Finding Trendsetters in Information Networks

Diego Saez-Trumper[1] Giovanni Comarela[2]

Virgilio Almeida[2] Ricardo Baeza-Yates[3] Fabricio Benevenuto[4]

[1]Universitat Pompeu Fabra, Barcelona
[2]Universidade Federal Minas Gerais, Brazil
[3]Yahoo! Research, Barcelona
[4]Universidade Federal de Ouro Preto, Brazil

Beijing, August, 2012

What is a Trendsetter?

What is a Trendsetter?

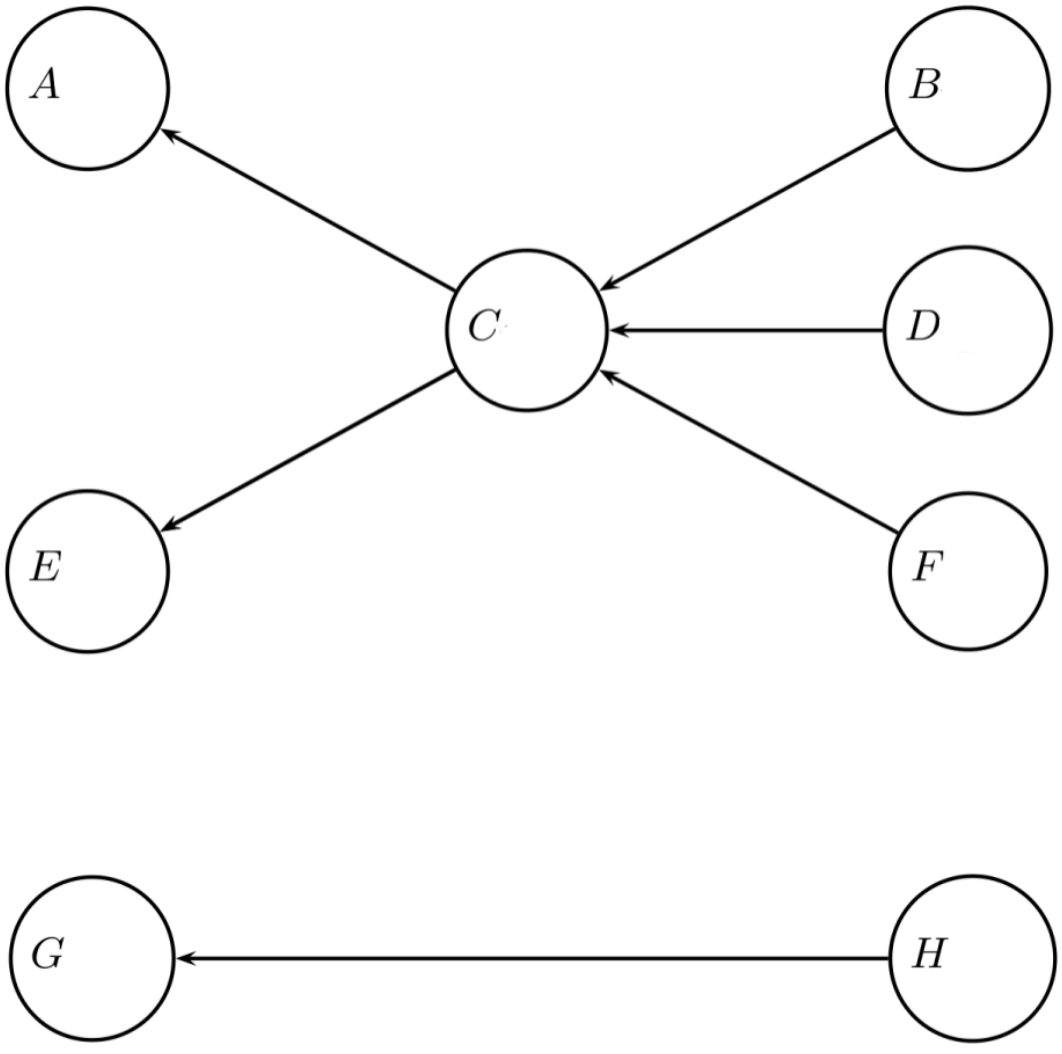
Trendsetters are people:

- Adopt and spread new trends before these trends become popular.
- Propagate these trends over the network.

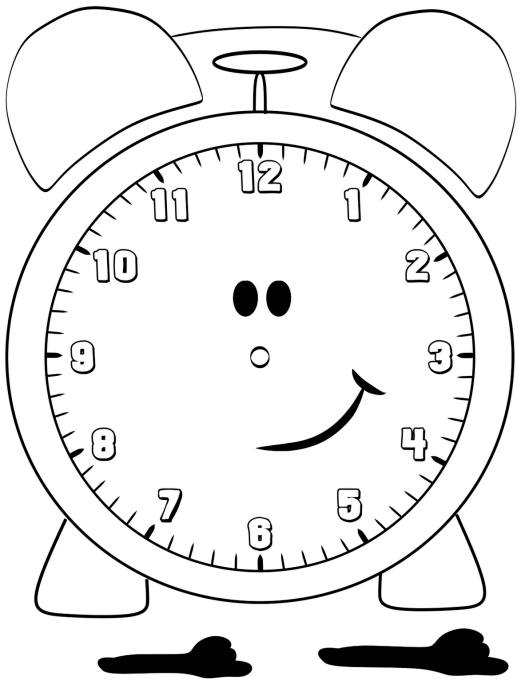
	node indegree	threshold for adoption of new ideas
follower hubs	high	high
innovation hubs	lower	low

Finding trendsetters in a graph

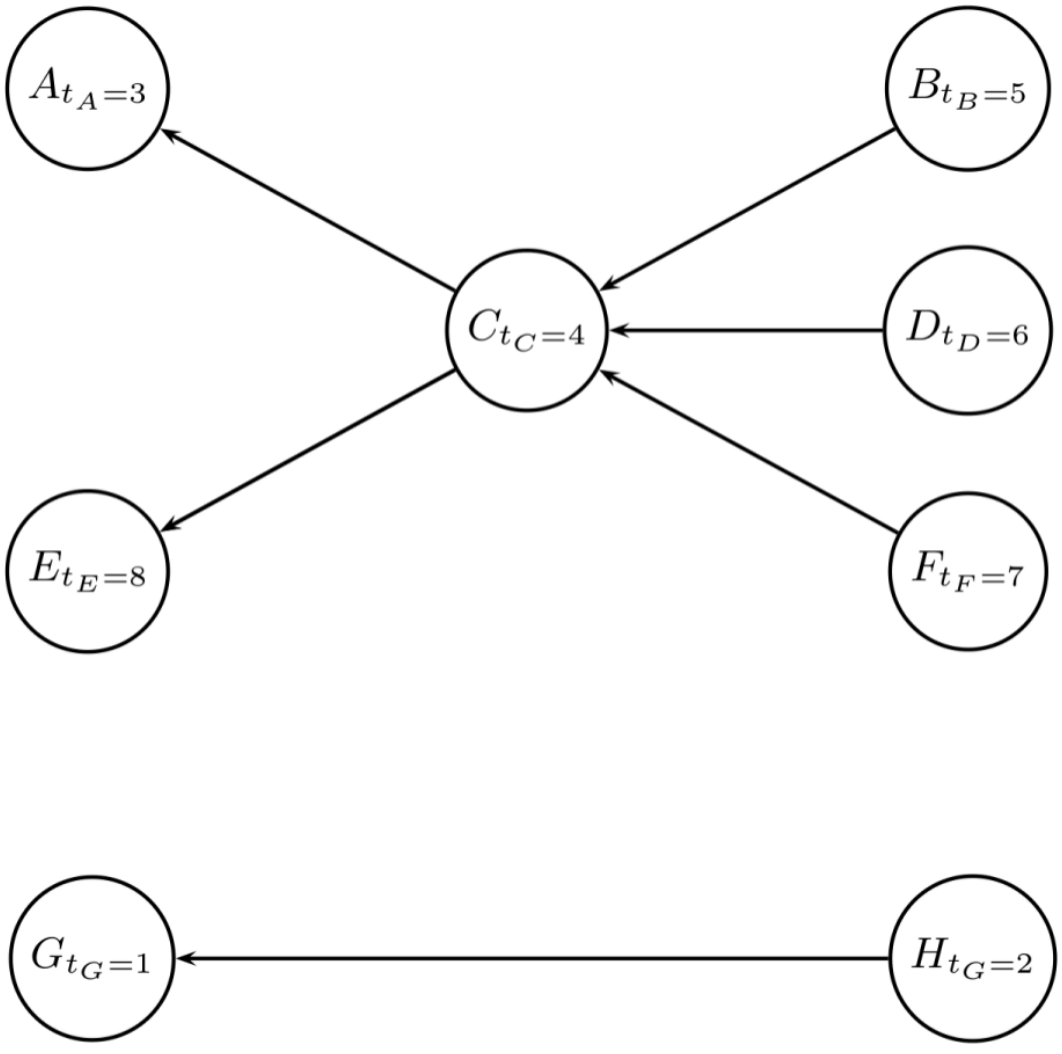
Who are the trendsetters?



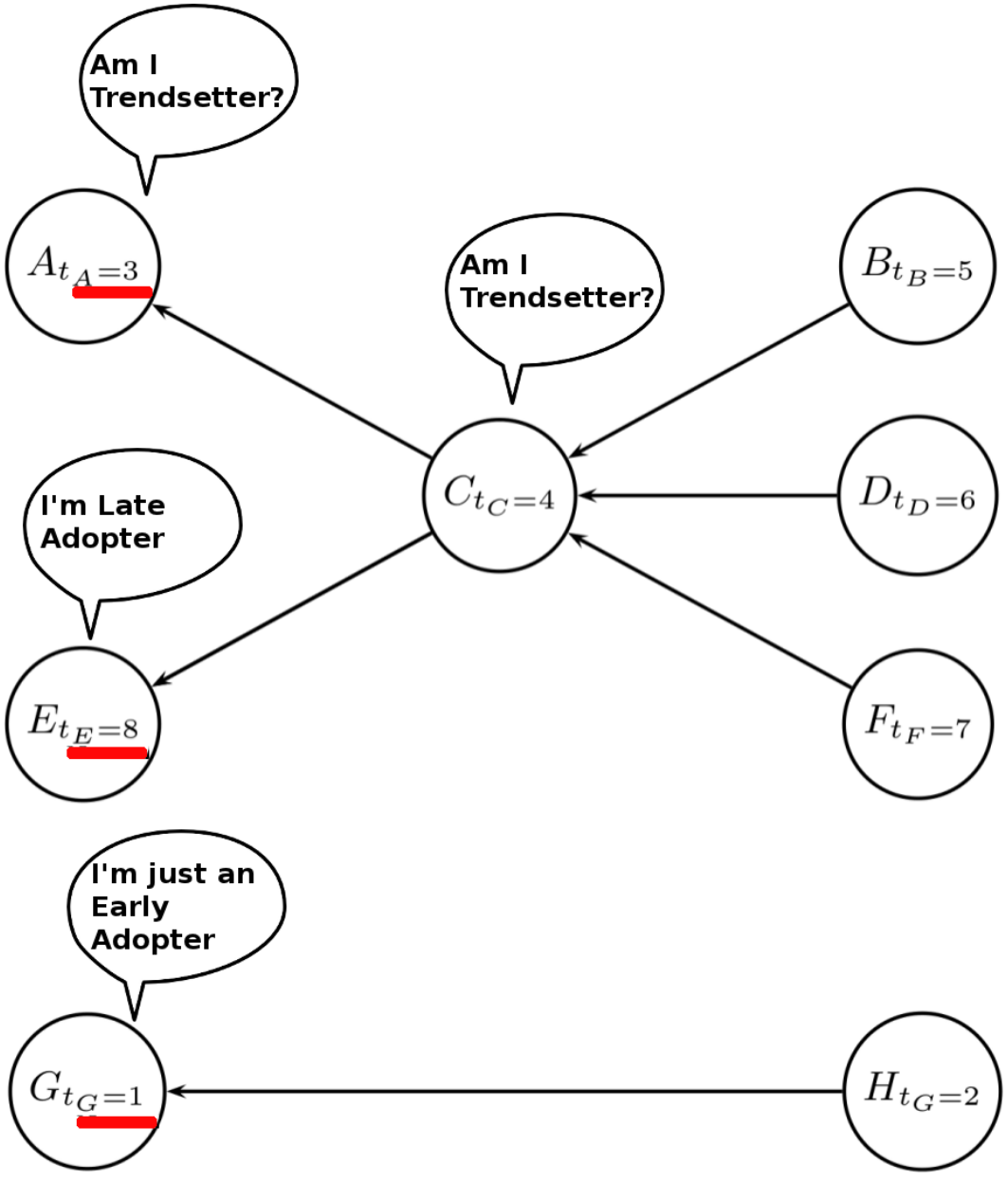
TIME!



Time



Time


Am I Trendsetter?
$A_{t_A=3}$
Am I Trendsetter?
$B_{t_B=5}$
$C_{t_C=4}$
$D_{t_D=6}$
I'm Late Adopter
$E_{t_E=8}$
$F_{t_F=7}$
I'm just an Early Adopter
$G_{t_G=1}$
$H_{t_G=2}$

How to find Trendsetters?

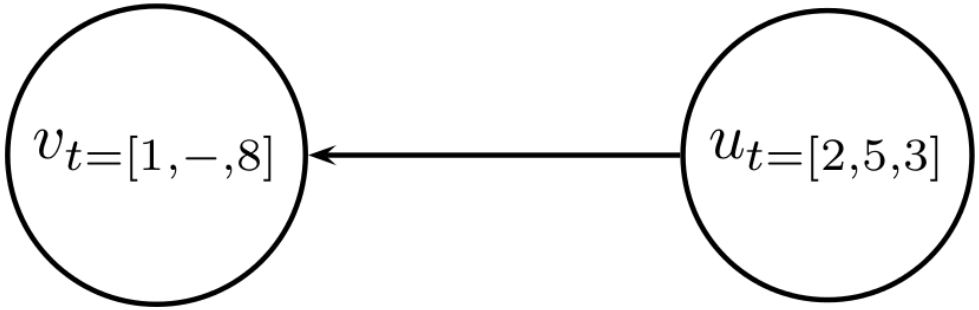
Weight edges and run PageRank

Topics and Influence Model

Topics

- Topic: collection of trends (Urls, memes, #hashtags, quotes, etc)
- We denote this set by a vector: $\{h_1 , . . . , h_n\}$.
- For each node we store the timestamp when she adopt a trend h_1

Weight Edges

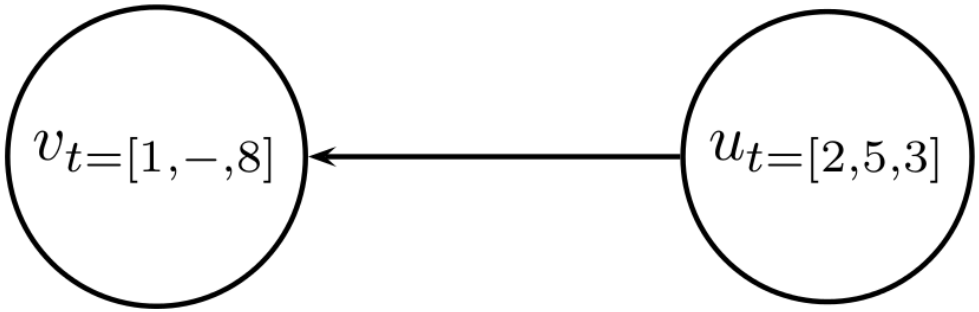


$$s_1(v)_i = \begin{cases} 1, & \text{if } t_i(v) > 0, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$s_2(u,v)_i = \begin{cases} e^{-\frac{\Delta}{\alpha}}, & \text{if } t_i(v) > 0 \text{ and } t_i(v) < t_i(u), \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

for $i = 1, \ldots, n_k$, where $\Delta = t_i(u) - t_i(v)$ and $\alpha > 0$.

Weight Edges



$$s_1(v) = [1, 0, 1] \qquad s_2(u, v) = [e^{-\frac{4-1}{\alpha}}, 0, 0]$$

- Next,

$$I_k^*(u, v) = \left(\frac{s_1(v) \cdot s_2(u, v)}{||s_1(v)|| \times ||s_2(u, v)||} \right) \times \left(\frac{L(s_2(u, v))}{n_k} \right), \qquad (3)$$

TS Ranking

Definition

The trendsetters (TS) rank of node v in a network $G_k(N_k, E_k)$, denoted by $TS_k(v)$, is given by:

$$TS_k(v) = d\ D_k(v) + (1 - d) \sum_{w \in In_{G_k}(v)} TS_k(w) I_k(w, v), \qquad (4)$$

Evaluation

Baseline

- In-degree ranking
- PageRank

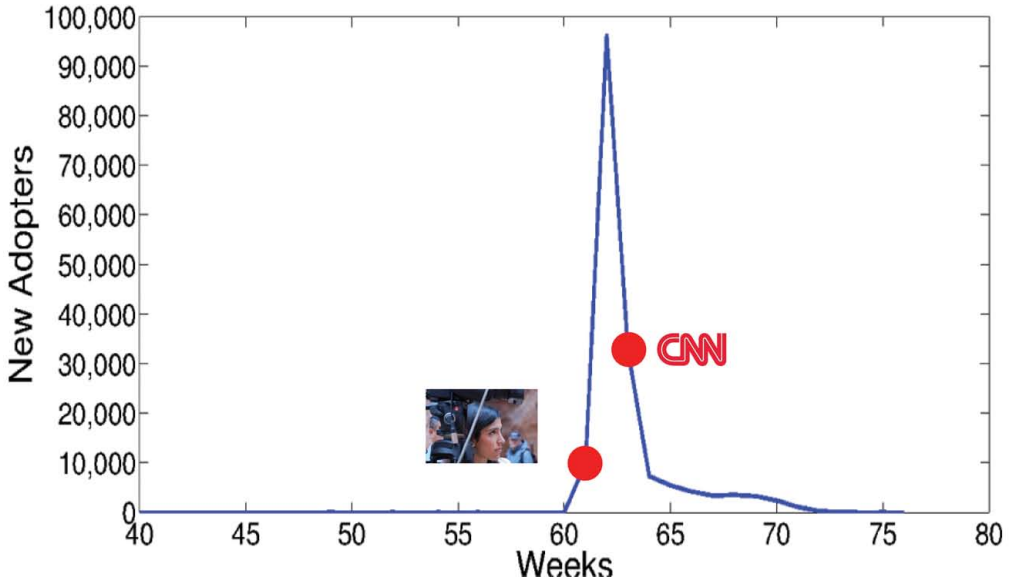
Dataset

- Twitter until August 2009.
- Over 50 Millions users with all their *followers* and *followees*.
- 1.6 Billions tweets
- We use #tags as trends.

Example:
Iran Elections on Twitter

Example

- Iran Elections: {#iran, #iranelections,#tehran}



- TS : @Lara (“Reporting from the Middle East”)
- PR : @cnnbr (“CNN Breaking News”)

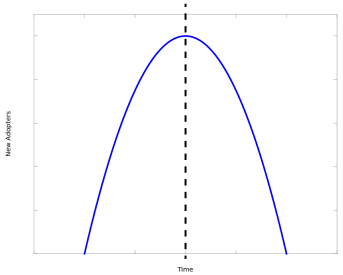
Baseline

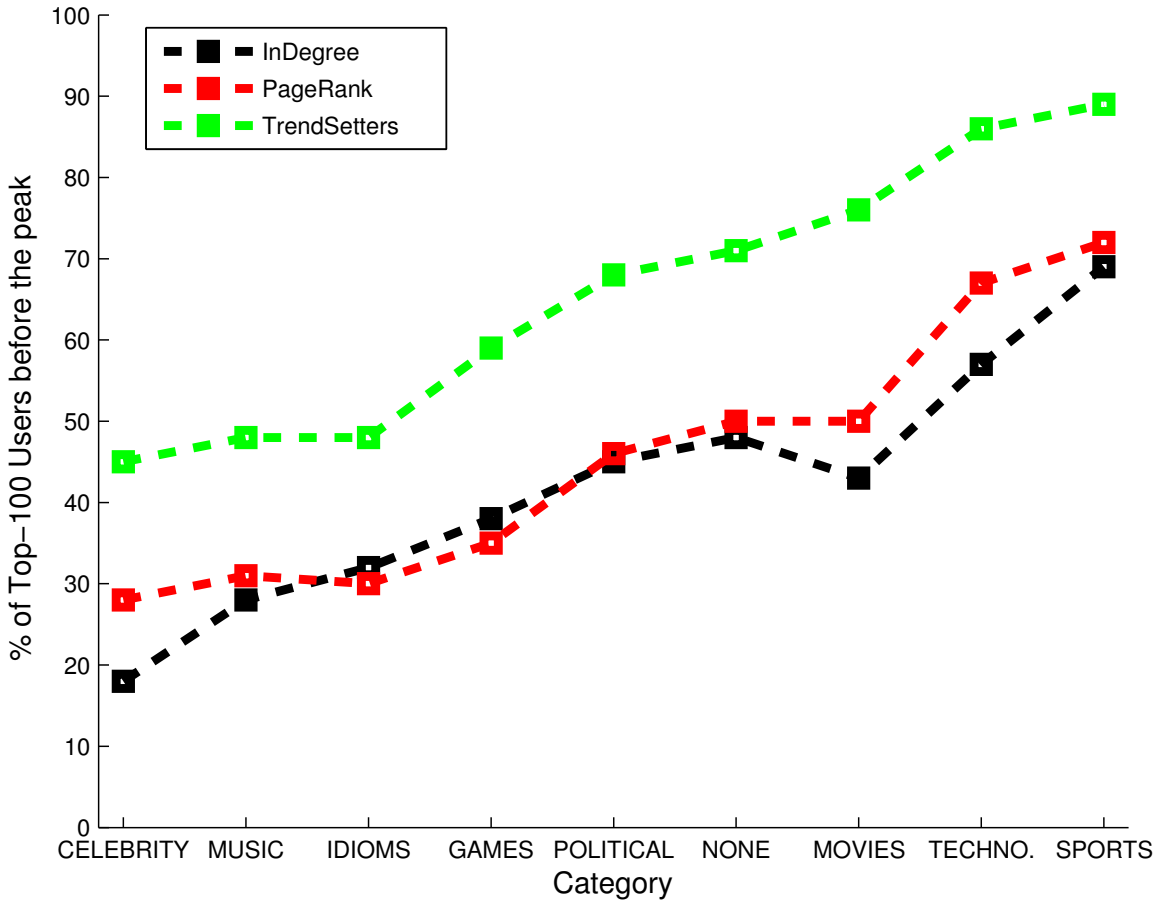
- We use the #tag classification made by Romero *et al.*

Category	#Topics	Example of Hashtags	#Tweets
Celebrity	16	#michaeljackson, #niley	1,036,101
Games	13	#mafiawars, #ps3 #	2,556,437
Idioms	35	#musicmonday,#followfriday	7,882,209
Movies	29	#heroes,#tv	1,769,945
Music	33	#lastfm, #musicmonday	2,785,522
None	153	#quotes, #sale	2,227,971
Political	39	#honduras, #Iranelection,	8,156,786
Sports	27	#soccer, #rugby	1,914,061
Technology	41	#twitter,#android	7,459,471

Trendsetters: early adopters?

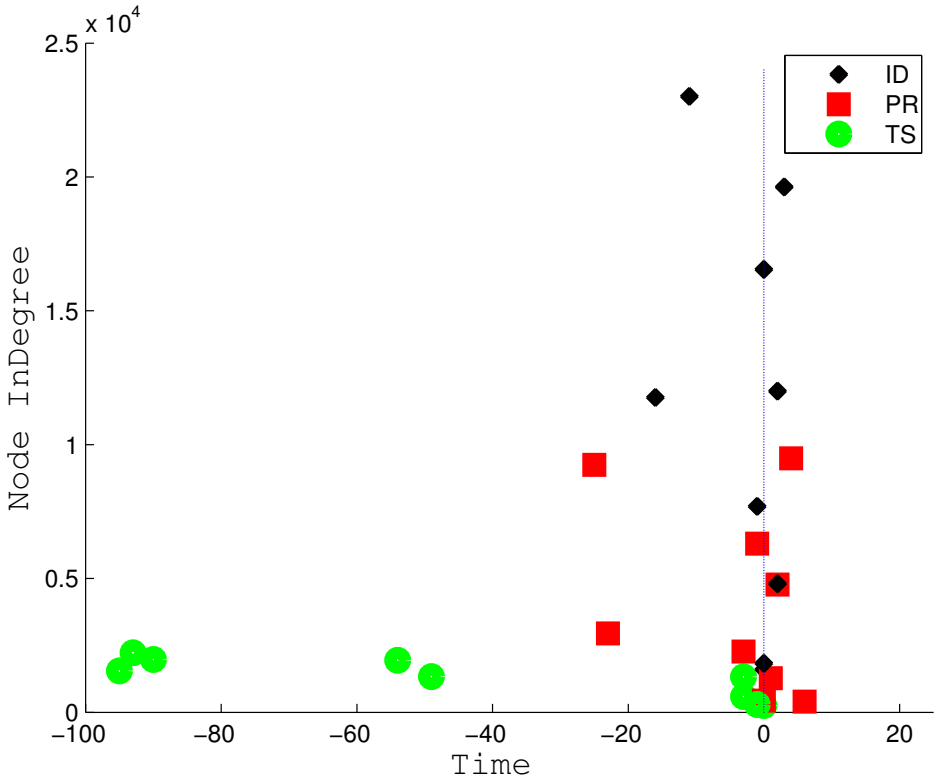
Experiments I





In-degree vs adoption time

Experiments II



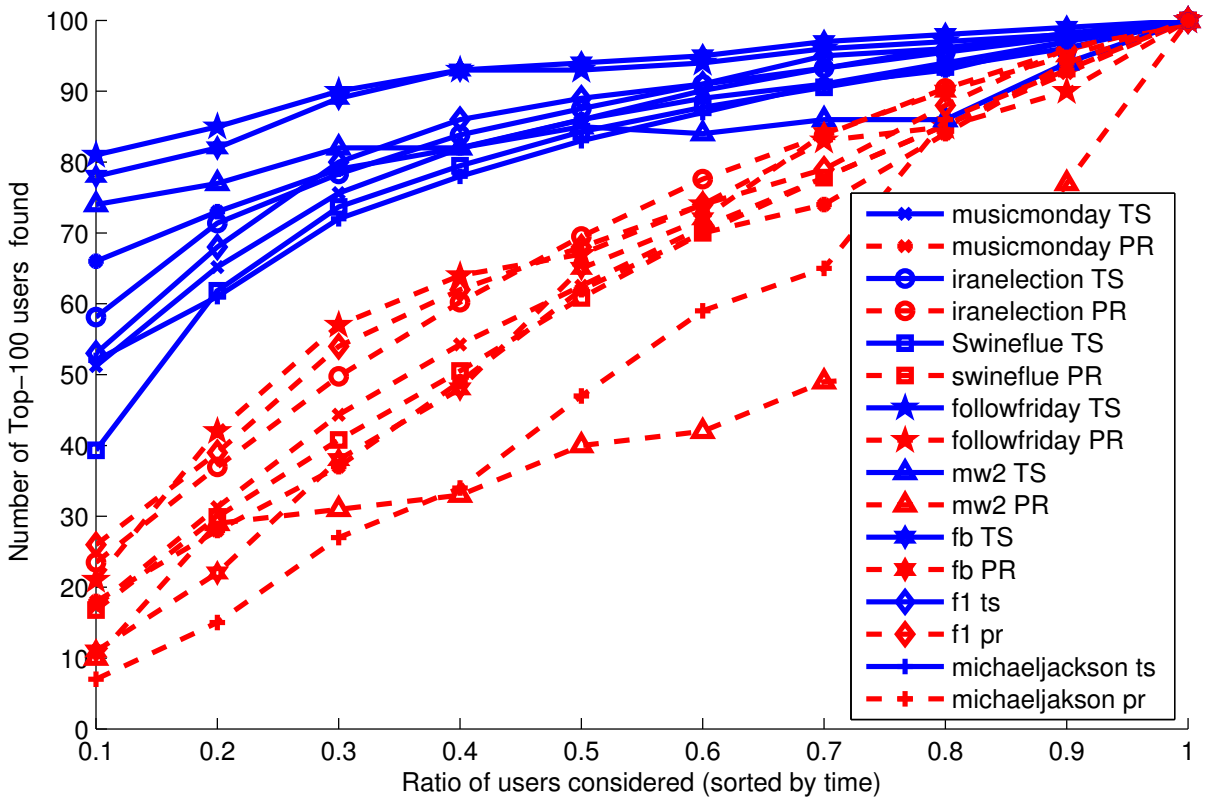
Influenced Followers Ratio

Influenced Followers Ratio

$IF_k(v)$ is the fraction of followers of v that adopted at least one trend of the topic k after v.

Category	(%)ID	(%)PR	(%)TS
POLITICAL	0.013	0.084	**0.174**
CELEBRITY	0.015	0.089	**0.148**
MUSIC	0.013	0.096	**0.160**
GAMES	0.022	0.058	**0.115**
SPORTS	0.004	0.054	**0.098**
IDIOMS	0.001	0.034	**0.088**
NONE	0.011	0.001	**0.085**
TECHNOLOGY	0.006	0.054	**0.078**
MOVIES	0.006	0.043	**0.067**

Ranking with Partial Information



Final Remarks

- Usually, follower hubs (celebrities) are **late adopters**.
- Trendsetters have lower in-degree, but they spread new ideas.

Code

- Code available (diego.saez [at] upf.edu)
- Implementation in python + C

Questions

Questions

Questions

We also group the topics by shape, using the KSC algorithm.