Information Cascades

Ana Paula
Fabrício Benevenuto

Following the crowd

- Pessoas são influenciadas por outras
  - Produtos que compram
  - Atividades que participam
  - Tecnologias que utilizam
  - Etc.

Por que uma das mídias prevaleceu?

Blu-ray Disc

HD DVD

Movimento de manadas
Um experimento

Movimento de manadas
Um experimento

- Há uma decisão a ser feita
- Pessoas tomam decisões sequencialmente e observam as decisões tomadas anteriormente
- Cada pessoa tem uma informação privada que ajuda a tomar a decisão
- A pessoa pode tomar a decisão baseada na informação privada ou na informação fornecida pelos outros

Movimento de manadas
Um experimento

- Primeiro aluno:
  - Decisão simples baseada em informação privada

- Segundo aluno
  - Confirma a decisão do primeiro aluno
  - Ou decide. Espera-se que siga a informação privada

- Terceiro aluno
  - Se os dois anteriores fizeram escolhas diferentes, ele precisa desempatar. Tende a seguir informação privada.
  - Senão tende a seguir a maioria e ignorar informação privada

- Quarto aluno em diante
  - Tende a seguir a maioria

Bayes' Rule: A Model of Decision Making under Uncertainty

- Qual a probabilidade de termos a maioria vermelha, dado que o papel que eu tirei e as escolhas das outras pessoas?
- Qual a probabilidade de se escolher o melhor restaurante, dadas as revisões que eu li sobre eles?

"We will compute the probabilities of various *events*, and use these probabilities to reason about decision making."

Conditional Probability

"… we need to consider the probability of A *given that some other event B has occurred*.

…

We will refer to this quantity as the *conditional probability of A given B* and denote it by **Pr[A|B]**."

Sample Space

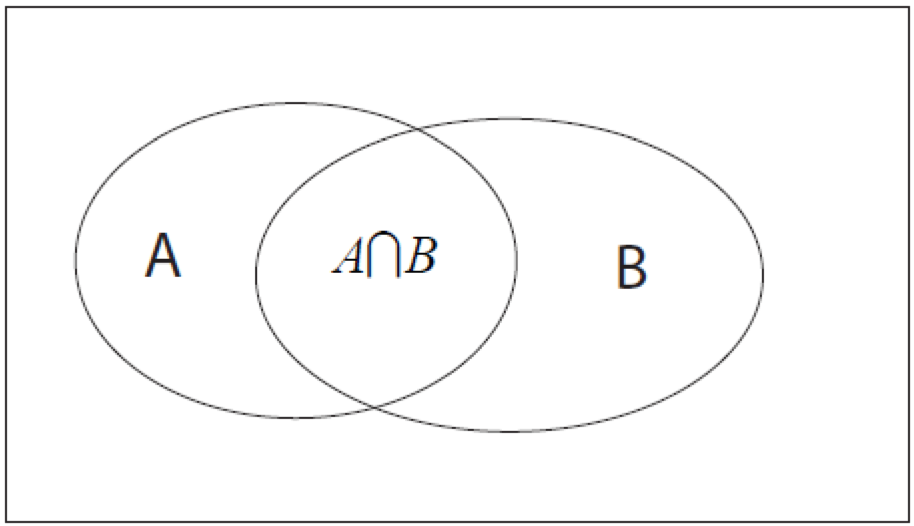


Figure 16.1: Two events A and B in a sample space, and the joint event $A \cap B$.

Conditional Probability of A given B and the Bayes' Rule

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}. \tag{16.1}$$

Similarly, the conditional probability of B given A is

$$\Pr[B \mid A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{\Pr[A \cap B]}{\Pr[A]}, \tag{16.2}$$

where the second equality follows simply because $A \cap B$ and $B \cap A$ are the same set.

Rewriting (16.1) and (16.2), we have

$$\Pr[A \mid B] \cdot \Pr[B] = \Pr[A \cap B] = \Pr[B \mid A] \cdot \Pr[A], \tag{16.3}$$

and therefore, dividing through by $\Pr[B]$,

$$\Pr[A \mid B] = \frac{\Pr[A] \cdot \Pr[B \mid A]}{\Pr[B]}. \tag{16.4}$$

Examples of Bayes' Rule

To maximize her chance of winning the monetary reward for guessing correctly, she should guess majority-blue if

$$\Pr\left[\textit{majority-blue} \mid \textit{what she has seen and heard}\right] > \frac{1}{2}$$

$$\Pr\left[\textit{majority-blue}\right] = \Pr\left[\textit{majority-red}\right] = \frac{1}{2}.$$

$$\Pr\left[\textit{blue} \mid \textit{majority-blue}\right] = \Pr\left[\textit{red} \mid \textit{majority-red}\right] = \frac{2}{3}.$$

Decompõe o denominador em 2 partes: maioria azul e maioria vermelha

$$\begin{aligned}
\Pr\left[\textit{blue}\right] &= \Pr\left[\textit{majority-blue}\right] \cdot \Pr\left[\textit{blue} \mid \textit{majority-blue}\right] + \\
&\quad \Pr\left[\textit{majority-red}\right] \cdot \Pr\left[\textit{blue} \mid \textit{majority-red}\right] \\
&= \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}.
\end{aligned}$$

$$\Pr\left[\textit{majority-blue} \mid \textit{blue}\right] = \frac{\Pr\left[\textit{majority-blue}\right] \cdot \Pr\left[\textit{blue} \mid \textit{majority-blue}\right]}{\Pr\left[\textit{blue}\right]}.$$

$$\Pr\left[\textit{majority-blue} \mid \textit{blue}\right] = \frac{1/3}{1/2} = \frac{2}{3}.$$

Examples of Bayes' Rule

$$\Pr[\textit{majority-blue} \mid \textit{blue, blue, red}]$$

so as to make a guess about the urn. Using Bayes' Rule we get

$$\Pr[\textit{majority-blue} \mid \textit{blue, blue, red}] = \frac{\Pr[\textit{majority-blue}] \cdot \Pr[\textit{blue, blue, red} \mid \textit{majority-blue}]}{\Pr[\textit{blue, blue, red}]}. \tag{16.7}$$

Since the draws from the urn are independent, the probability $\Pr[\textit{blue, blue, red} \mid \textit{majority-blue}]$ is determined by multiplying the probabilities of the three respective draws together:

$$\Pr[\textit{blue, blue, red} \mid \textit{majority-blue}] = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{27}.$$

To determine $\Pr[\textit{blue}, \textit{blue}, \textit{red}]$, as usual we consider the two different ways this sequence could have happened — if the urn is majority-blue, or if it is majority-red:

$$\begin{aligned}
\Pr[\textit{blue, blue, red}] &= \Pr[\textit{majority-blue}] \cdot \Pr[\textit{blue, blue, red} \mid \textit{majority-blue}] + \\
&\quad \Pr[\textit{majority-red}] \cdot \Pr[\textit{blue, blue, red} \mid \textit{majority-red}] \\
&= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{6}{54} = \frac{1}{9}.
\end{aligned}$$

Plugging all this back into Equation (16.7), we get

$$\Pr[\textit{majority-blue} \mid \textit{blue, blue, red}] = \frac{\frac{4}{27} \cdot \frac{1}{2}}{\frac{1}{9}} = \frac{2}{3}.$$

When do the Cascades begin?

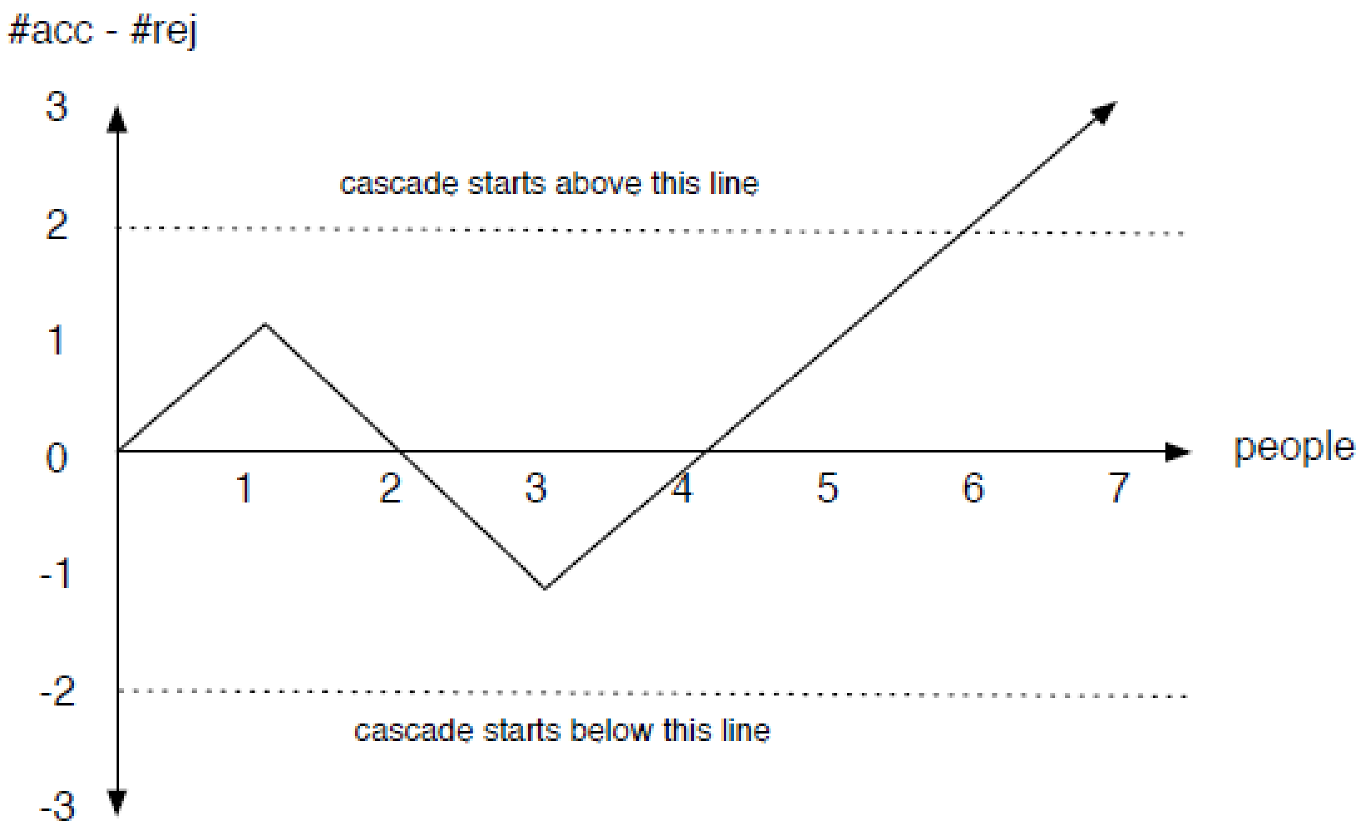


Figure 16.3: .

Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market

Matthew J. Salganik[1,2,*], Peter Sheridan Dodds[2,*], Duncan J. Watts[1,2,3,*]

± Author Affiliations

[*] To whom correspondence should be addressed. E-mail: mjs2105@columbia.edu (M.J.S.); pd315@columbia.edu (P.S.D.); djw24@columbia.edu (D.J.W.)

ABSTRACT

Hit songs, books, and movies are many times more successful than average, suggesting that “the best” alternatives are qualitatively different from “the rest”; yet experts routinely fail to predict which products will succeed. We investigated this paradox experimentally, by creating an artificial “music market” in which 14,341 participants downloaded previously unknown songs either with or without knowledge of previous participants' choices. Increasing the strength of social influence increased both inequality and unpredictability of success. Success was also only partly determined by quality: The best songs rarely did poorly, and the worst rarely did well, but any other result was possible.

Lessons from Cascades

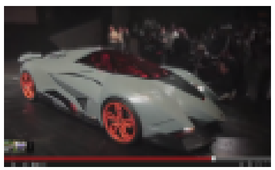
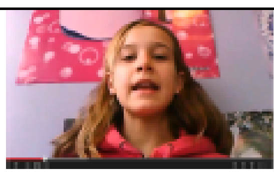
- Cascades can be wrong
  - Divergence on the literature (Book: The Wisdow of the Crowds)
- Cascades can be based on very little information
- Cascades are fragile

Cascades on Real Data

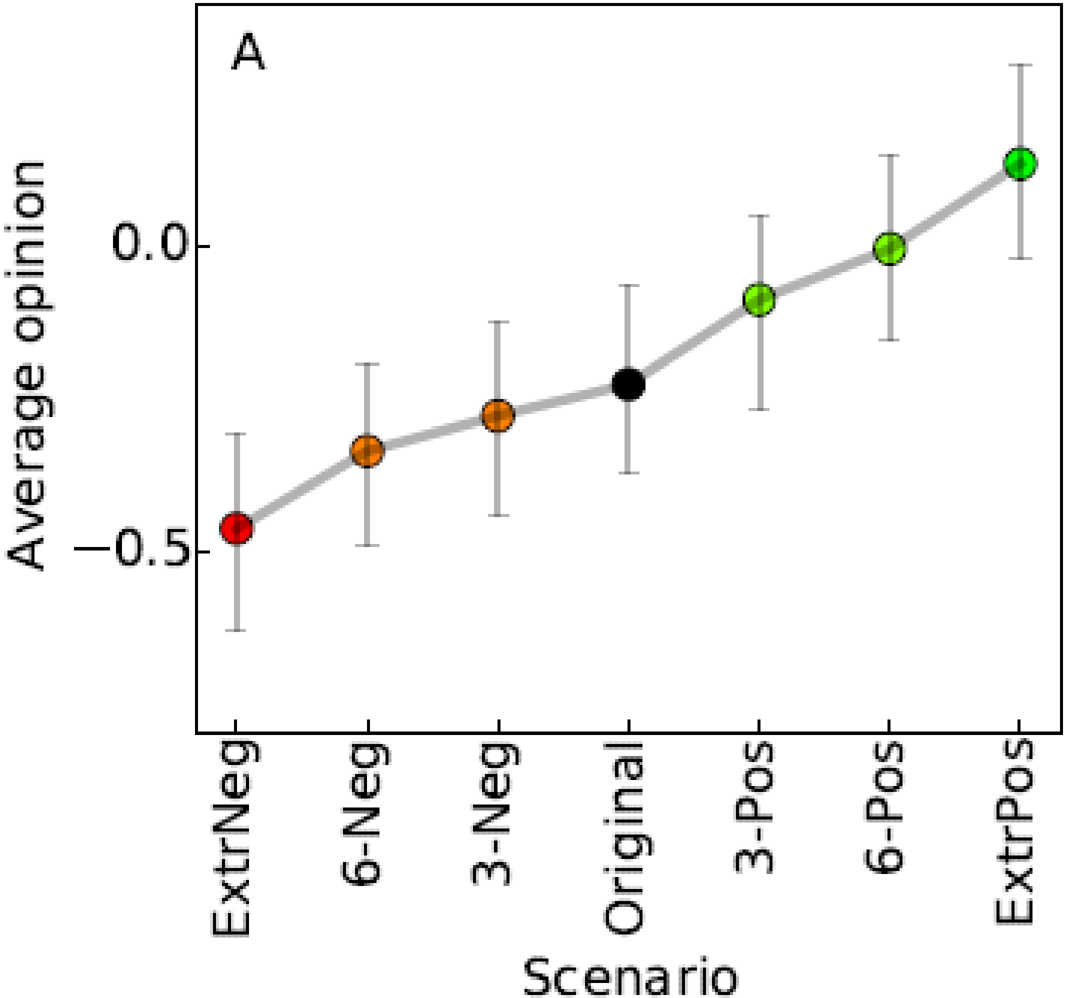
An experimental study of opinion influenceability
Przemyslaw A. Grabowicz, Francisco Romero-Ferrero, Theo Lins, Gonzalo G. de Polavieja, Fabrício Benevenuto, Krishna P. Gummadi

http://arxiv.org/abs/1512.00770.

Manipulação do YouTube

Video	Thumbnail	Positive comments	Negative comments
Lamborghini		i lov lambogini myt muther hav lambogini shit-tere The car is ahead of its time in 20 years all cars will look like this just like how Lamborghini always surprises with its futuristic designs, since the Countach came out =)	omg monster? Who design this? I will kill him I seriously hope not. This is fucking disgusting.
A girl singing		U are soooooooo good i wish i could sing like tht u are the best i bet ur friends and family are sooo proud of u u should go on Britain's Got Talent i wish i was ur bestfriend how old are u? xxxxx... wow! you are good! AMAZINGGGGGGGG!!!!!!!!!!	Again you are good but that's kinda selfish and rude telling people your the best well you're not the best and also a little flat... listen sweetie any who says their the best hon-estly arent so yeah i can be honest but u dont want me to be i mean would u feel better if i lied

Manipulação do YouTube



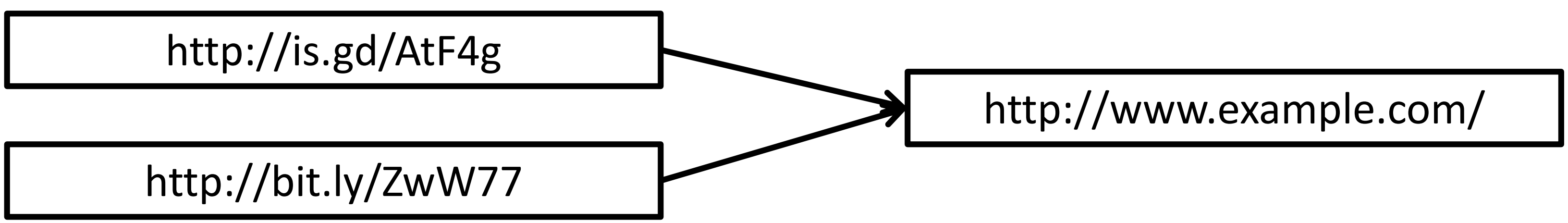
ACM IMC 2011

On Word-of-Mouth Based Discovery of the Web.

Tiago Rodrigues, Fabricio Benevenuto, Meeyoung Cha,
Krishna Gummadi, Virgílio Almeida.

Extração de URLs

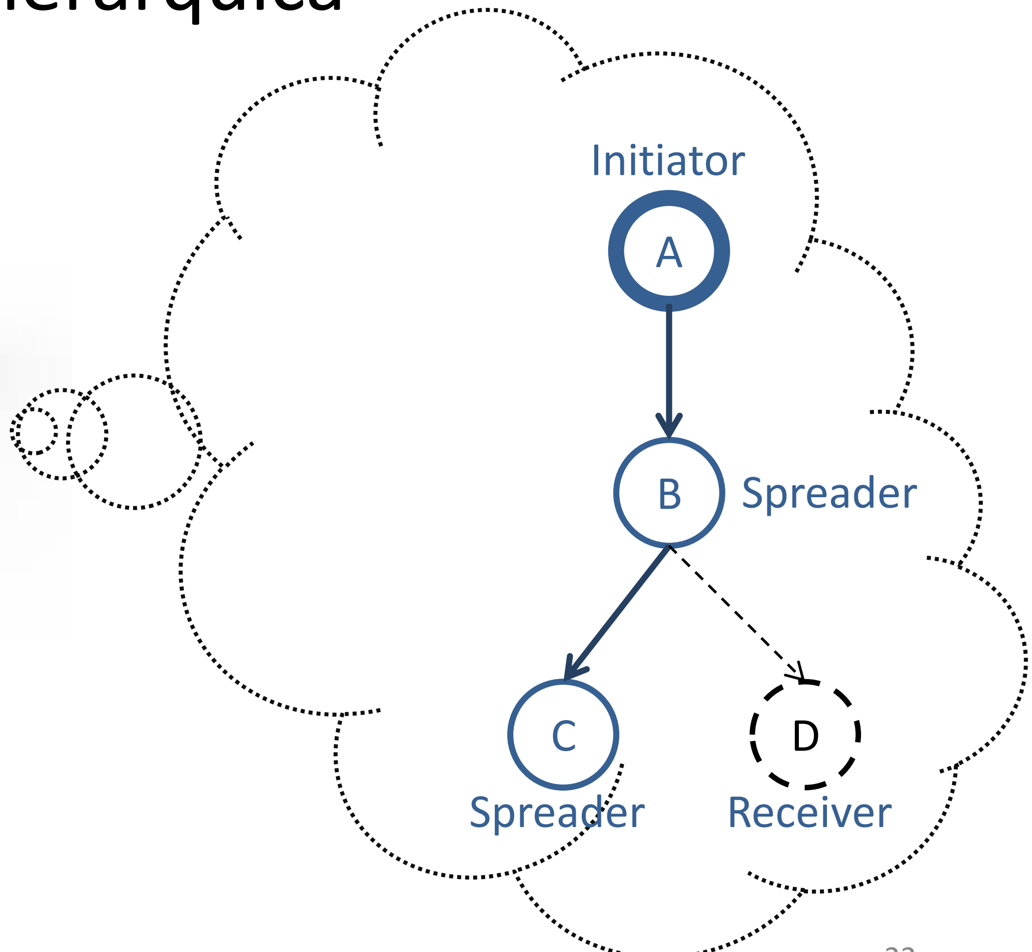
- Tradução de URLs encurtadas



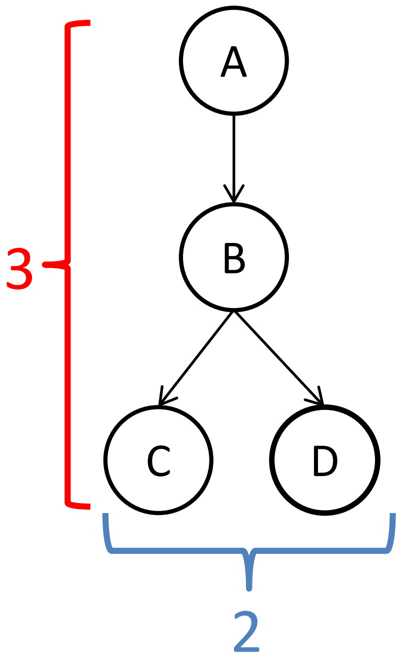
Modelando propagação de informação

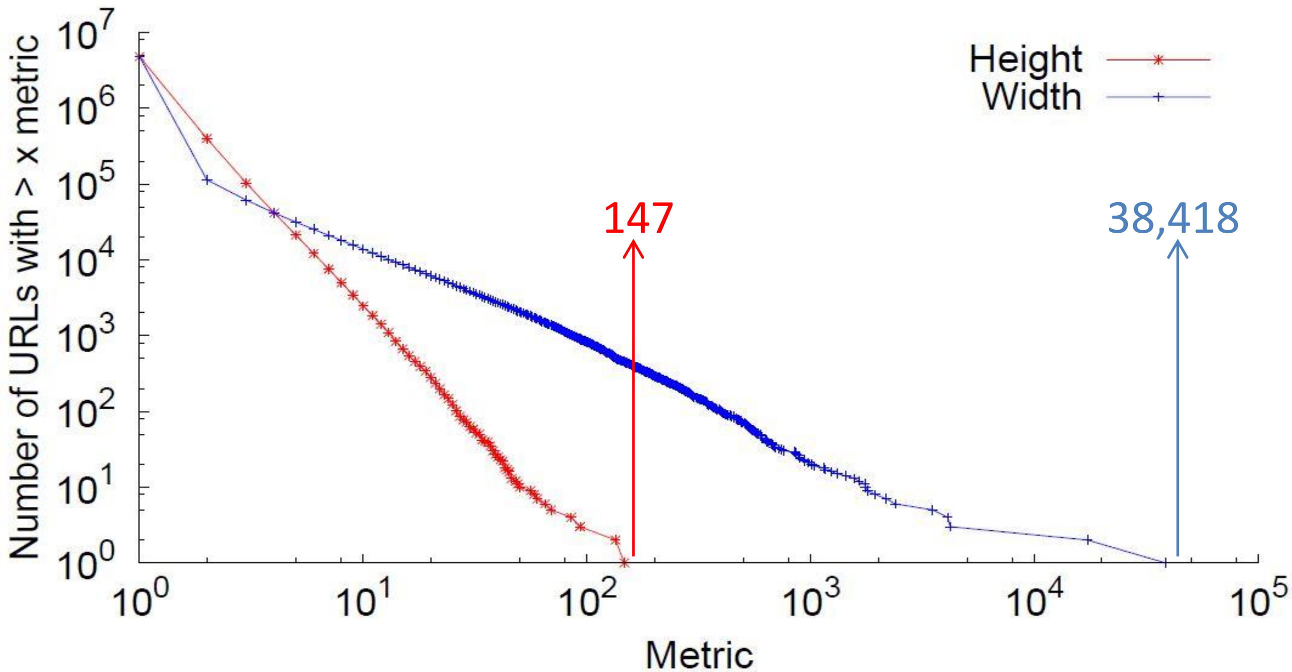
- Modelo de Árvore Hierárquica

Audience



Qual a estrutura típica de árvores de propagação?



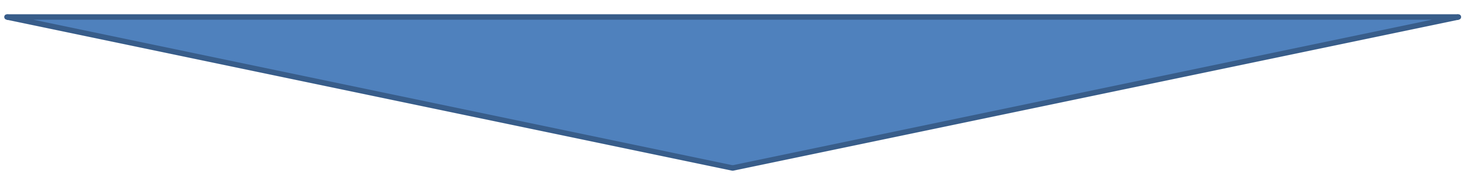


Árvores são muito mais largas do que altas

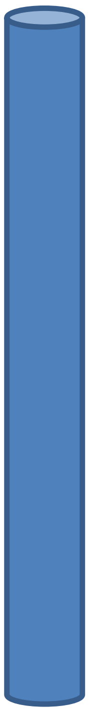
Propagação no Twitter Cascades vs. Propagação por E-mails

- D. Liben-Nowell and J. Kleinberg
  - Tracing Information Flow on a Global Scale using Internet Chain-Letter Data, PNAS, 2008

Data Source	Nodes	Height	Width
PNAS [31]	18,119	288 (med)	82
Twitter	26,227	23 (max)	17,255
PNAS (estimated)	980	16 (med)	4
Twitter (900-1,100)	980 ± 0.02	20 ± 0.25	398 ± 0.27
PNAS (estimated)	162	3 (med)	1
Twitter (100-300)	162 ± 0.02	10 ± 0.03	86 ± 0.03



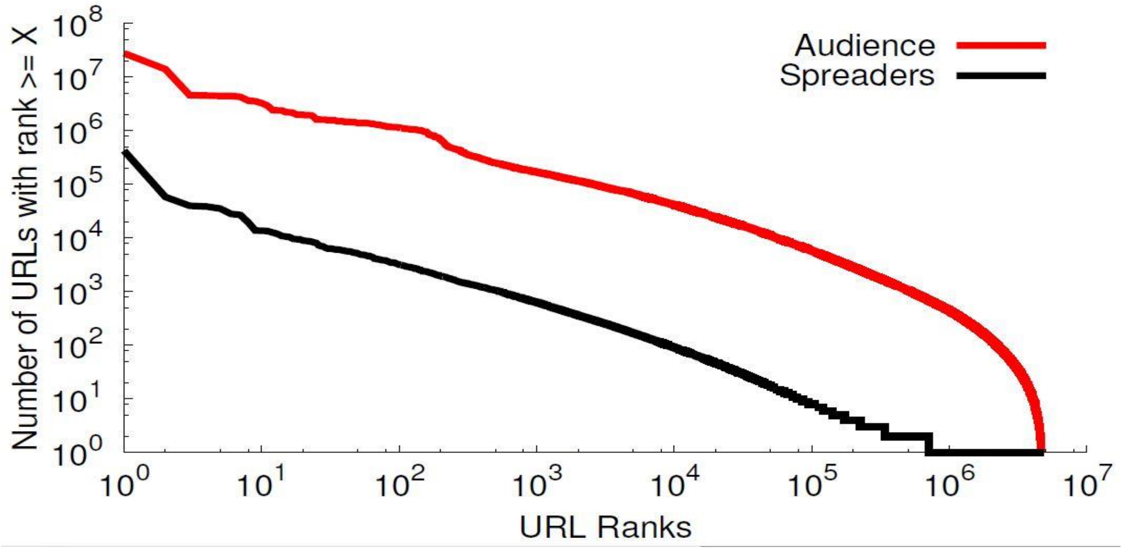
Twitter



e-mail

Quão importante é o papel da propagação na audiência?

- Mais popular: 426.820 espalhadores e audiência de 28M de usuários



- Propagação pode resultar em grandes cascatas

O que essas análises inspiraram

- Ideia: Encontrar usuários que adotam e espalham amplamente novas tendências
  - Algoritmo para identificação de trendsetters

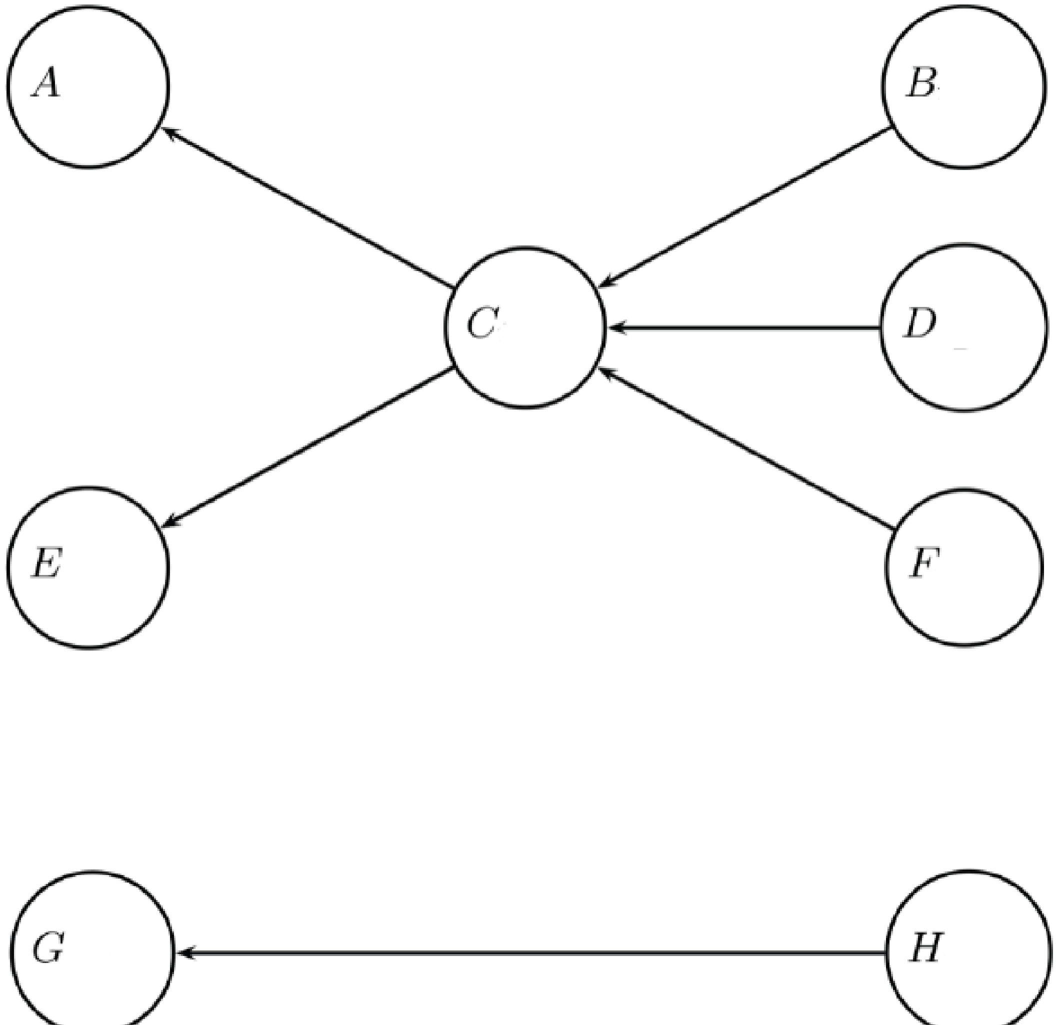
- Como encontrá-los?
  - Algoritmo do Pagerank com peso baseado no tempo de postagem
  - Modelo de identificação de tópicos

ACM KDD 2012

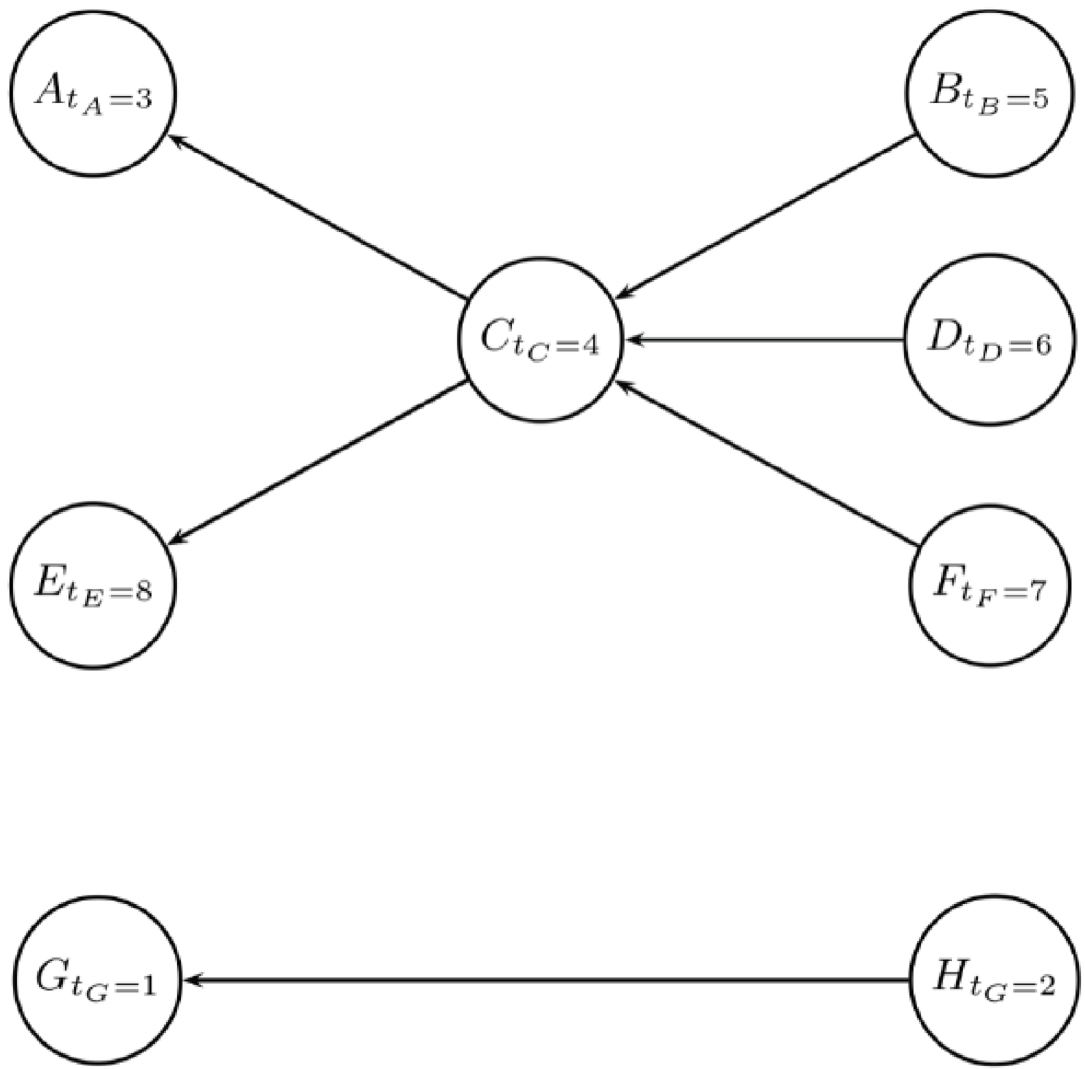
Finding Trendsetters in Information Networks.

Diego Saez-Trumper, Giovanni Comarela, Virgílio Almeida,
Ricardo Baeza-Yates, Fabrício Benevenuto.

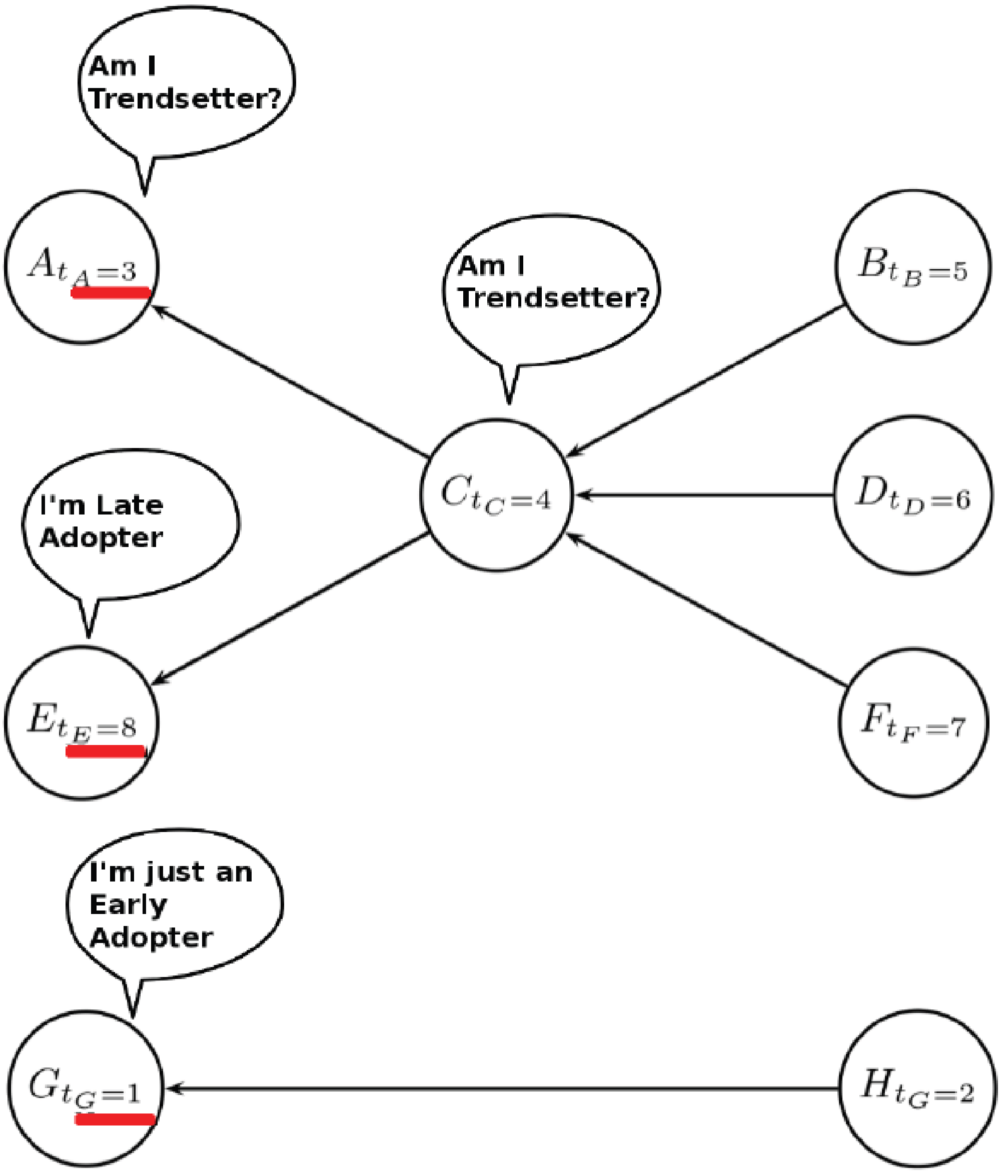
Quem são os trendsetters nesse grafo?



Quem são os trendsetters nesse grafo?



Quem são os trendsetters nesse grafo?



Como encontrá-los?

- Algoritmo do Pagerank com peso baseado no tempo de postagem

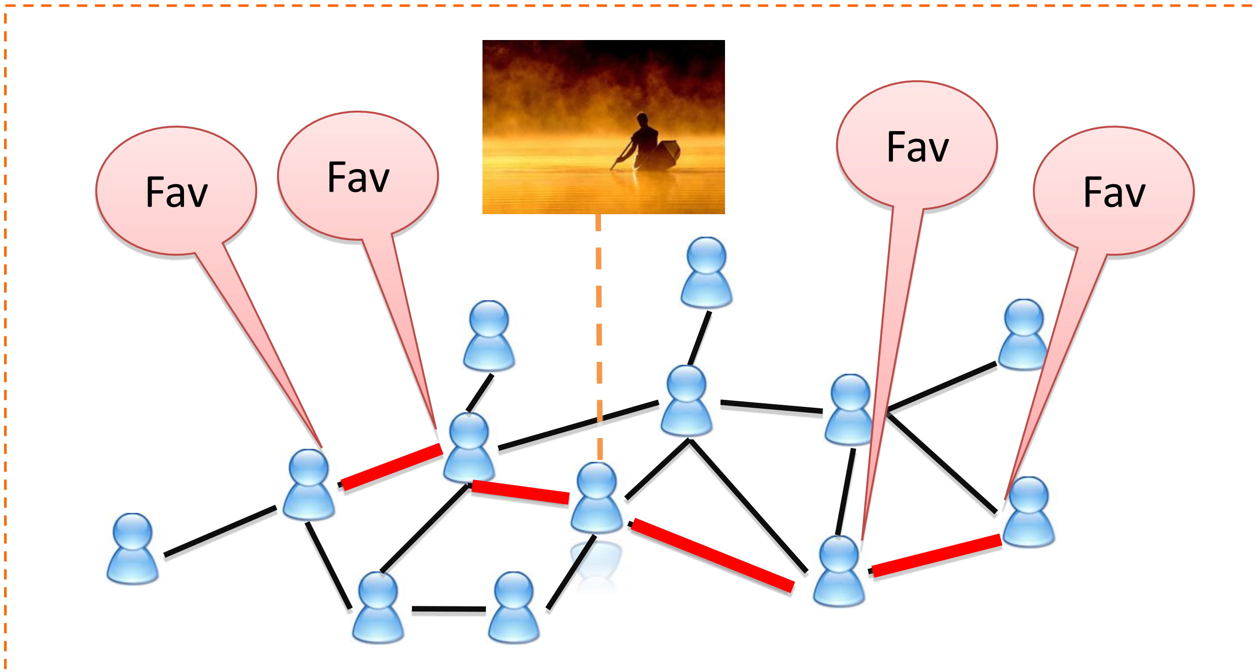
- Modelo de identificação de tópicos

Computer Networks 2012

Delayed Information Cascades in Flickr: Measurement, Analysis, and Modeling.

Meeyoung Cha, Fabrício Benevenuto,
Young-Yeol Ahn and Krishna Gummadi.

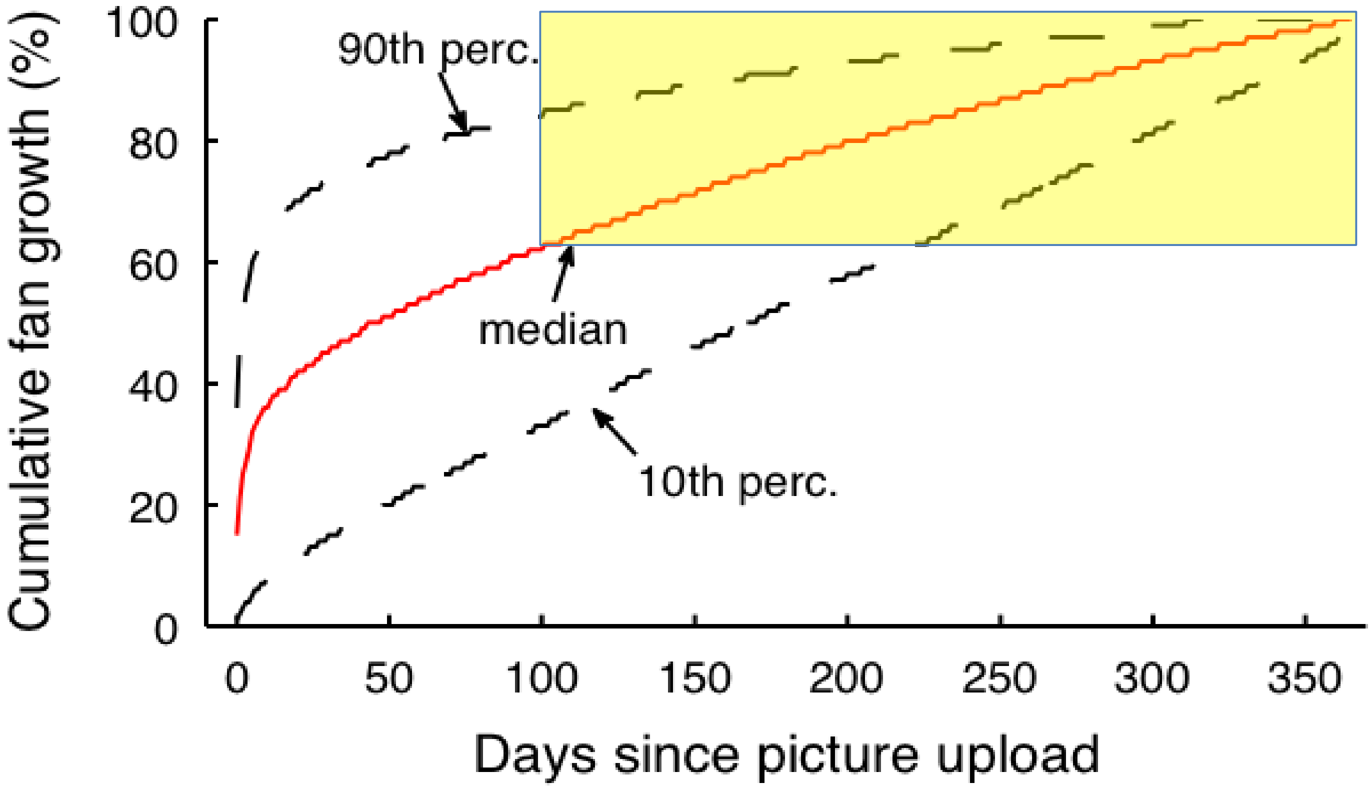
Propagação de informação no Flickr



- Rede de amizades possui propriedades small-world
  - A topologia é apropriada para a propagação de informação
- A informação se propaga lentamente

Padrão de Crescimento Agregado

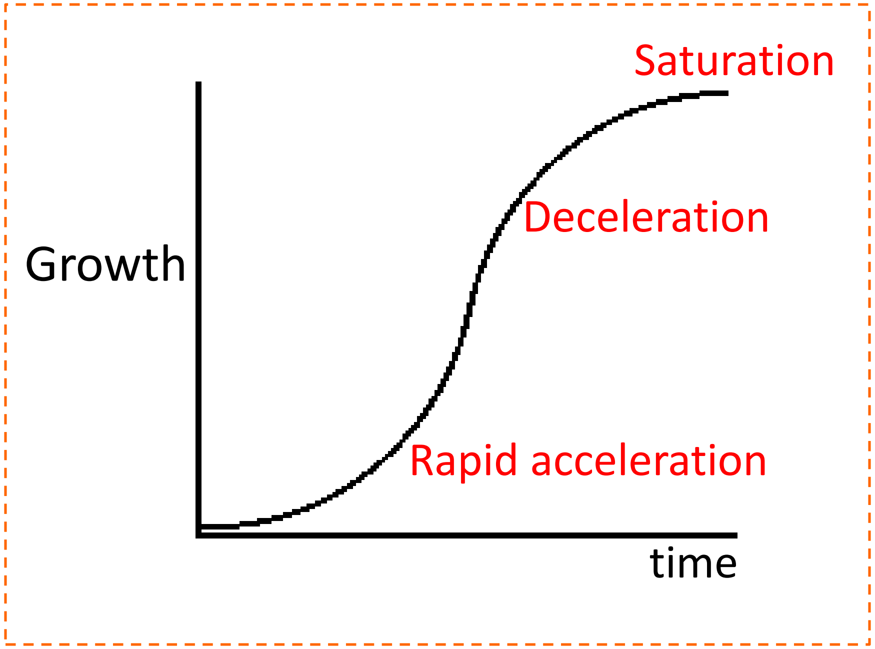
- 5,346 photos (Mais velhas do que 1 ano com mais de 100 fans)



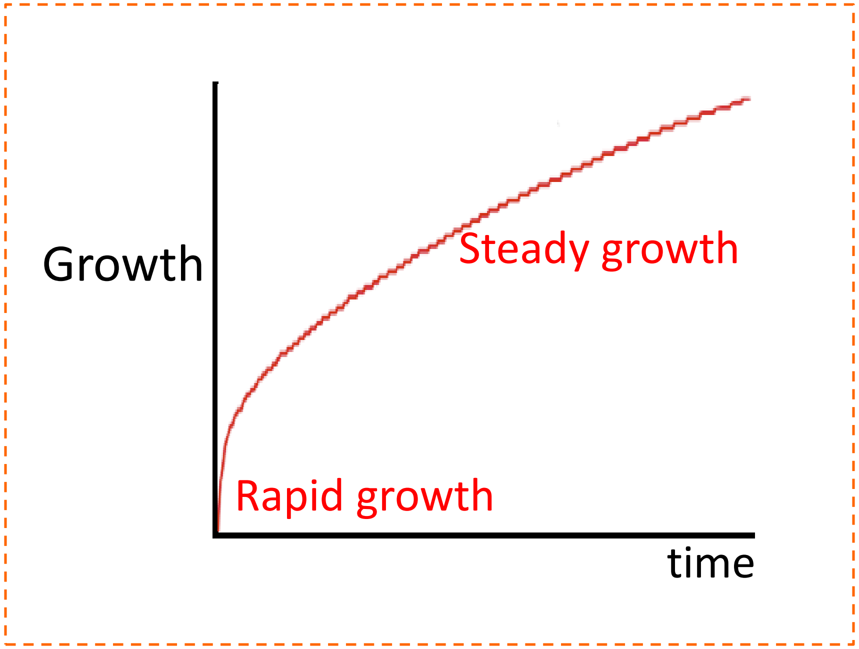
Fotos ganham popularidade lenta e continuamente.
Usuários do Flickr demoram até descobrir novas fotos.

Comparando padrões de crescimento: teoria vs. prática

- Teorias Populares (Diffusion of innovations, Bass diffusion model, SIR model) sugerem padrão de crescimento com curva em S



Teoria

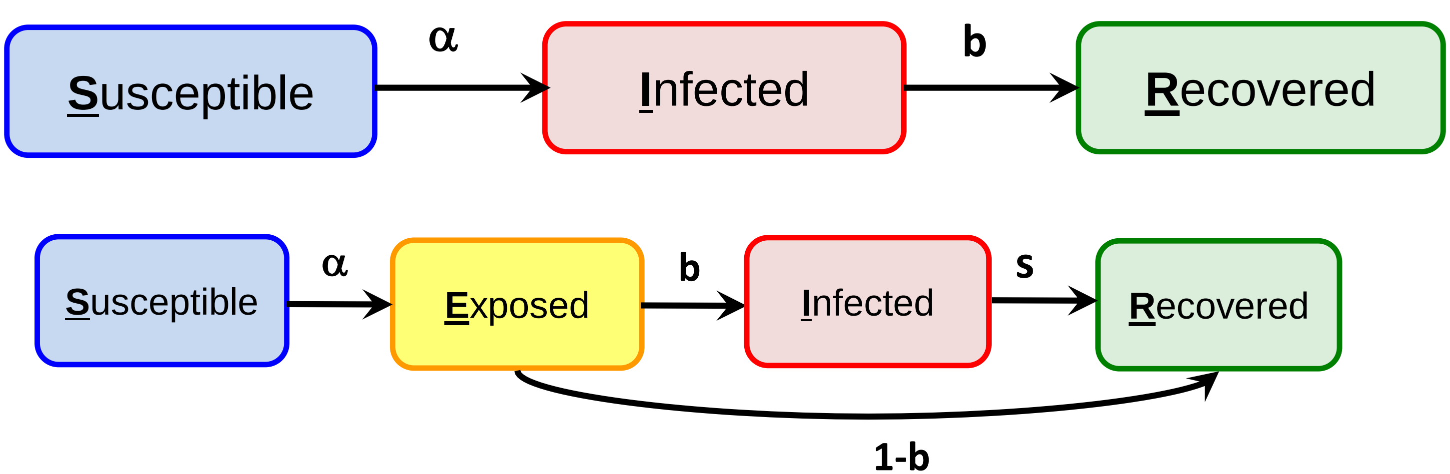


Dados reais do Flickr

Teste de diversas hipóteses

- Crescimento do sistema?
  - Padrão observado mesmo quando novos usuários foram desconsiderados

- Hipótese: Excesso de informação é uma fonte de atraso na propagação de informação
  - Frequência do login no Flickr
  - Usuários só acessam as primeiras páginas de conteúdo



ICDM 2013

Prominent Features of Rumor Propagation in Online Social Media

Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang

Rumores se espalham rápido!

23/04/2013
1:07PM

2:13PM