

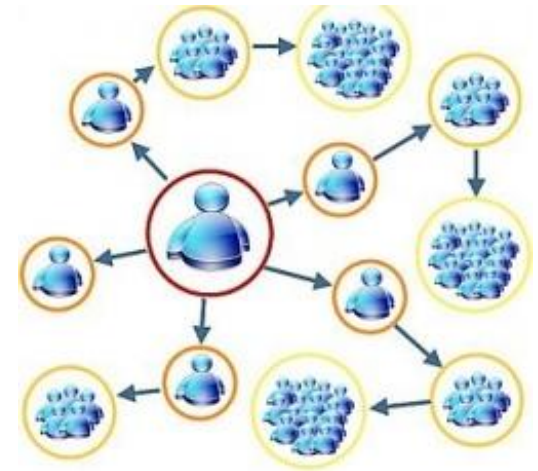
Influence and Passivity in Social Media

Daniel M. Romero, Wojciech Galuba,
Sitaram Asur, and Bernardo A. Huberman

Social Computing Lab, HP Labs

Introduction

- Explosive growth of social media: Facebook, Twitter, Digg, etc
- Anyone has the potential of spreading a message
- Competition for attention
- Sources rely on the crowd to spread their message and make it popular
- Who is influential?



Outline

1. Twitter background and data
2. Description of the algorithm to assign an influence score to each user on Twitter
3. Algorithm evaluation through popularity prediction
4. Discussion of possible extensions and applications of the algorithm
5. Conclusion

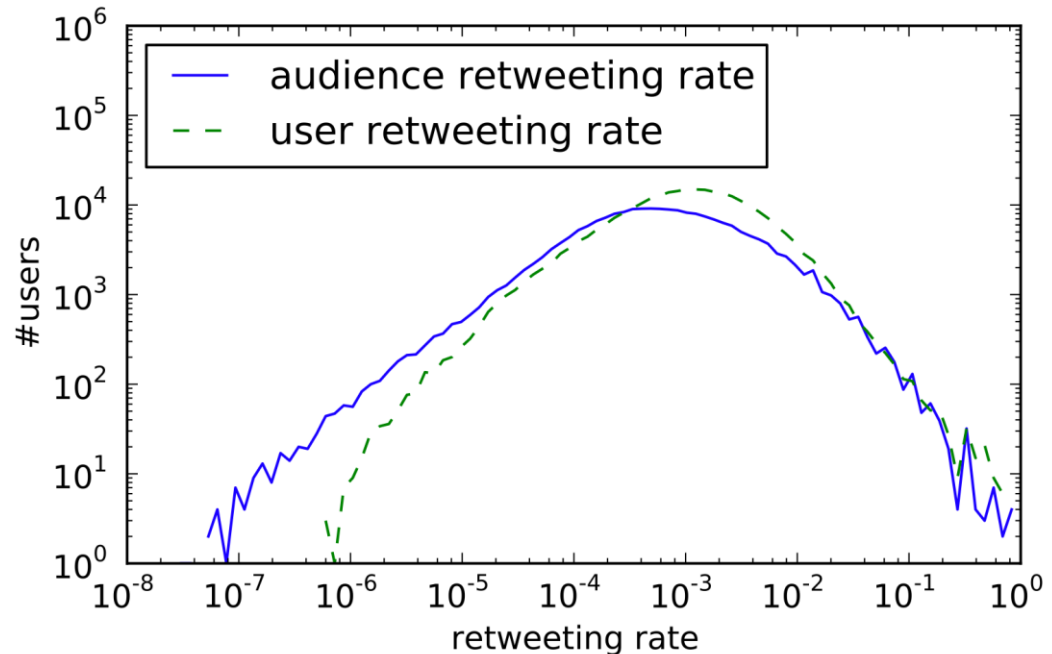
twitter Background

- Microblogging service
- 200 million users as of March 2011
- User *follow* other users
- Many business, celebrities, news organizations, and others use Twitter
- *Tweet*: A short post of at most 140 characters
- *Retweet*: A tweet originally posted by a user and reposted by another user

The Data

- 300 hours starting on Sept. 10th 2009
- All tweets containing a URL
- 22 million tweets (1/15th of all)
- 15 million unique URLs
- 2.5 million users posted at least one tweet containing a URL

Passivity



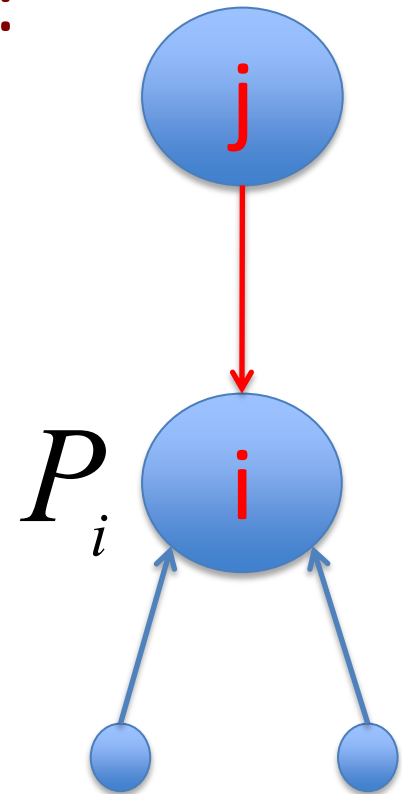
- Twitter users rarely pass information forward to others
- Retweeting rates vary among users but it is very low overall
- On average, users retweet only 1 in 318 URLs they are exposed to

Influence Score Assumptions

A user's influence score depends on:

1. The **passivity** of the those she influences.

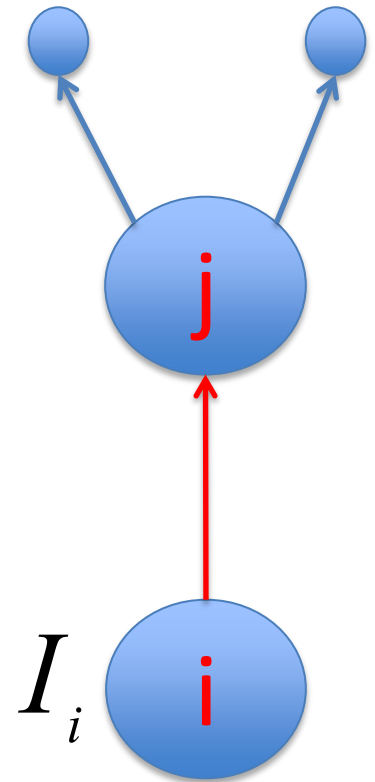
2. How much they **accept** her influence compared to everyone else's



Passivity Score Assumptions

A user's passivity score depends on:

1. The **influence** of those who she's exposed to.
2. How much she **rejects** their influence compared to everyone else.



IP Algorithm (Input)

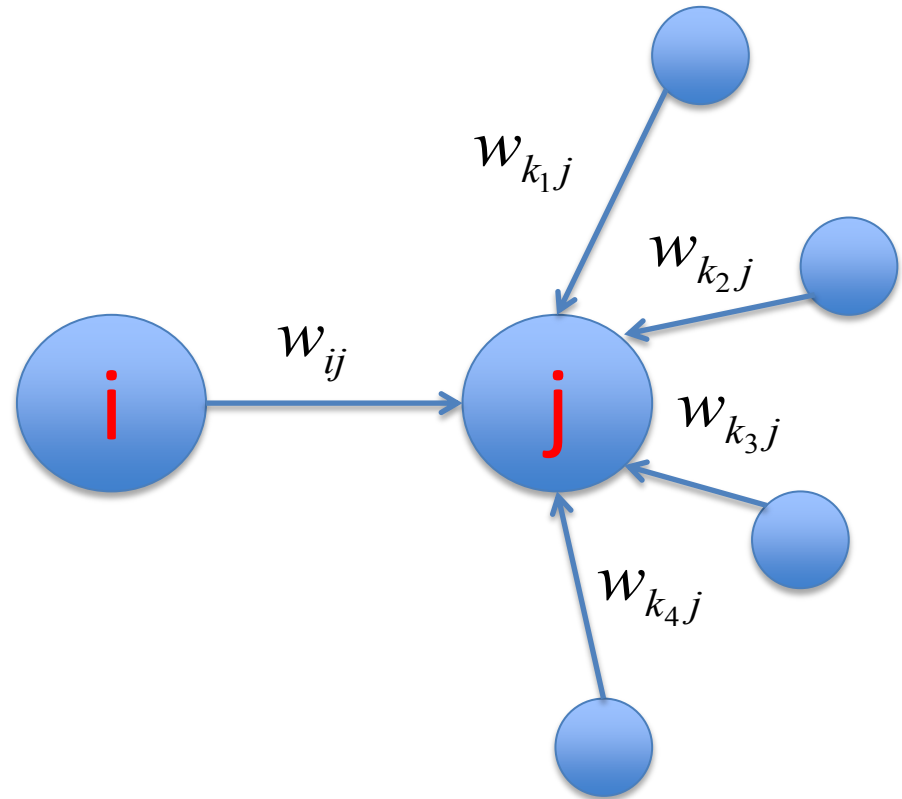
- A weighted influence graph $G = (N, E, W)$
- Nodes N : Users who retweeted or were retweeted at least once
- Edges E : The edge (i, j) exists if user i retweeted j at least once
- Weights W : The weight of edge (i, j) is the ratio of tweets by user j retweeted by i .



Acceptance Rate

For each edge (i,j) , compute
acceptance rate

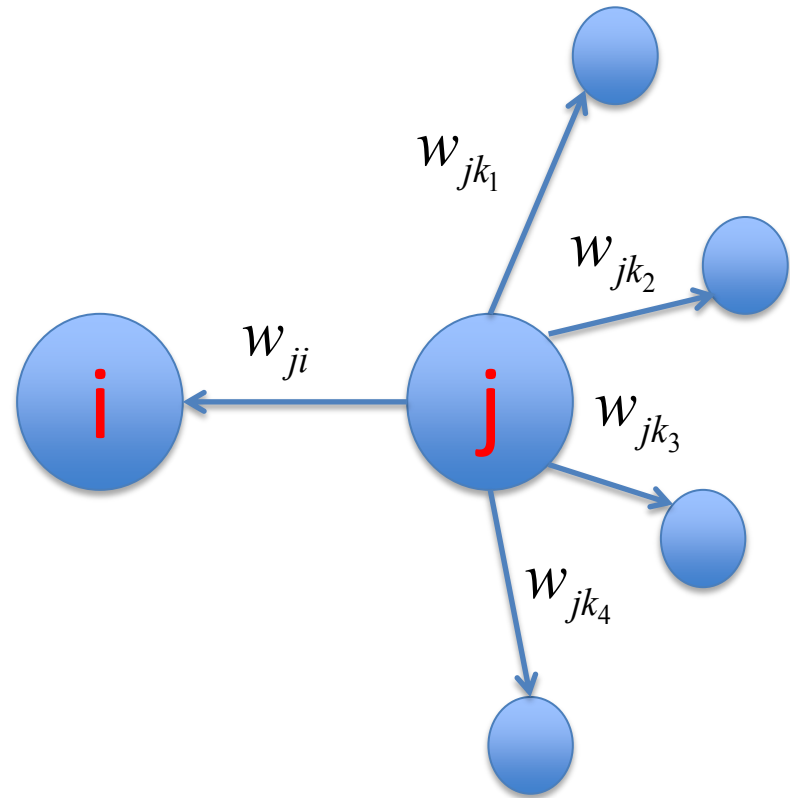
$$a_{ij} = \frac{w_{ij}}{\sum_{k:(k,j) \in E} w_{kj}}$$



Rejection Rate

For each edge (j,i) , compute
rejection rate

$$r_{ji} = \frac{1 - w_{ji}}{\sum_{k:(j,k) \in E} (1 - w_{jk})}$$

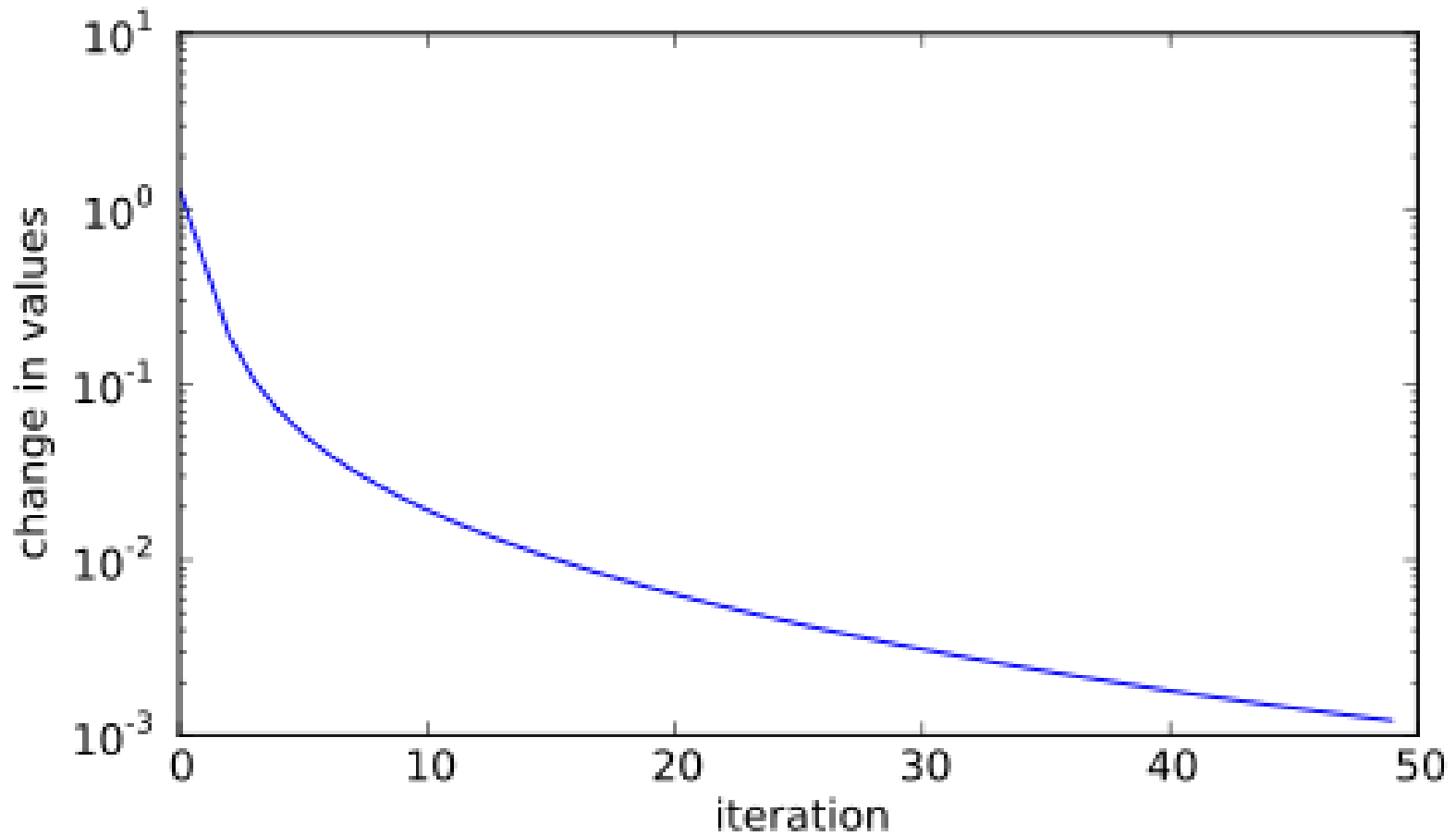


IP Algorithm Operations

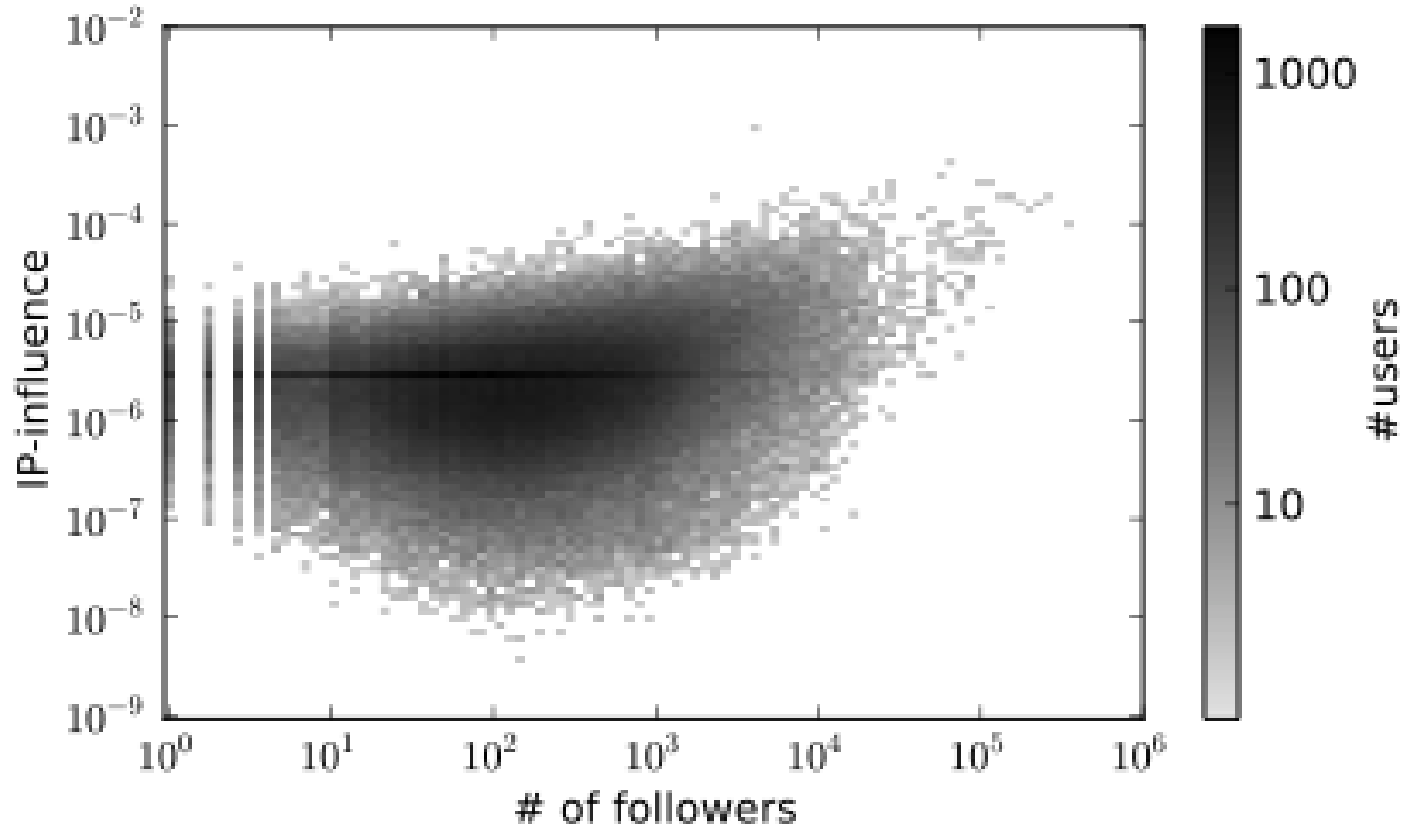
$$P_i \leftarrow \sum_{j:(j,i) \in E} r_{ji} I_j$$

$$I_i \leftarrow \sum_{j:(i,j) \in E} a_{ij} P_j$$

Algorithm Convergence



Influence VS. Popularity



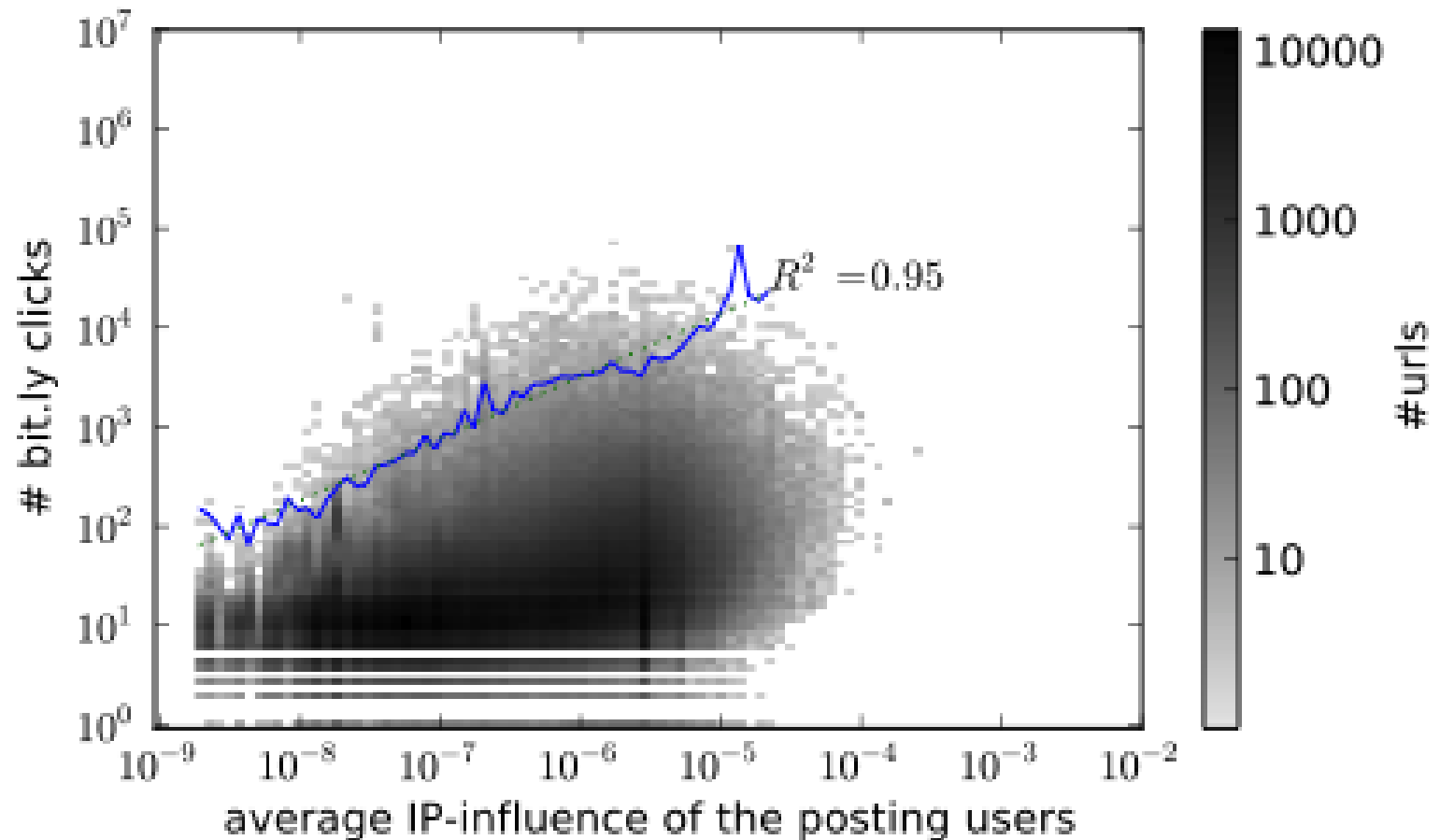
$$R^2 = .44$$

Evaluation

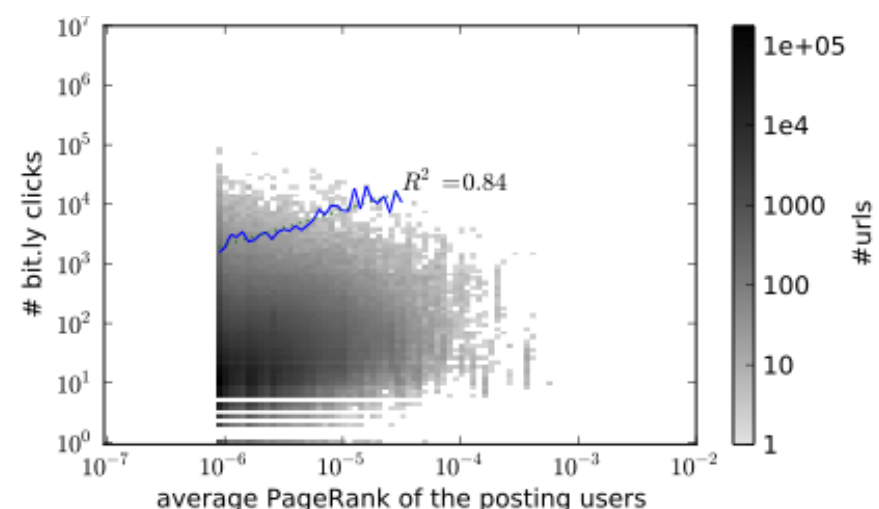
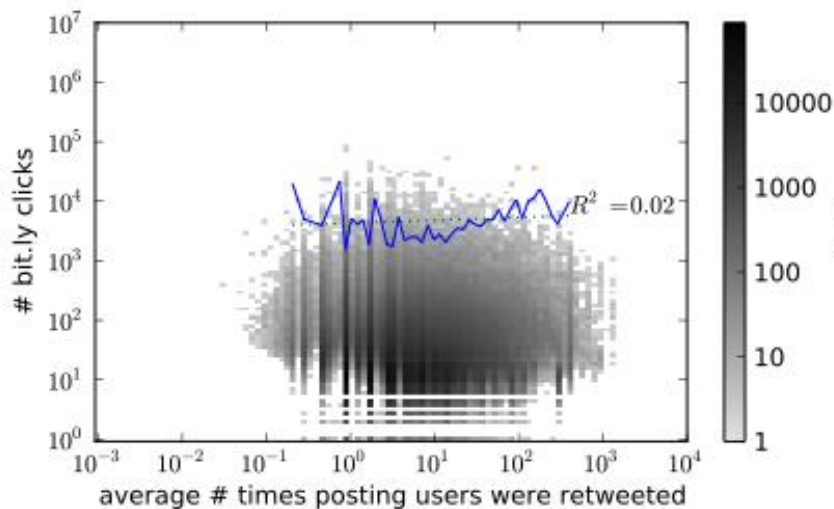
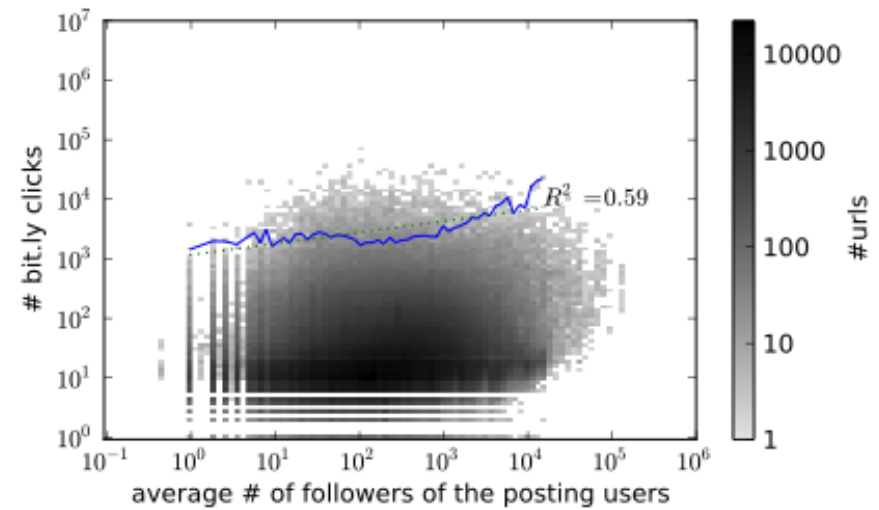
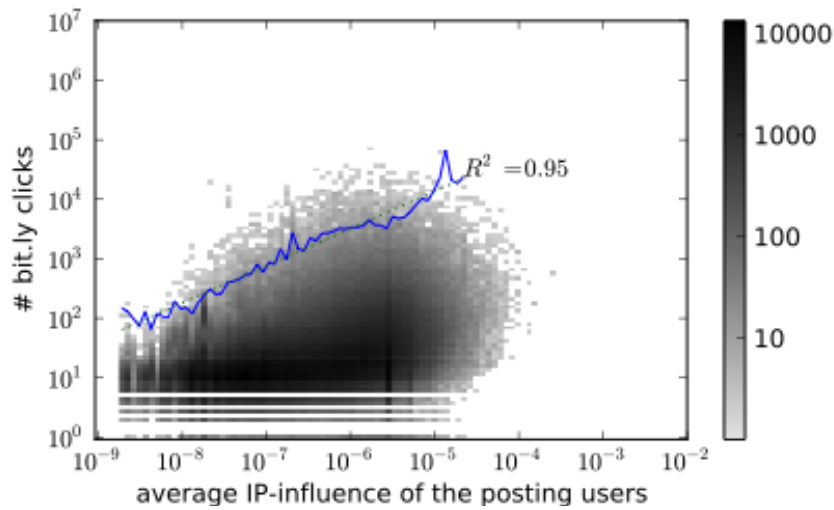


- No ground truth for influence
- We used URL traffic as an evaluation tool
- bit.ly is a URL shortening service that keeps track of URL traffic
- 3.2 million bit.ly URLs in our data set
- For each URL we find the average IP-influence of the users that mentioned it

IP-Influence VS. URL Traffic Upper Bound



Comparison to Other Measures of Influence



Low Influence High Popularity

Many Followers \Rightarrow Influence

User name	Category	Rank by # followers	Rank by IP-influence
thatkevinsmith	Screen Writer	33	1000
nprpolitics	Political News	41	525
eonline	TV Channel	42	1008
marthastewart	Television Host	43	1169
nba	Sports	64	1041
davidgregory	Journalist	106	3630
nfl	Sports	110	2244
cbsnews	News Channel	114	2278
jdickerson	Journalist	147	4408
newsweek	News Magazine	148	756

Table 3: Users with many followers and low relative influence

Most Passive Users

Who are the most passive users?

Spammers

redscarebot	Keyword Aggregator
drunk_bot	Suspended
tea_robot	Keyword Aggregator
condos	Listing Aggregator
wootboot	Suspended
raybeckerman	Attorney
hashphotography	Keyword Aggregator
charlieandsandy	Suspended
ms_defy	Suspended
rpattinsonbot	Keyword Aggregator

Table 2: Users with the most IP-passivity

Conclusion

- Passivity in social media is hard but necessary to overcome to be influential
- Weak correlation between influence and popularity
- IP-influence can be a good predictor on upper bound of URL traffic

Future Work

- Content specific ranking
- Content ranking
- Extensions to other online social networks
- Application