
Information Spread on Twitter: How Does Mention Help?

Soumajit Pramanik, Ayan Das & Bivas Mitra

**Department of Computer Science & Engineering,
Indian Institute of Technology, Kharagpur, India**

**Dynamics On & Of Complex Networks (DOOCN)-VII
ECCS Satellite Workshop, 2014**

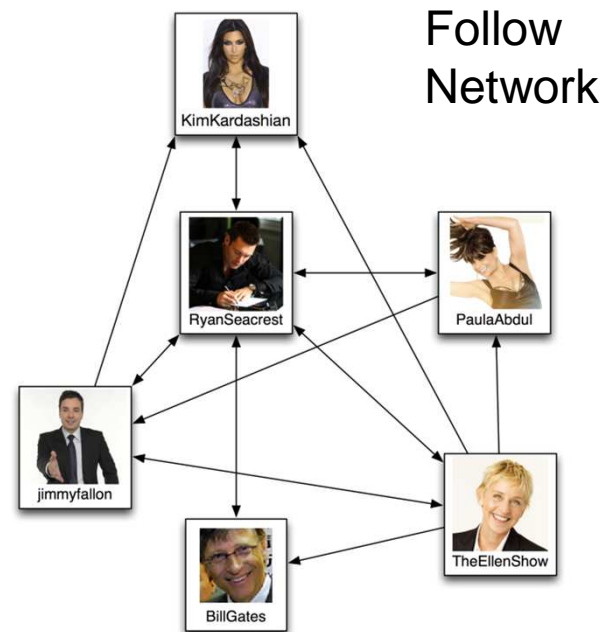


Introduction



■ Follow Links in Twitter

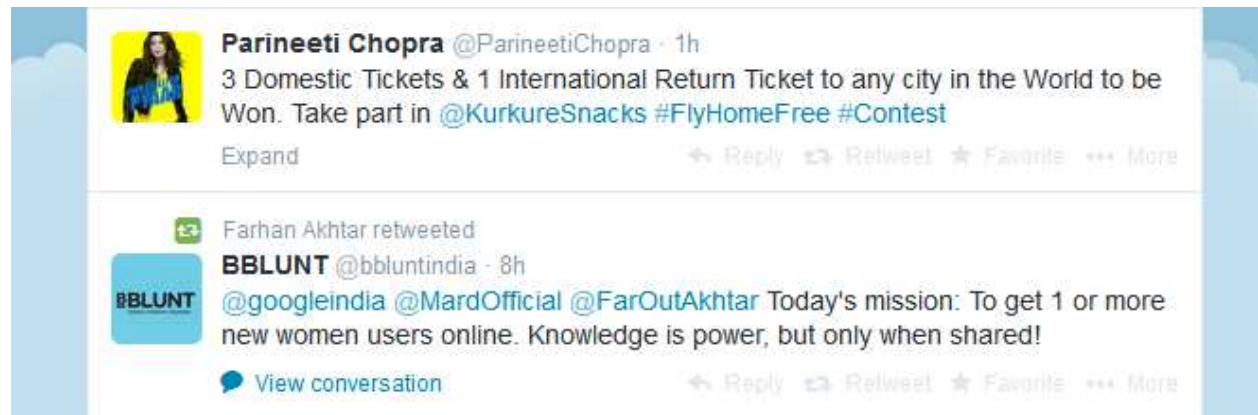
- A user can be followed by any number of users
- All tweets by the user are shown in the timeline of her followers.



Introduction (Continued)

■ Mention Links in Twitter

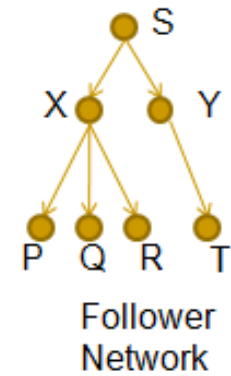
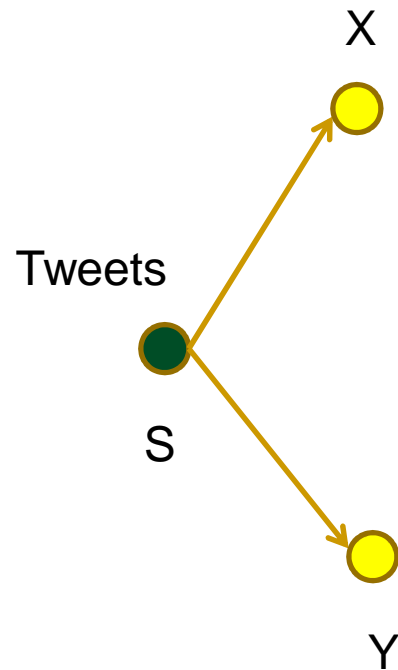
- ❑ A mention is any Twitter update that contains "@username" anywhere in the body of the Tweet.
- ❑ Non-Followers can also be mentioned



Introduction (Continued)

Information Diffusion via Follow Links

S has two followers X & Y

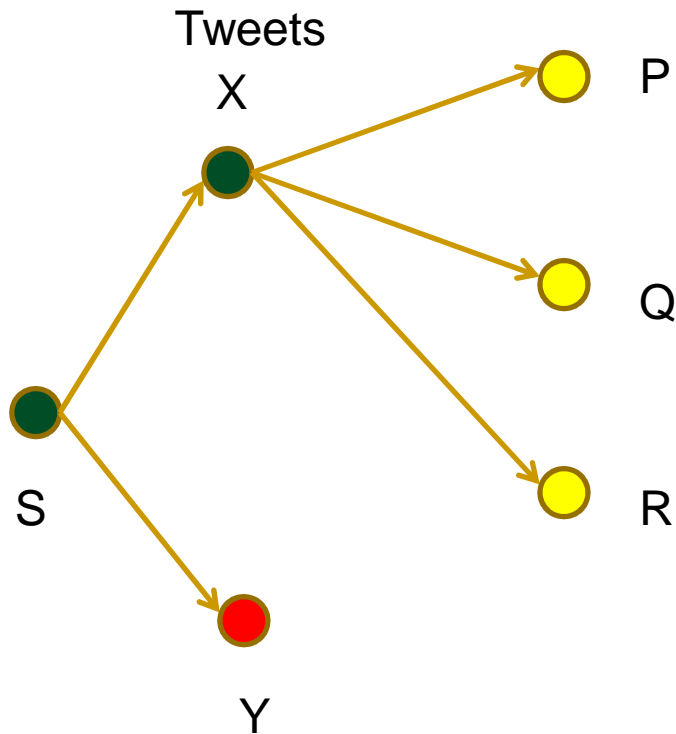


Hashtag is a unit of information in Twitter

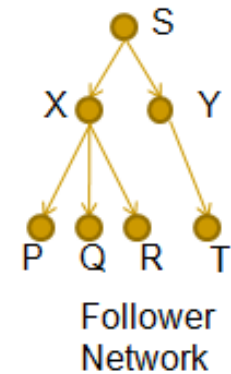
Introduction (Continued)

Information Diffusion via Follow Links

X has 3 followers- P, Q & R



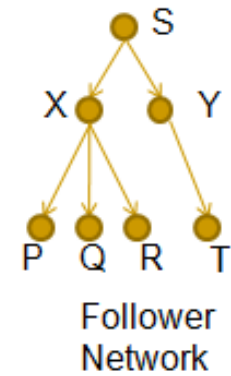
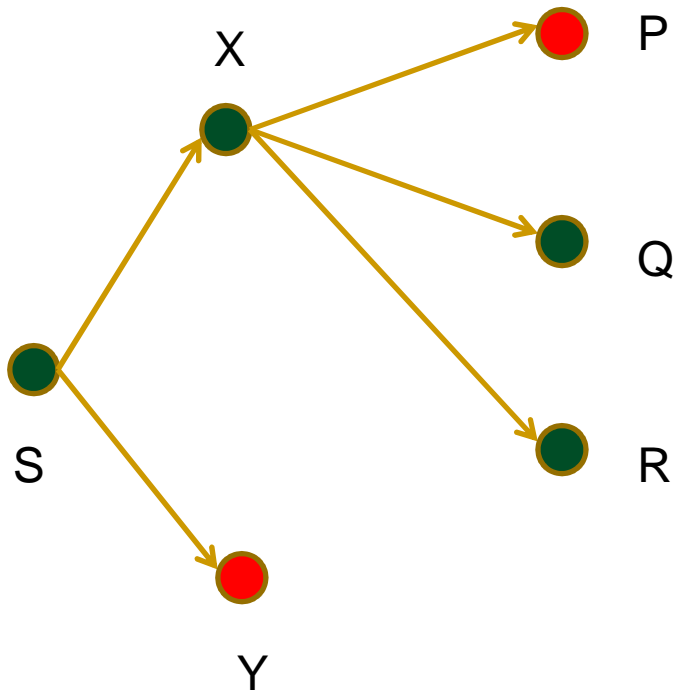
X decides to re-tweet
Y decides not to re-tweet



Introduction (Continued)

Information Diffusion via Follow Links

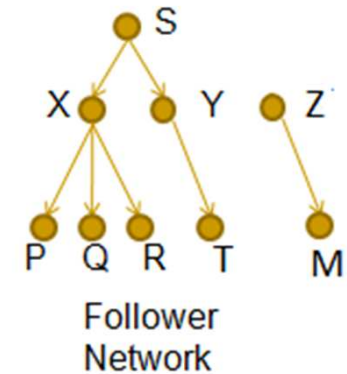
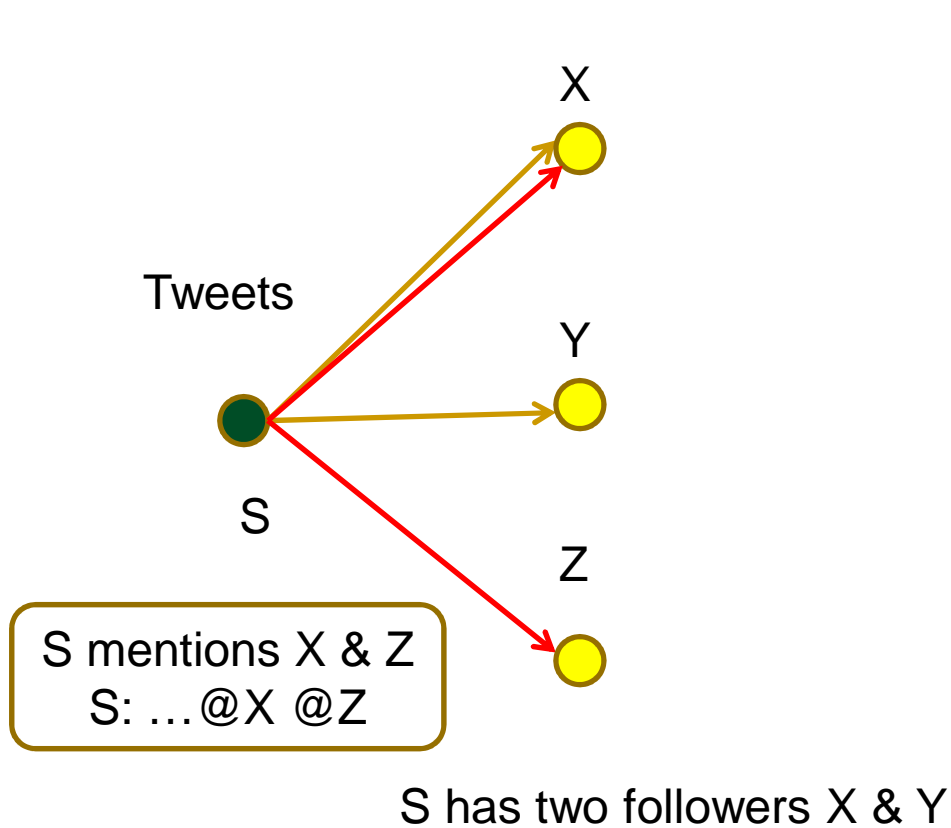
P decides not to re-tweet
Q & R decide to re-tweet



And in this way the information propagates....

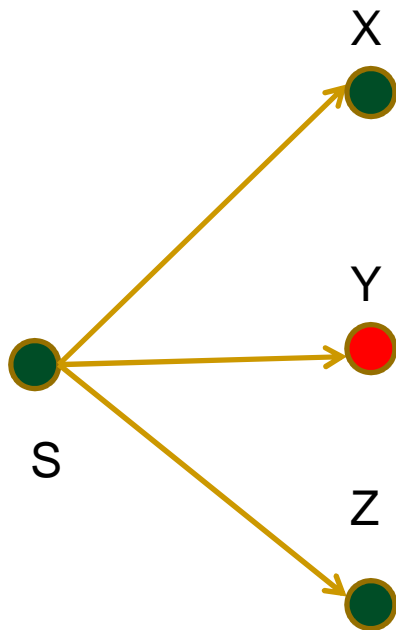
Introduction (Continued)

Information Diffusion via Follow & Mention Links

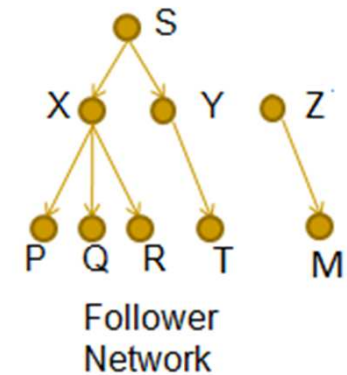


Introduction (Continued)

Information Diffusion via Follow & Mention Links

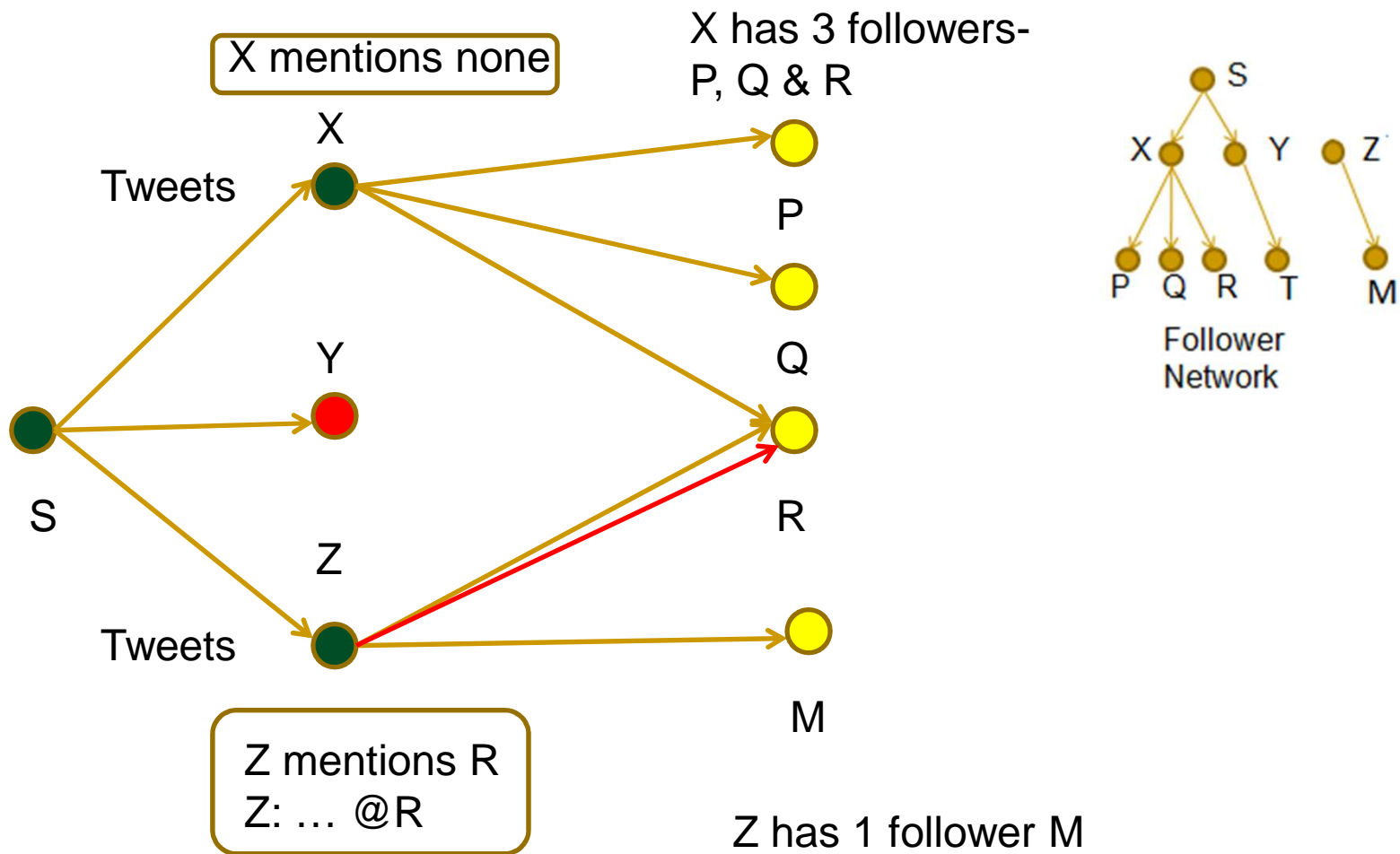


X, Z decide to re-tweet
Y decides not to



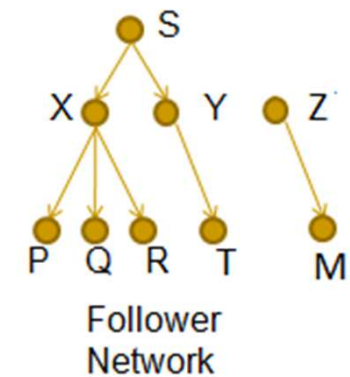
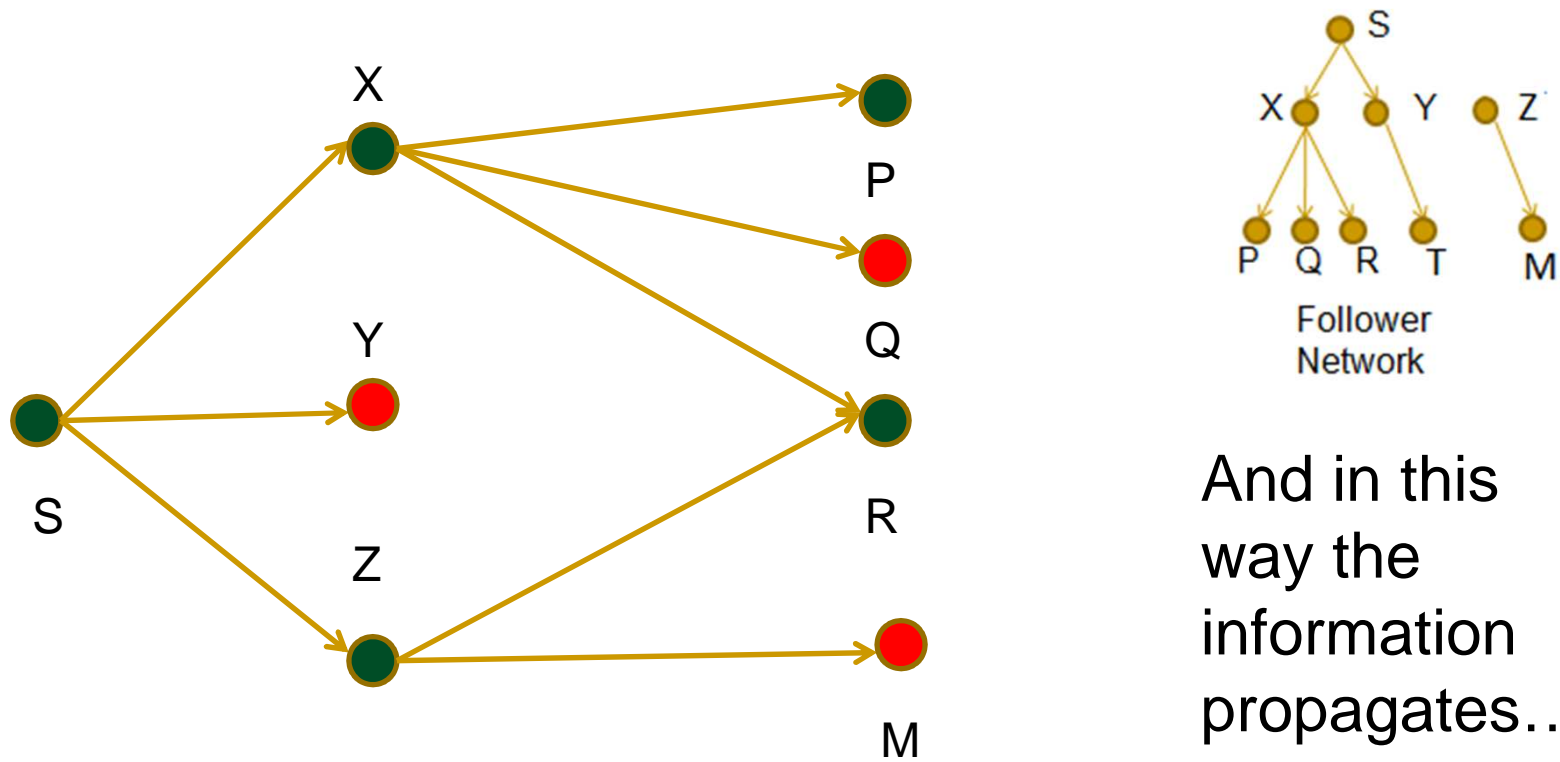
Introduction (Continued)

Information Diffusion via Follow & Mention Links



Introduction (Continued)

Information Diffusion via Follow & Mention Links

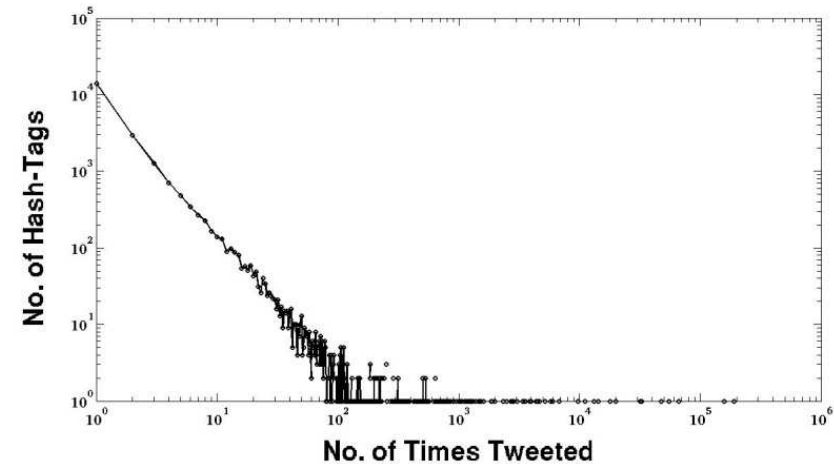
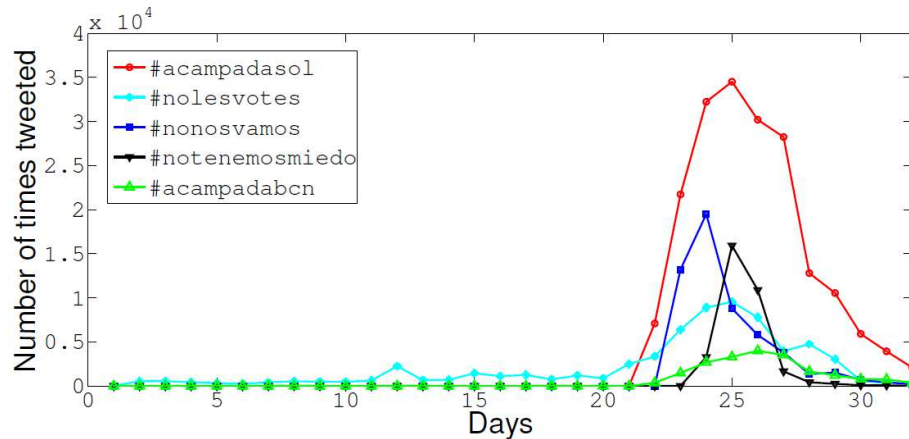


And in this way the information propagates...

P, R decide to propagate
Q, M decide not to propagate

Introduction (Continued)

Hashtag Popularity (Number of users tweeted)



(15m Dataset)

Observation:

1. Different hashtags have different temporal pattern of popularity
2. Few hash-tags are highly popular , but most are not

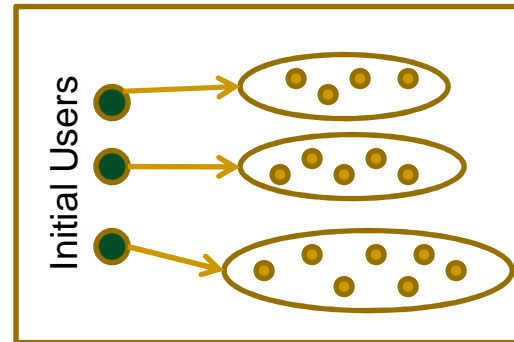
Research Question:

Investigate the Key-factors controlling the popularity of a hashtag

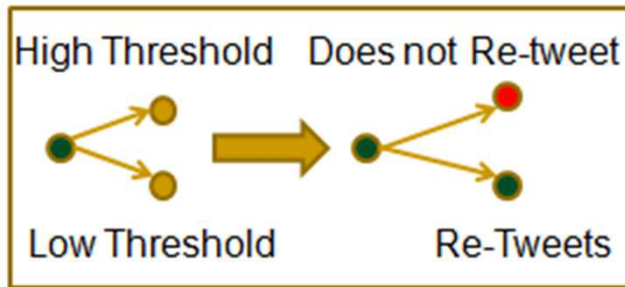
Introduction (Continued)

Factors Influencing Popularity of a hashtag

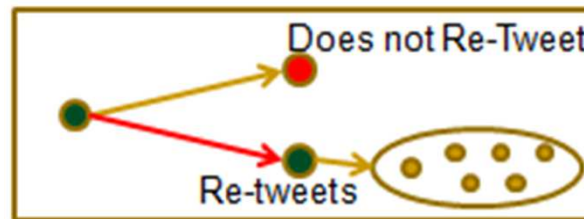
- Number of Initial Users (gets information from external sources)



- Passivity-Threshold of Users

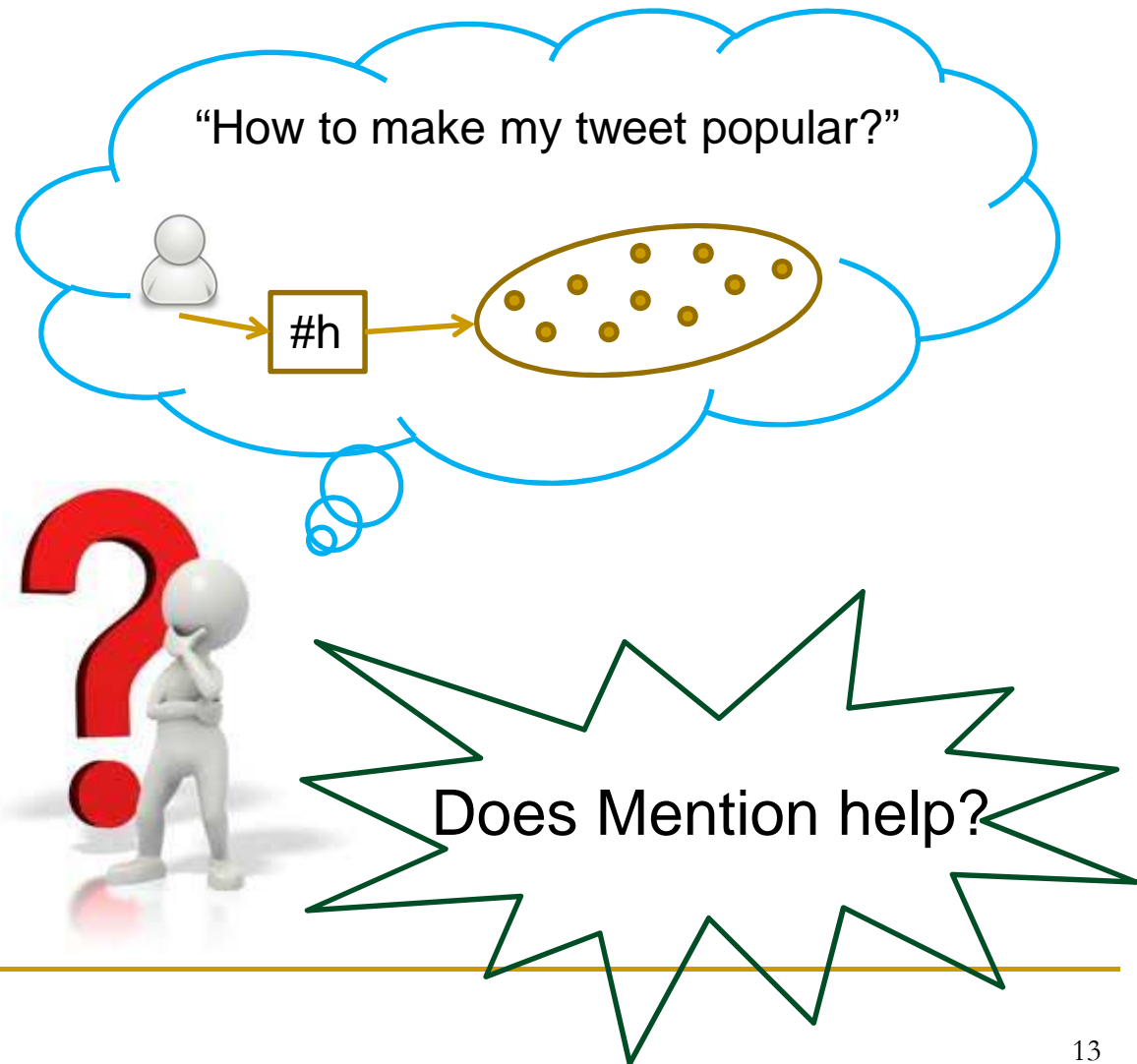


- Mention Usage



Problem Statement

How to increase the popularity of a Hashtag?



Problem Statement (Continued)

Available Options:

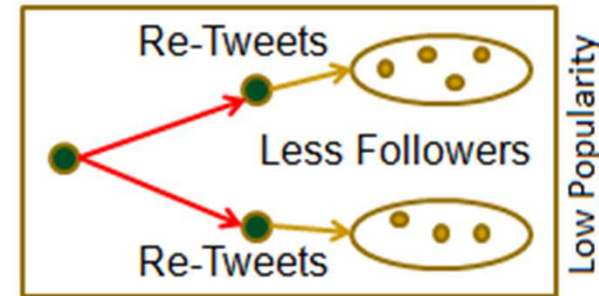
- Mention Friends:

- Pro:

- High probability of re-tweet

- Con:

- Low popularity if friends are not popular (less number of followers).



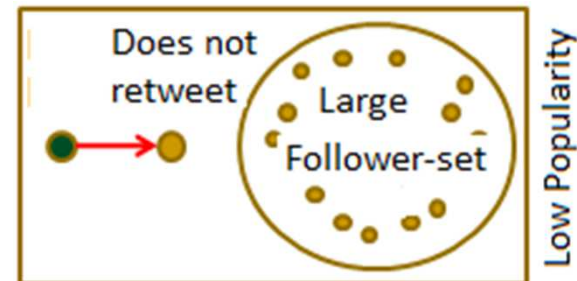
- Mention Celebrities

- Pro:

- High popularity if they re-tweet

- Con:

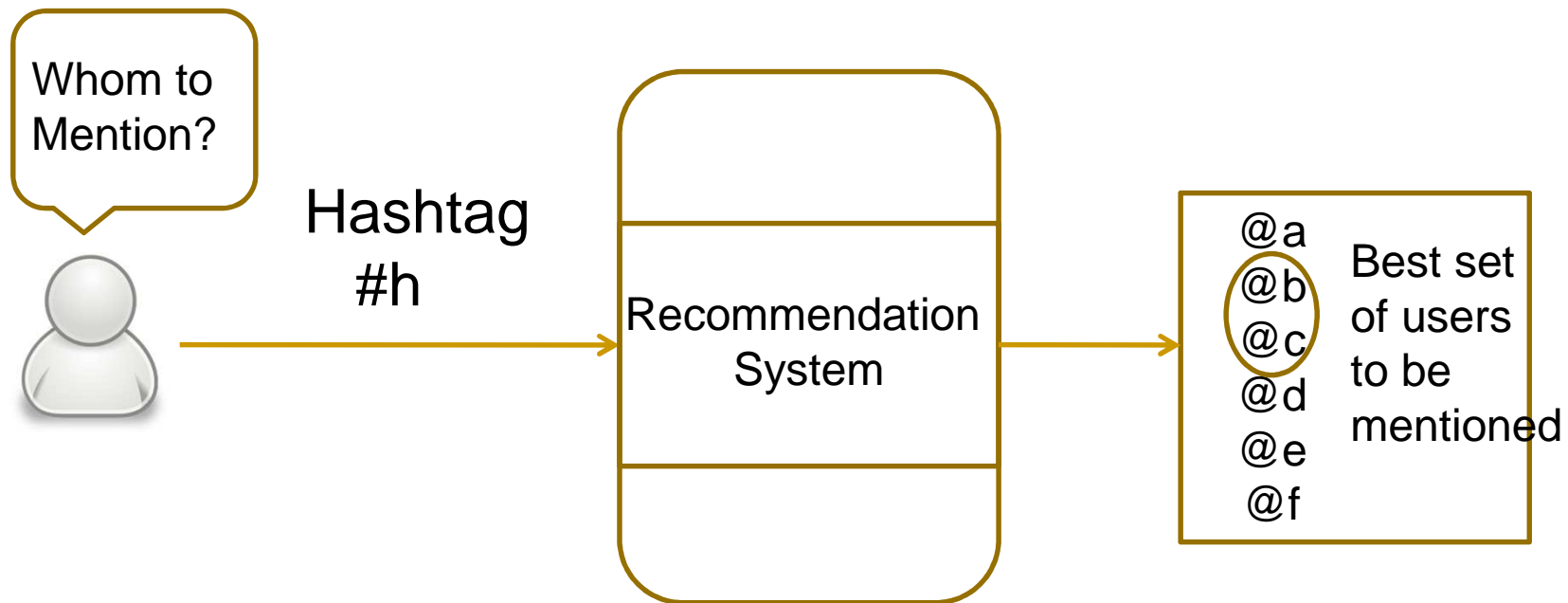
- Low probability of re-tweet (high number of followers)



Need to maintain a BALANCE

Objective

Recommendation System



Parineeti Chopra @ParineetiChopra · 1h

3 Domestic Tickets & 1 International Return Ticket to any city in the World to be Won. Take part in @KurkureSnacks #FlyHomeFree #Contest

Expand

Reply Retweet Favorite More

Outline

- **Data-study & Measurements**
 - Dataset
 - Representation
 - Dependency on Mention

 - **Model for hashtag Propagation**
 - Description
 - Set Parameters from Dataset
 - Validation
 - Insights

 - **Conclusion**
-

Data-study & Measurements

Dataset

- 15m Dataset → Contains information about tweets posted during the revolutionary movements in Spain during May 2011



■ Data:

- user_id ; timestamp ; hashtag_list ; mention_list
- Follow Links

■ Statistics:

- Total users: 87569
- Total Tweets: 529393
- Hash-Tags: 22376

Filtered relationships only to those who sent at least a message in that **topic**; or were mentioned by someone who did.

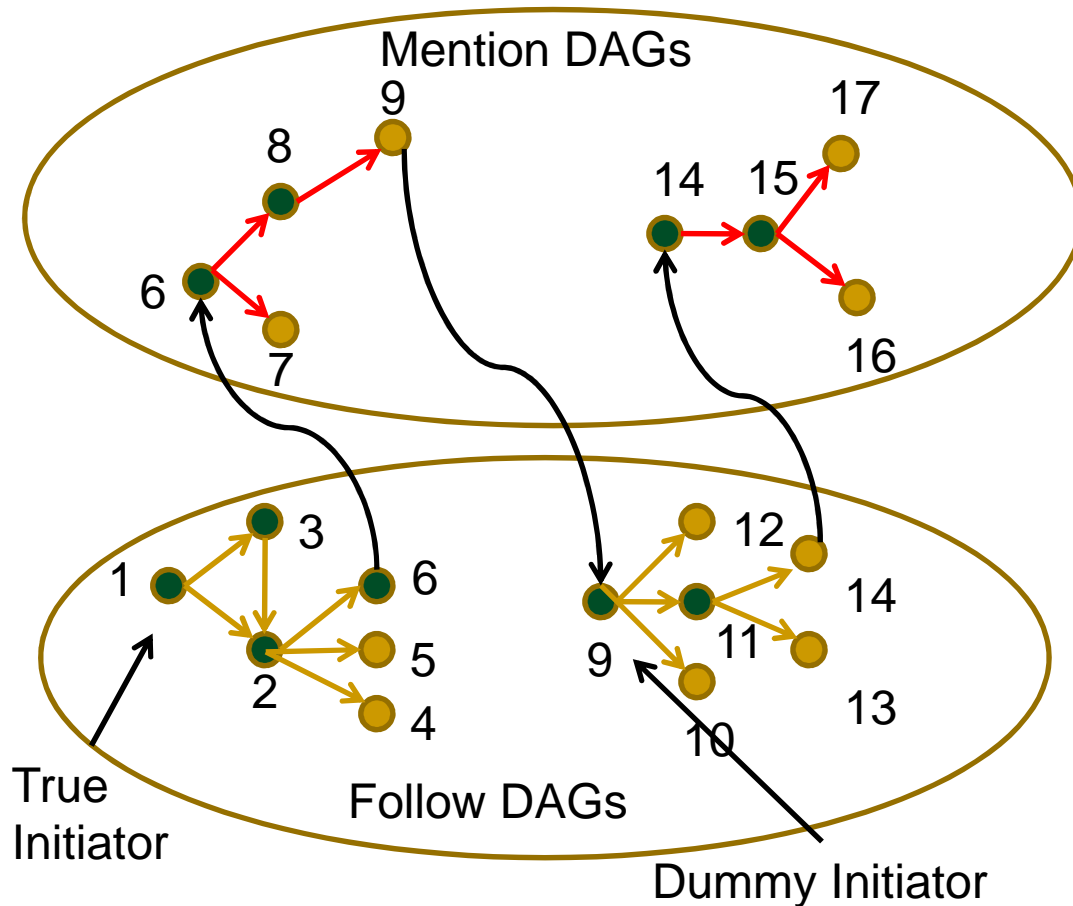
Source: <http://cosnet.bifi.es/research-lines/online-social-systems/15m-dataset>

Y. Moreno et al. The dynamics of protest recruitment through an online network. Scientific reports, 1, 2011.17

Other datasets: Arab-spring datasets

Data-study & Measurements

Representation



True Initiators

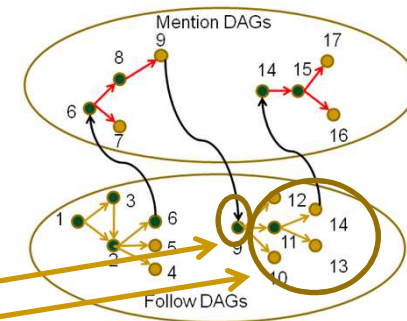
Those who get the information from external Source and spread in the network
e.g. 1

Dummy Initiators

Those who get the Information only through Mention links and spread In the network
e.g. 9

Data-study & Measurements

Dependency on Mention



- For each hashtag '#h',
 - Set of dummy initiators (set A)
 - Set of users who belong to only the DAGs rooted by dummy initiators (set B)
 - Set of users who have tweeted #h (set C)

- The users in sets A & B would not have got the information ('#h') without "Mention"

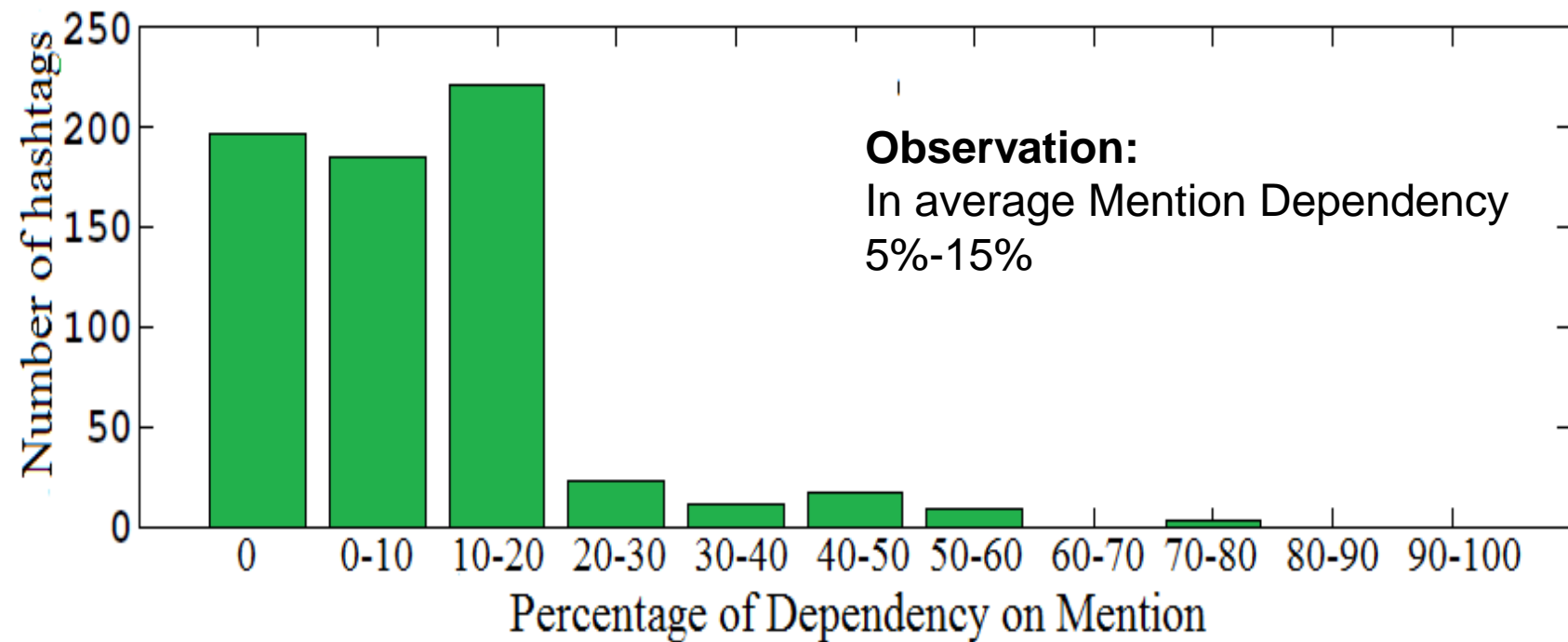
- Dependency of '#h' on Mention =

Fraction of tweeting users who would not have received the hashtag without Mention

$$= (|A|+|B|)/|C|$$

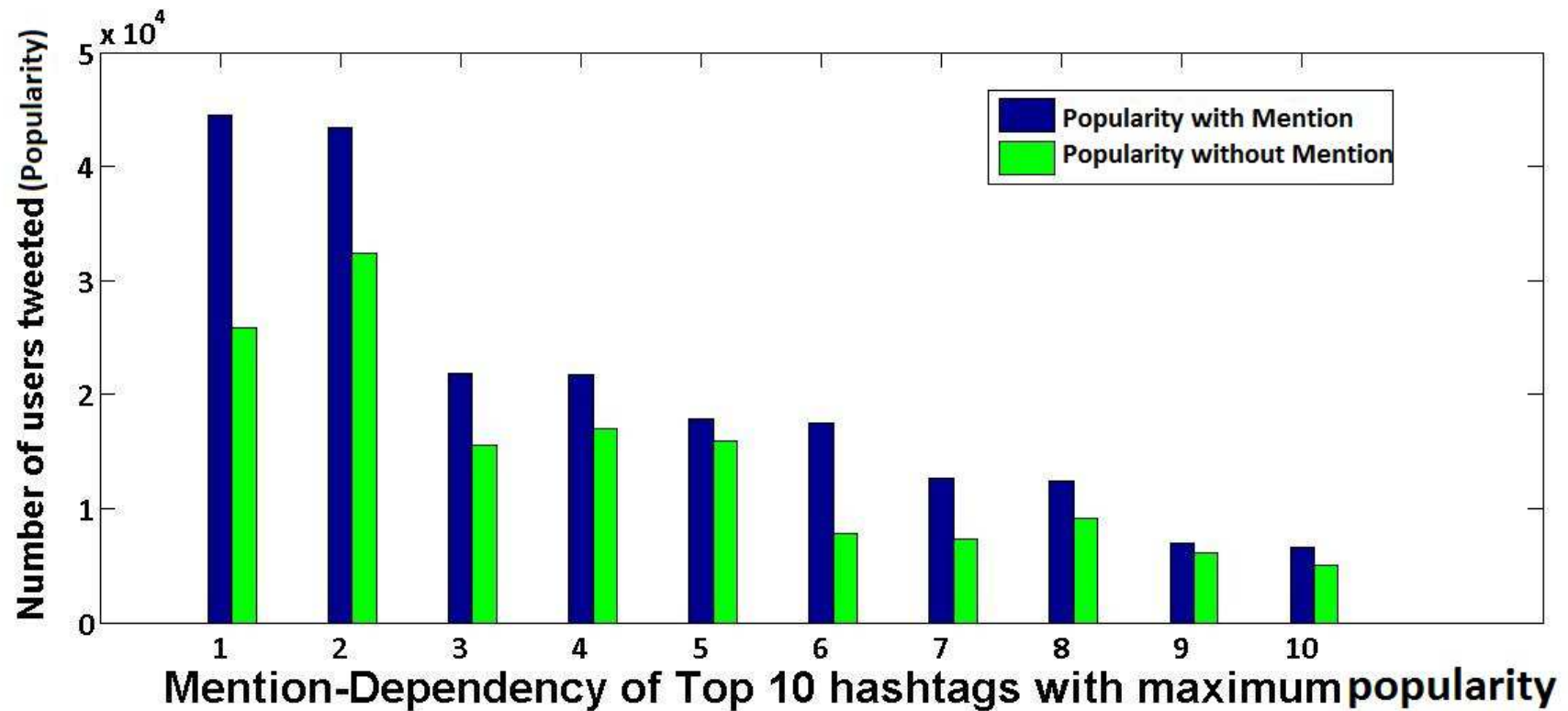
Data-study & Measurements

Dependency on Mention



Data-study & Measurements

Dependency on Mention



Observation:

Most of the highly popular hashtags are heavily dependent on Mention

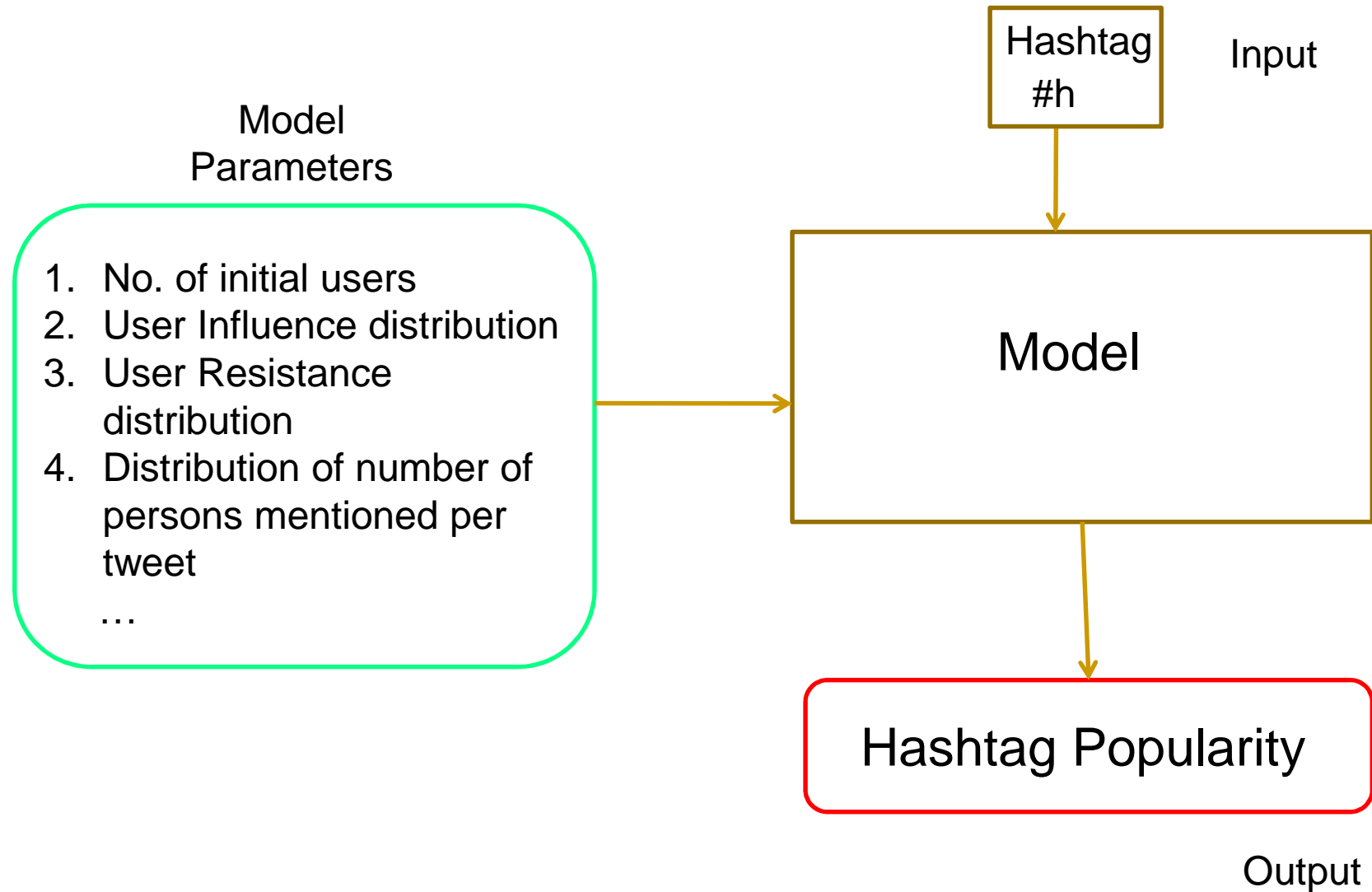
Outline

- Data-study & Measurements
 - Dataset
 - Representation
 - Dependency on Mention

- Model for hashtag Propagation
 - Description
 - Set Parameters from Dataset
 - Validation

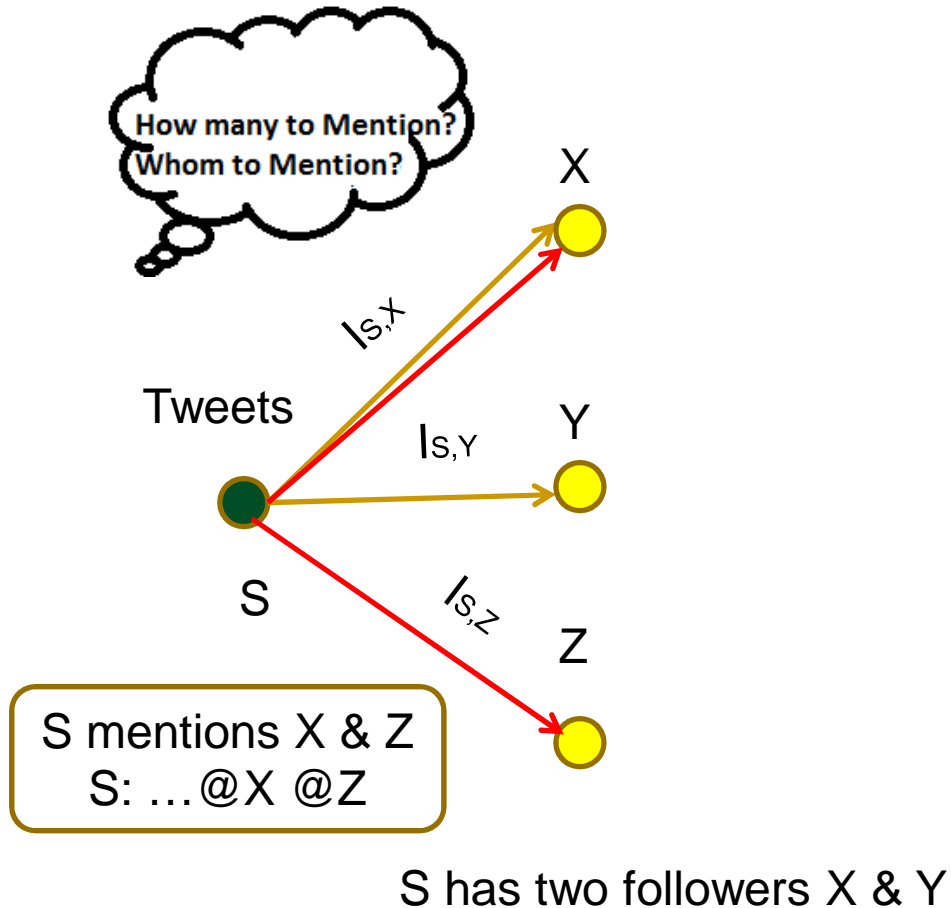
- Conclusion

Model for Hashtag Propagation



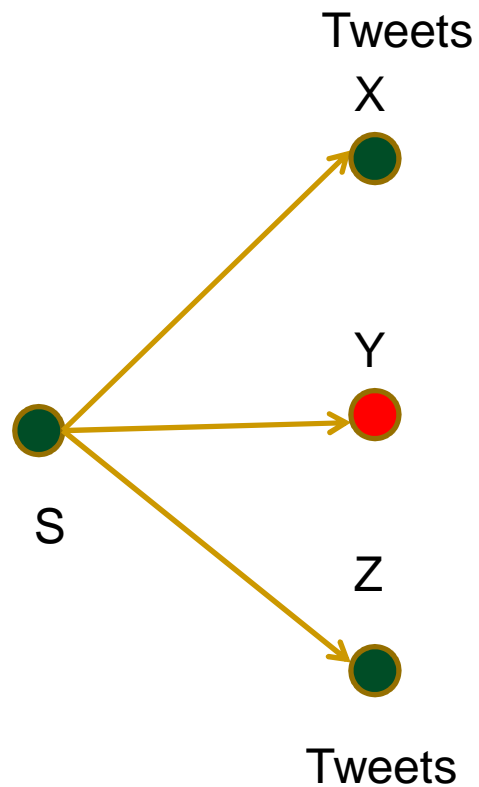
Model for Hashtag Propagation

Intuition Behind the Model



Model for Hashtag Propagation

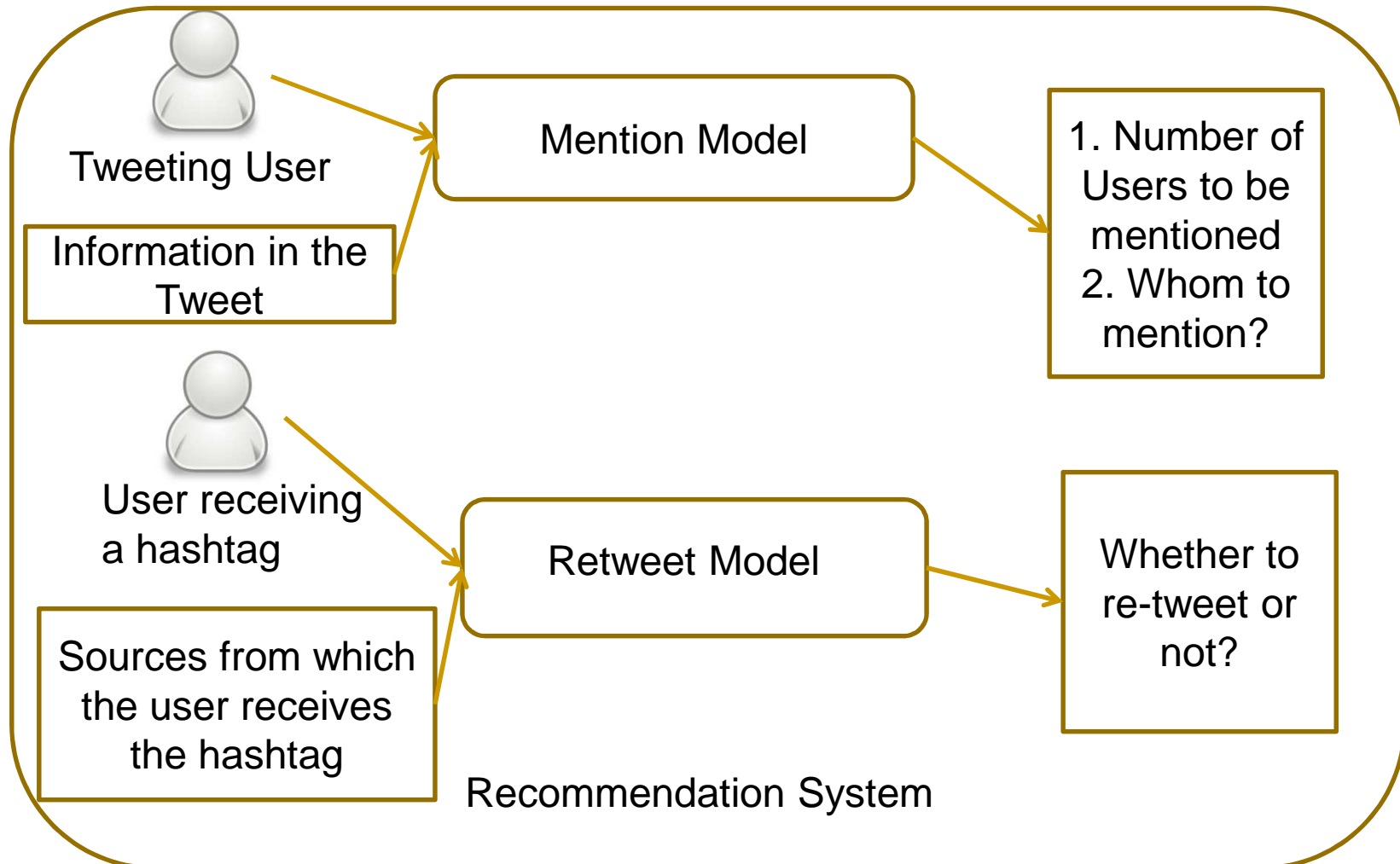
Intuition Behind the Model



X, Z decide to propagate
Y decides not to propagate

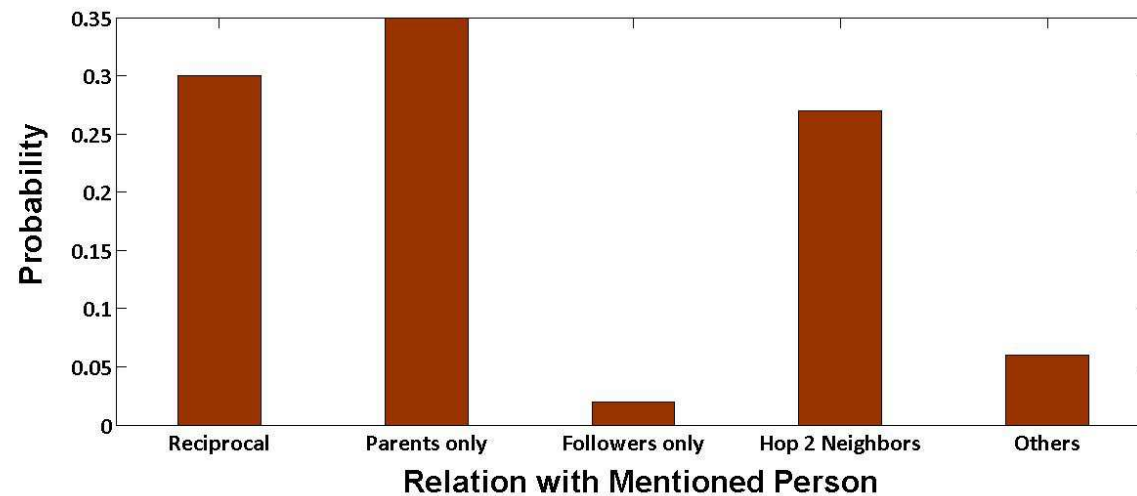
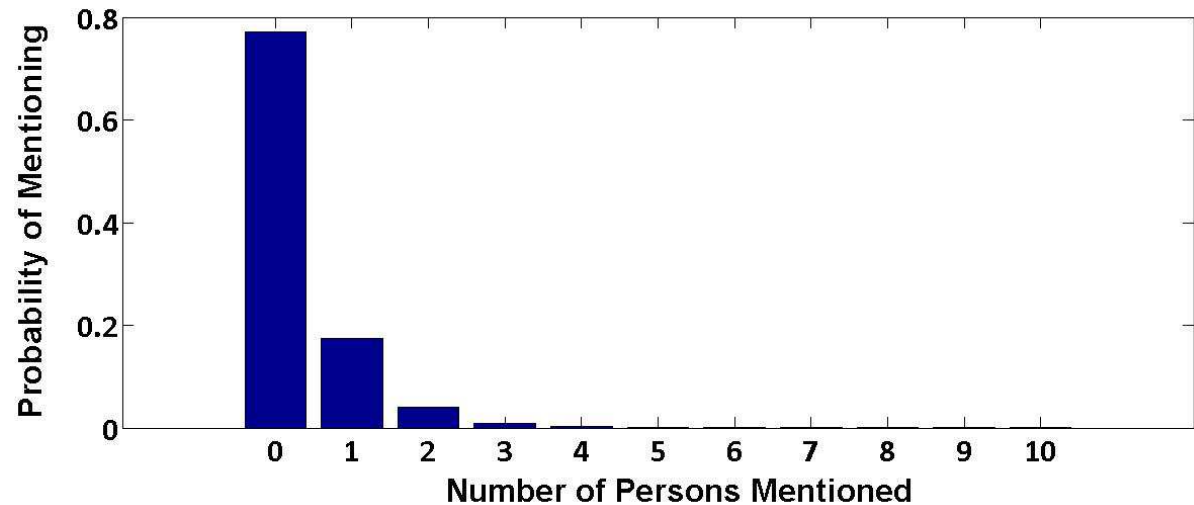
Model for Hashtag Propagation

Model Components



Mention Model

- Distributions calculated from the Dataset



Retweet Model (Linear Threshold Based)

- We calculate the weightage of each link a user (say, u1) gets
 - Factors:

- Influence of the user (say, u2) from whom the link is coming (Calculated using PageRank)
- Importance of the link based on the
 - Type of Link (Follow/Mention/Mixed)
 - Social Tie between u1 & u2 (Reciprocity) (calculated from dataset)
- Time-gap between u2's tweet and the current time

- We also calculate the passivity/resistance of each user

$$1 - \frac{\text{Number of times tweeted}}{\text{Number of times received any hashtag}}$$

Validation

1. Simulate model with a fixed number of initial users with different parameter values and get the popularity values
2. From dataset, collect hashtags with almost same number of initial users as the simulation
3. Check whether popularities of those real hashtags from the dataset fall within the range of simulated popularity values

Hashtag	Real Original Tweeters	Real Final Popularity	Predicted Popularity for 1100 initial users
#worldrevolution	1091	1866	[1341-3733]
#acampadasol	1125	3449	[1341-3733]

Outline

- Data-study & Measurements
 - Dataset
 - Representation
 - Dependency on Mention

- Model for hashtag Propagation
 - Description
 - Set Parameters from Dataset
 - Validation

- Conclusion

Conclusion

- “Mention” definitely plays a key-role in deciding the popularity of hashtags
 - Using insights from simulations of our model, our recommendation System should try to
 - Suggest minimum number of users (due to character limitation of tweets)
 - So that maximum popularity can be achieved
-

*Thank
You*