An empirical approach towards a whom-to-mention Twitter app

Soumajit Pramanik\textsuperscript{1}, \textbf{Maximilien Danisch}\textsuperscript{2}, Qinna Wang\textsuperscript{2}
Jean-Loup Guillaume\textsuperscript{3}, Bivas Mitra\textsuperscript{1}

1- CNeRG, IIT Kharagpur, India
2- Complex Networks team, LIP6, UPMC-CNRS, Paris, France
3- L3I, University of La Rochelle

Supported by:
The French National Research Agency - contract CODDDE ANR-13-CORD-0017-01 and ECIS - Cefipra Indo-French Targeted Program
This is joint work with:

Soumajit Pramanik (IIT), Maximilien Danisch (LIP6), Dr. Qinna Wang (LIP6)

Prof. Jean-Loup Guillaume (ULR)

Prof. Bivas Mitra (IIT)
What is Twitter?

“Twitter is a social microblogging system which has become one of the most important ways for sharing information online.”

What kind of information?

- **very local**: I’ll have a coffee at our bakery in 10min
- **local**: IITkgp party on Friday night, all invited
- **global**: Nicolas Sarkozy is back to politics
- **very global**: There has been a terrorist attack in Paris
Retweets = propagation

Where do retweets come from?

Someone has to see your tweet:

- One of your followers
- One of the followers of someone that retweeted your tweet
- Someone you mentioned @
- Someone who saw your tweet searching for keywords or #
- Someone who saw it somewhere else on the Internet
Follow network
You can do something with @
Our goal

How to share an information/opinion with the greatest number of interested people?

→ use Twitter + use mention = use our App

We target:

1. (very) global tweets and
2. (very) local tweets
Outline

1. Related work on mention recommendation
2. Data study to get insights
3. Online application
4. Experiments “faky app”
5. Conclusion & perspectives
Three related papers

1- Whom to mention: expand the diffusion of tweets by @ recommendation on micro-blogging systems.
Wang et al. WWW2013

**PROS**

- First in-depth study of mention in microblogs
- Extract features (interest match, user relationship and user influence)
- A ranking function is trained
- Address recommendation overload problem

**CONS**

- No real-time experiments
- No usable app provided
Three related papers

2- Locating targets through mention in Twitter.
Tang et al. World Wide Web 2014

PROS

- Extract four categories of features
  (content, social, location and time)

CONS

- The goal is to mention someone interested and trigger interactions (mentions, replies, retweets, friendship) but not to maximize the spreading
- No real-time experiments
- No usable app provided
Three related papers

3- Who will retweet this? Automatically identifying and engaging strangers on Twitter to spread information. Lee et al. IUI2014

**PROS**

- More features extracted for machine learning methods and relevance is evaluated
- Waiting-time model
- Real-time experiments

**CONS**

- Goal is to mention someone so that he retweets, not to maximize the spreading
- Not real users (“Public Safety News” and “Bird Flu News”)
- No usable app provided
What we suggest

Make a whom-to-mention recommendation system

- to maximize the spread of a tweet,
- working only on real time experiments,
- with a usable app,
- and improve the app every time it is used.
Outline

1. Related work on mention recommendation
2. Data study to get insights
3. Online application
4. Experiments “faky app”
5. Conclusion & perspectives
World cup dataset

The dataset consists of all tweets made during May, June or July 2014 and containing hashtags specific to the 2014 soccer world cup.

27,916,100 tweets made by 6,020,228 unique users.

- 15,249,762 retweets (55%)
- 937,201 answers (3%)
- 11,729,137 original tweets (42%)
Data study

![Graph showing the relationship between the number of mentions and the average number of retweets. The graph depicts a peak around 3 mentions, with fluctuations around 10 and 14 mentions.]
Data study
Data study

![Graph showing the relationship between the number of followers of a user and the retweet probability times the number of followers. The x-axis represents the number of followers of the mentioned user, ranging from $10^0$ to $10^7$, and the y-axis represents the retweet probability times the number of followers, ranging from $10^{-2}$ to $10^4$. The graph indicates a positive correlation between the number of followers and the retweet probability times followers.]
Outline

1. Related work on mention recommendation
2. Data study to get insights
3. Online application
4. Experiments “faky app”
5. Conclusion & perspectives
Online application

Users’ database?

- (very) global tweets $\rightarrow$ use a database of “random” users
- (very) local tweets $\rightarrow$ target non reciprocal followees
Online application

Target users that

- are popular: $f_P(u)$
- are susceptible to retweet the tweet: $f_{RT}(u, t)$

$\rightarrow$ i.e. have a high $S(u, t) = f_P(u).f_{RT}(u, t)$

Approximation

Neglect overlap (as we mention several users) and maximize the number of people that will see the tweet at hop one.

- $\sum_u S(u, t)$
- $f_P(u) =$ number of followers of $u$
- Learn $f_{RT}(u, t)$ using logistic regression
Logistic regression

Principle:
Regression (or binary classification of vectors) by fitting parameters of a logistic function: $V_\Theta(X) = \frac{1}{1+\exp(\Theta^T X)}$ to optimize a convex quality function.

Advantages:
- Returns the probability for a user to retweet;
- A few parameters to store: convenient for an online app.
Map into the knapsack problem

The goal is to use the remaining place as efficiently as possible:

• Solve the knapsack problem with:
  • total budget: $140 - L_{\text{tweet}}$
  • user weight: $2 + L_{\text{screenname}}$
  • user value: $S(u, t)$
Improving the app

Every time the app is used we

1. collect users to increase the database
2. check which mentioned user retweets and improve $f_{RT}(u, t)$
Live Demo

Mention Recommendation System

Username
Soumajit
Tweet
Loved the movie #HappyNewYear starring Shah Rukh Khan and Deepika Padukone
Keywords
Movie, Happy New Year, Shah Rukh, Deepka

Get!

Node | Rank
--- | ---
deevikapadukone | 9.90256613445
iamsrk | 9.31518241019
SrBachchan | 9.06844864743
arrahman | 8.49729229928
sonunigam | 8.182208032923
Outline

1. Related work on mention recommendation
2. Data study to get insights
3. Online application
4. Experiments “faky app”
5. Conclusion & perspectives
Current work: faky app

- Select tweets containing mention(s) made recently on Twitter by random users and retweet them from our fake account replacing the mention(s). → Try to have more retweets!
- We chose the tweets based on keywords and a manual selection in order to have only global tweets.
- 50 tweets lead to:
  - 4 retweets,
  - 2 followers,
  - 5 favourites,
  - 6 mentions,
  - 1 mentioned user blocked us.
Outline

1. Related work on mention recommendation
2. Data study to get insights
3. Online application
4. Experiments “faky app”
5. Conclusion & perspectives
Discussion

Impact of number of mentions

Is knapsack worth it?
Tweeting several times the same tweets mentioning different users

Extreme case

Do not use friendship... use mention! → Twitter2.0

Should not target “bad” users

Sybils (users creating many fake accounts)
Social capitalists (FMFY/IFYFM) http://www.bit.ly/DDPapp