018 a Folha de S.Paulo publicou um artigo da jornalista Patrícia Campos Mello cujo título era: **Empresários bancam campanha contra o PT pelo WhatsApp.**
- ▶ Mas pode uma campanha no WhatsApp explicar os erros nas previsões dos institutos de pesquisa?!

O que o WhatsApp tem a ver com o erro de predição?

- ▶ Mauro Paulino, diretor do instituto DataFolha sugere a resposta a essa pergunta. Num texto publicado em sua conta no Twitter ele diz:
- ▶ *“PESQUISAS ELEITORAIS evidenciaram a impulsão da onda nos momentos finais. RJ, MG e DF são claros exemplos. Ao se comparar as fotos das vésperas, registradas por Ibope e Datafolha, em comparação com a foto das urnas, o fenômeno é claramente explicitado.”*

Uma onda nos últimos dias?

- ▶ Se Mauro Paulino tem razão, houve uma mudança brusca nas intenções de voto nos últimos dias antes da eleição.
- ▶ Conjectura-se que essa mudança foi provocada por uma campanha feita nas redes sociais.
- ▶ Verificar a plausibilidade dessa conjectura é um desafio científico.
- ▶ É preciso modelar estatisticamente como uma onda que muda a opinião de uma grande massa de eleitores em um tempo curtíssimo pode ocorrer numa rede social.

Um quadro conceitual

1. Grafos aleatórios
2. Conjuntos de opiniões evoluindo ao longo do tempo
3. Formação e evolução temporal de redes sociais
4. Detecção de comunidades em redes sociais
5. **Formação rápida de consenso em redes sociais.**
6. Identificação de campanhas de opinião a partir de amostras parciais.

Como modelar uma Rede Social

O modelo tem duas componentes básicas:

1. Uma descrição das **interações** entre os membros da rede (doravante chamados *atores*).

Observação: **Grafo** é o nome do objeto matemático que descreve as interações entre os atores.

2. Uma **função probabilística** descrevendo como a opinião dos demais atores afeta a opinião de cada ator específico.

Um modelo simples de Rede Social

- ▶ Um conjunto finito de atores que interagem entre si.
- ▶ Cada ator emite ao longo do tempo opiniões a respeito de um determinado assunto.
- ▶ Essas opiniões podem ser ou *a favor* (+1) ou *contra* (-1).
- ▶ O sentido de cada nova opinião emitida por um ator **depende probabilisticamente** das últimas opiniões emitidas por seus **influenciadores**.

Um modelo simples de Rede Social

- ▶ $\mathcal{A} = \{1, 2, \dots, N\}$: conjunto de atores, onde N é um número inteiro positivo.
- ▶ $\mathcal{O} = \{+1, -1\}$: conjunto de opiniões possíveis.
- ▶ $X_n(a) \in \mathcal{O}$: última opinião emitida pelo ator a , até o instante n .
- ▶ $A_n \in \mathcal{A}$: ator que emitiu uma opinião no instante n .

Um modelo simples de Rede Social

- ▶ $X_n(a) \in \mathcal{O}$: última opinião emitida pelo ator a , até o instante n .
- ▶ $X_n = (X_n(a) : a \in \mathcal{A})$: lista com as últimas opiniões emitidas pelo conjunto de atores, até o instante n .
- ▶ $X_0, X_1, X_2, \dots, ..$ sequência descrevendo a evolução das opiniões dos diversos atores ao longo do tempo.

Definindo o modelo $(X_n)_{n \geq 0}$

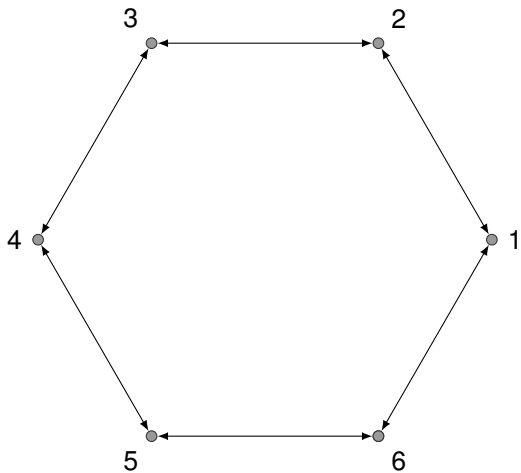
Recapitulando: Os ingredientes do modelo são:

- ▶ A especificação das relações de influência entre os atores;
- ▶ A especificação da função probabilística usada na escolha de cada nova opinião de um ator, em função das últimas opiniões de seus influenciadores.

Conjuntos de influenciadores. Exemplo 1

- ▶ Seja $\mathcal{A} = \{1, \dots, N\}$ o conjunto de atores da rede;
- ▶ Para cada ator $a \in \mathcal{A}$, denotamos $\mathcal{V}_{\rightarrow a}$ o conjunto de seus influenciadores,
- ▶ Se $a = 2, \dots, N - 1$, $\mathcal{V}_{\rightarrow a} = \{a - 1, a + 1\}$
- ▶ $\mathcal{V}_{\rightarrow 1} = \{N, 2\}$, $\mathcal{V}_{\rightarrow N} = \{N - 1, 1\}$.

Exemplo 1



Conjuntos de influenciadores. Exemplo 2

- ▶ Seja $\mathcal{A} = \{(a_1, a_2) : a_1 = 1, \dots, N, a_2 = 1, \dots, N\}$, o conjunto de atores da rede;
- ▶ Para cada ator $(a_1, a_2) \in \mathcal{A}$,

$$\mathcal{V}_{\cdot \rightarrow (a_1, a_2)} = \{(a_1+1, a_2), (a_1-1, a_2), (a_1, a_2+1), (a_1, a_2-1)\},$$

também com a convenção $N+1 = 1$.

Exemplo de função probabilística usada por um ator para escolher uma nova opinião

- ▶ Seja $U_n^a(+1)$ o número de influenciadores do ator a , cuja última opinião emitida até o instante n foi $+1$.
- ▶ Seja $U_n^a(-1)$, o número de influenciadores do ator a , cuja última opinião emitida até o instante n foi -1 .
- ▶ Se o ator a decidir emitir uma opinião no instante $n + 1$, ele escolherá a opinião $+1$, com probabilidade

$$p_n^a(+1) = \frac{U_n^a(+1)}{U_n^a(+1) + U_n^a(-1)}$$

- ▶ e escolherá a opinião -1 , com probabilidade

$$p_n^a(-1) = \frac{U_n^a(-1)}{U_n^a(+1) + U_n^a(-1)}$$

Um algoritmo gerando a sequência $(X_n)_{n \geq 0}$

1. Escolho arbitrariamente a lista inicial de opiniões

$$X_0 = (X_0(a), a \in \mathcal{A})$$

2. Para $n = 1, \dots, T$, onde $T \geq 1$ é um número inteiro arbitrário, faça:

2.1 Sorteie $A_n \in \mathcal{A}$ independentemente dos sorteios passados, com $\mathbb{P}\{A_n = b\} = 1/N$, para todo $b \in \mathcal{A}$

2.2 Tendo já gerado $X_{n-1} = (X_{n-1}(a) : a \in \mathcal{A})$ e já tendo sorteado $A_n = b$, escolha $O_n = o \in \mathcal{O}$, com probabilidade

$$\mathbb{P}\{O_n = o \mid X_{n-1}, A_n = b\} = p_n^b(o)$$

2.3 Para todo $a \in \mathcal{A}$, defina $X_n(a) = O_n$, se $a = A_n$ e $X_n(a) = X_{n-1}(a)$, se $a \neq A_n$.

Exercícios

1. Escreva um código implementando o algoritmo apresentado na transparência anterior.
2. Utilize esse código para simular redes sociais correspondendo aos exemplos 1 e 2, com $N = 10$ e $T = 100$ e $T = 1000$.
3. O que a simulação sugere em termos de constituição de consenso na rede?
4. Para quais valores das listas iniciais de opiniões X_0 , teremos $X_n = X_0$, para todo $n \geq 1$?