

Rumor Detection in Online Social Networks

Presented by
Kale Ashish Anil
14IT60R03

Under the Guidance of
Dr. Shamik Sural



Dept. of Computer Science and Engineering
Indian Institute of Technology, Kharagpur



Contents

- ▶ Introduction
- ▶ Twitter Terminology
- ▶ Related Work
- ▶ Overall System Architecture
- ▶ Data Collection
- ▶ Data Filtering
- ▶ Part-of-Speech Tagging
- ▶ Tweet Clustering
- ▶ Analysis – Temporal, Burst, Data
- ▶ Future Work

Introduction

- ▶ Online social networks allows rapid propagation of information
- ▶ No one to verify the authenticity of the information provided
- ▶ May lead to spread of rumors

- ▶ Rumor
 - ▶ Controversial
 - ▶ Fact-checkable statement

- ▶ Identify potential rumors in Twitter data
- ▶ Extract features relevant to rumor

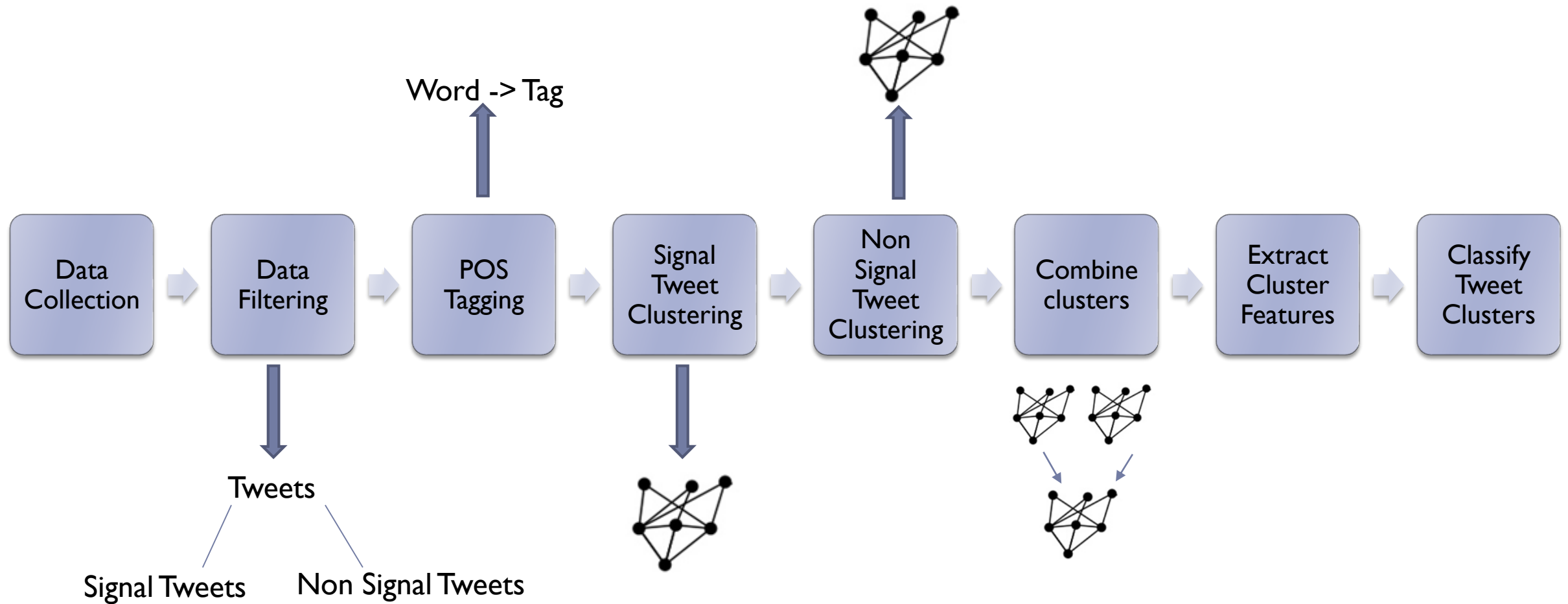
Twitter Terminology

- ▶ Twitter is a social communication tool where people broadcast short messages
- ▶ Tweet
 - ▶ Each message written on Twitter is called a tweet, limited to 140 characters in length
- ▶ Follower
 - ▶ Follow another user to see his or her updates on your Twitter home page
- ▶ User Mention
 - ▶ Twitter allows user to mention other user in a tweet using @ symbol
- ▶ Hashtag
 - ▶ Using # symbol a user can enrich the subject being discussed in the tweet

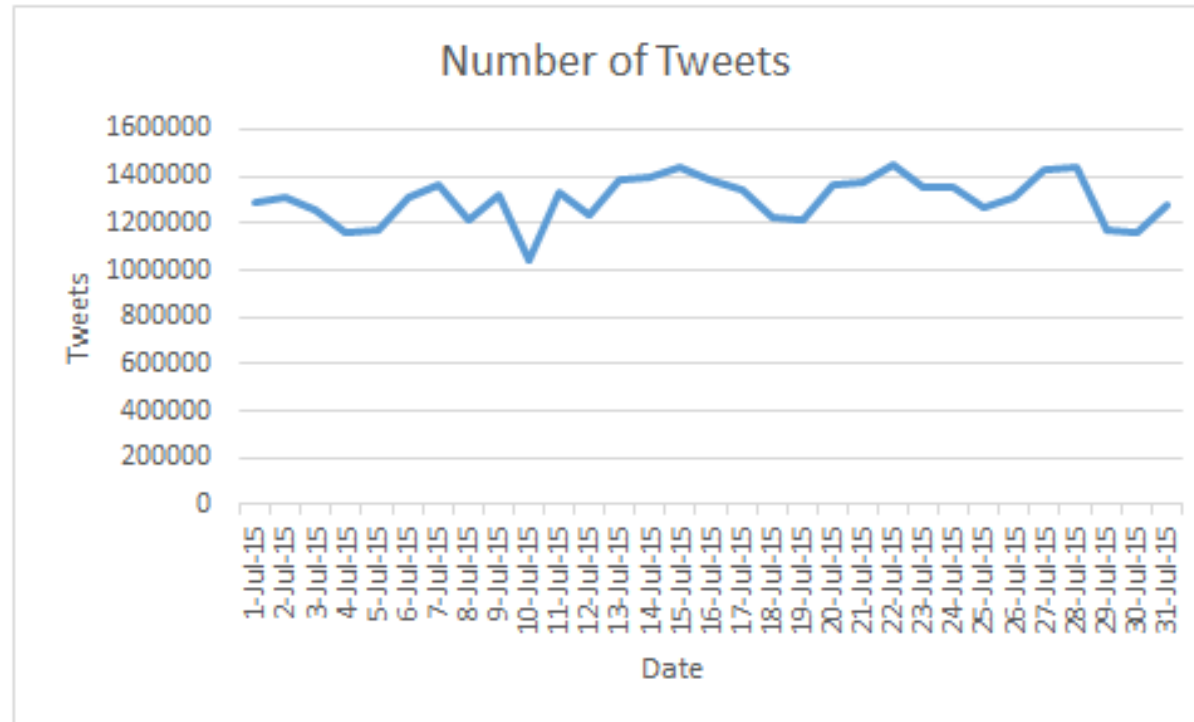
Related Work

- ▶ Qazvinian, Rosengren, Radev, Mei[1], 2011
 - ▶ Annotate a set of tweets
 - ▶ Uses network, content and Twitter features
 - ▶ Predict whether a new tweet contains a known rumor or not
 - ▶ The work is mainly targeted to retrieve a set of related rumors
 - ▶ Does not detect new types of rumors
- ▶ Zhao, Resnick, Mei[2], 2015
 - ▶ Partition tweets in 2 sets
 - ▶ Signal - Contains enquiry patterns eg. Really?, Is it true?, what?, rumor, debunk
 - ▶ Non-signal - Rest of the tweets
 - ▶ Form rumor cluster using Jaccard similarity between signal tweets

Overall System Architecture



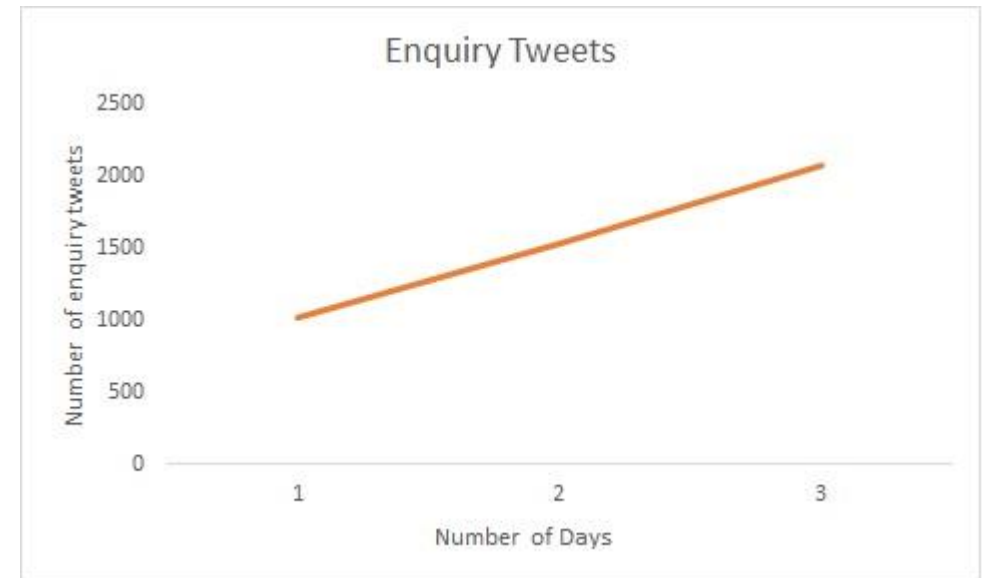
Data Collection



➡ 1.3 million tweets on average

Data Filtering

- ▶ Signal/Enquiry tweets – set of tweets enquiring about the fact
 - ▶ is it true?
 - ▶ what??
 - ▶ really!
 - ▶ debunk
 - ▶ It is not true
- ▶ Non-signal tweets – Other tweets

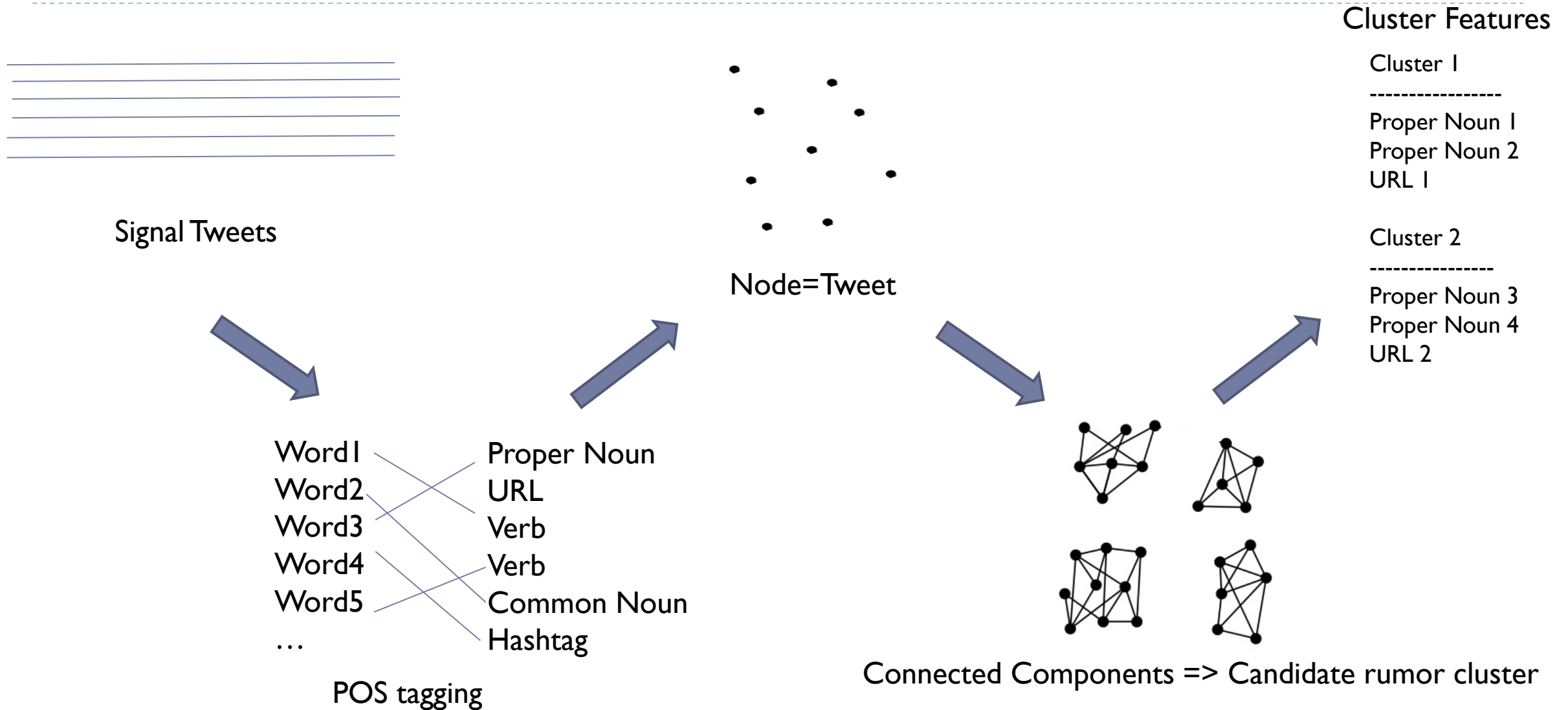


Part-of-Speech Tagging

- ▶ RT @brownjenjen : Ben Affleck denies affair rumors #rumor <http://t.co/qwrfe>

Part-of-Speech	Tag
RT	Re-tweet
@brownjenjen	User-mention
:	Discourse marker
Ben	Proper Noun
Affleck	Proper Noun
denies	Verb
affair	Common Noun
rumors	Common Noun
#rumor	Hashtag
http://t.co/qwrfe	URL

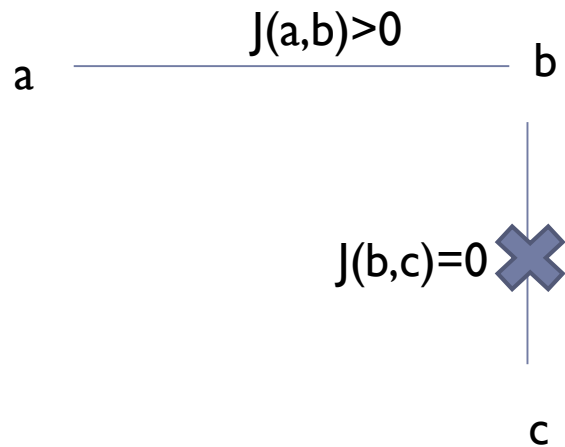
Signal Tweet Clustering



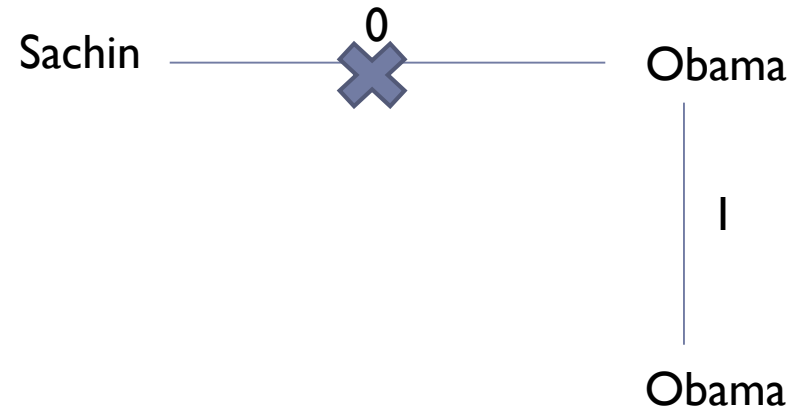
Tweet Similarity

Jaccard

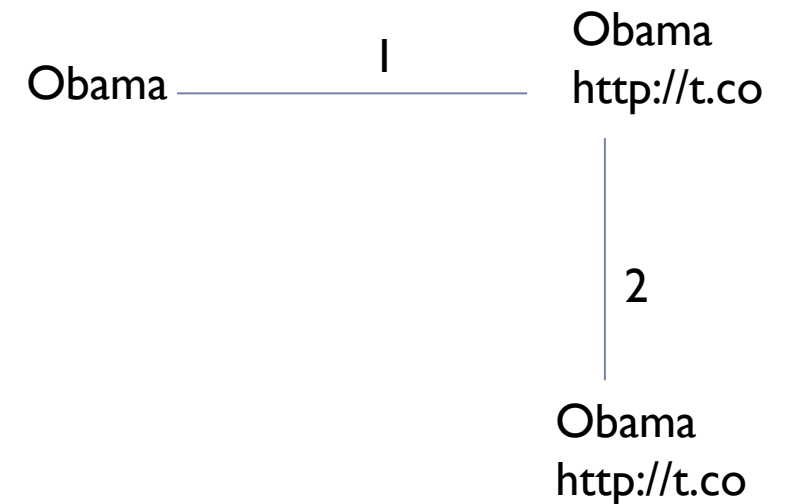
$$Jaccard(a,b) = \frac{|ngram(a) \cap ngram(b)|}{|ngram(a) \cup ngram(b)|}$$



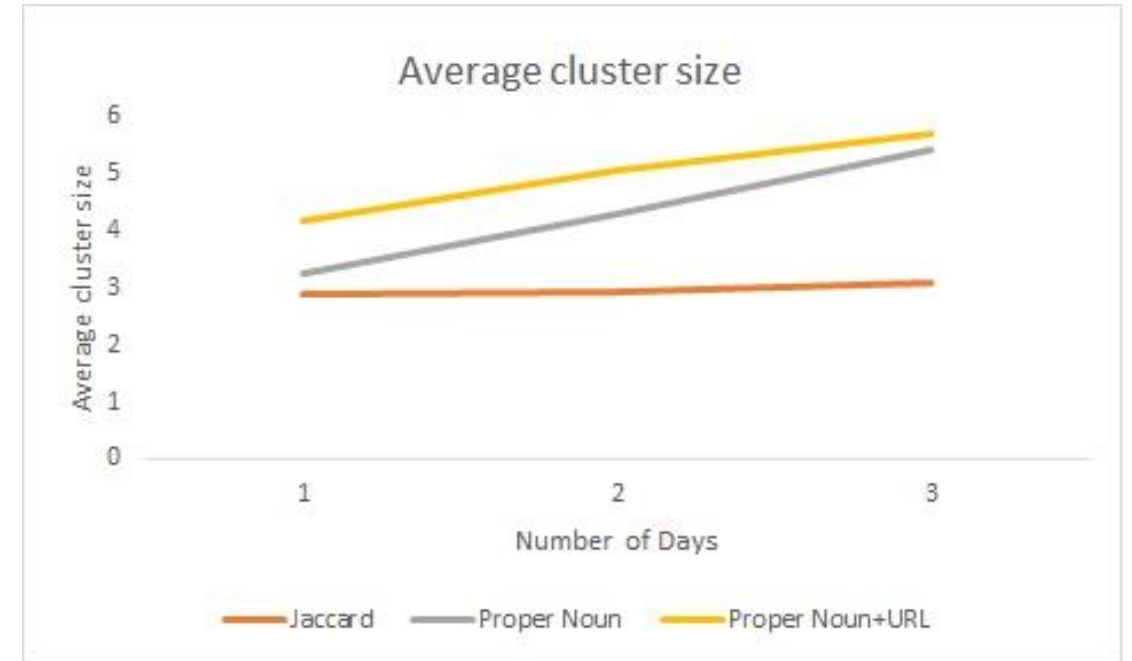
Proper Noun



Proper Noun + URL



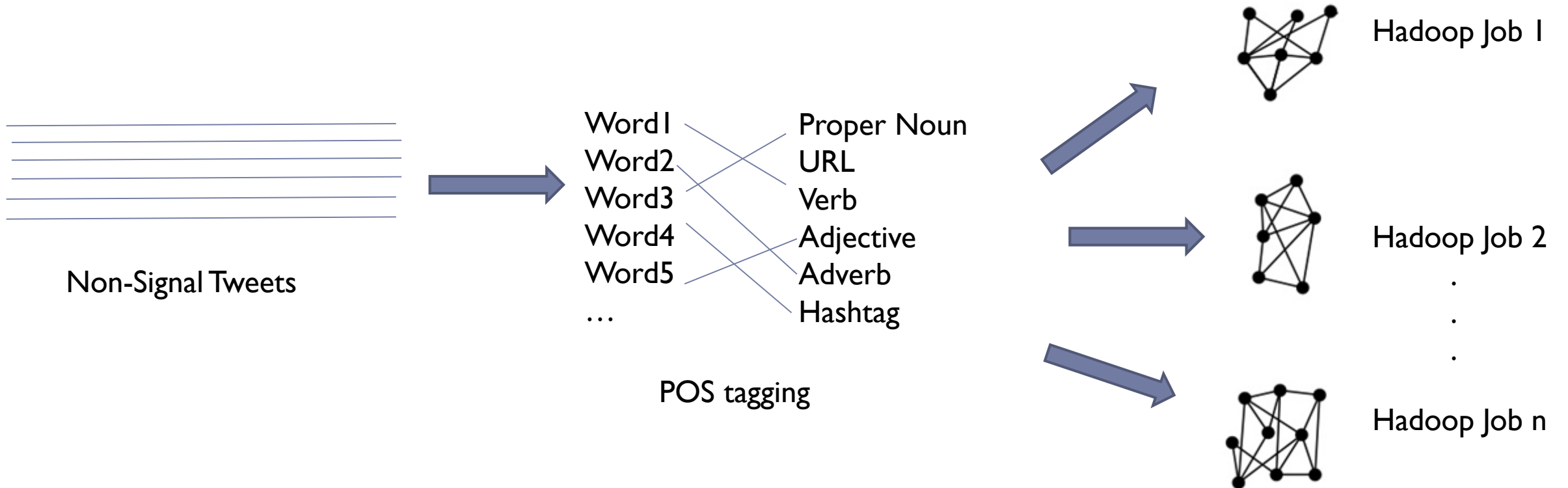
Tweet Similarity Performance



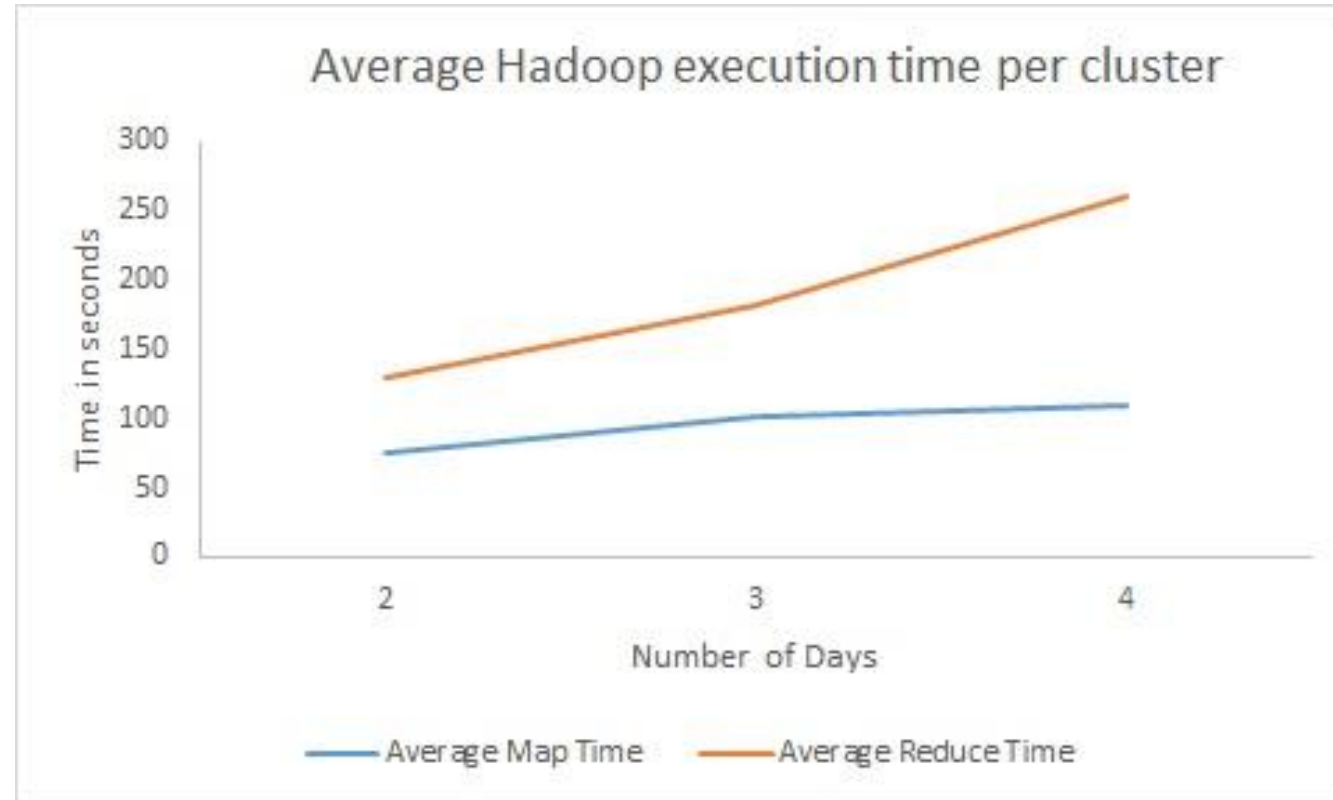
➡ Proper Noun+URL Tweet similarity measure gives overall good performance

➡ Average cluster size does not increase for Jaccard as word diversity increases

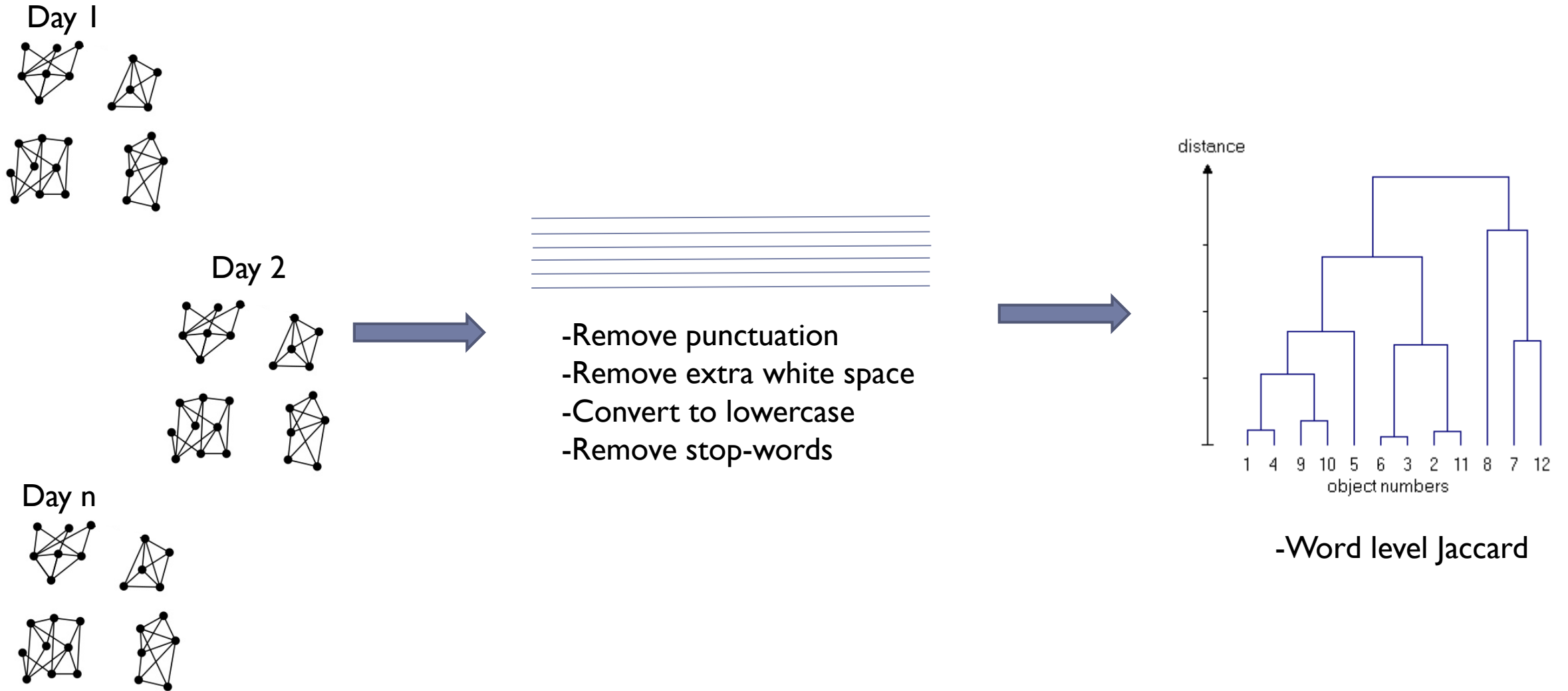
Non-Signal Tweet Clustering



Clustering Time

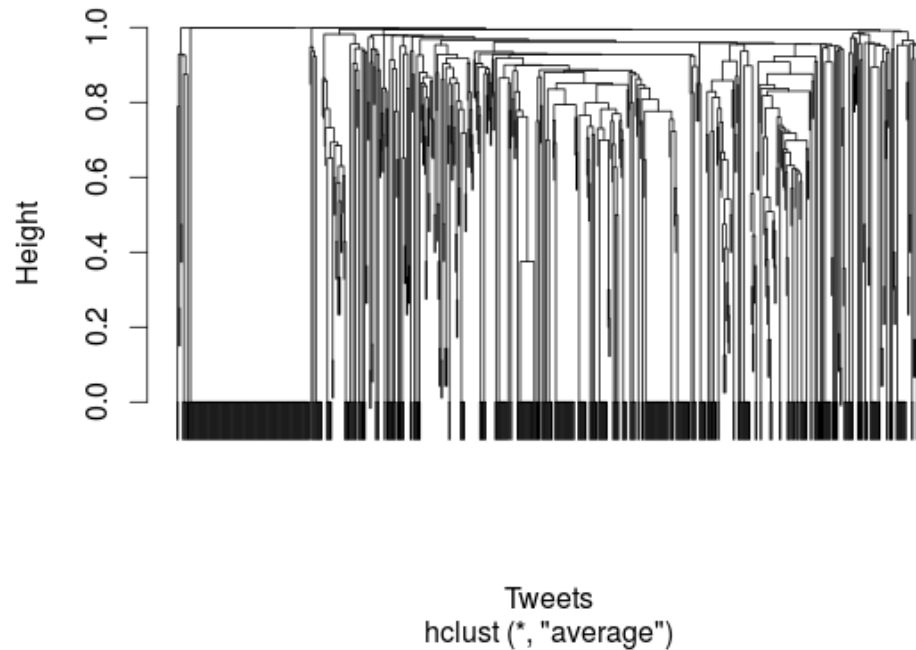


Combining Clusters

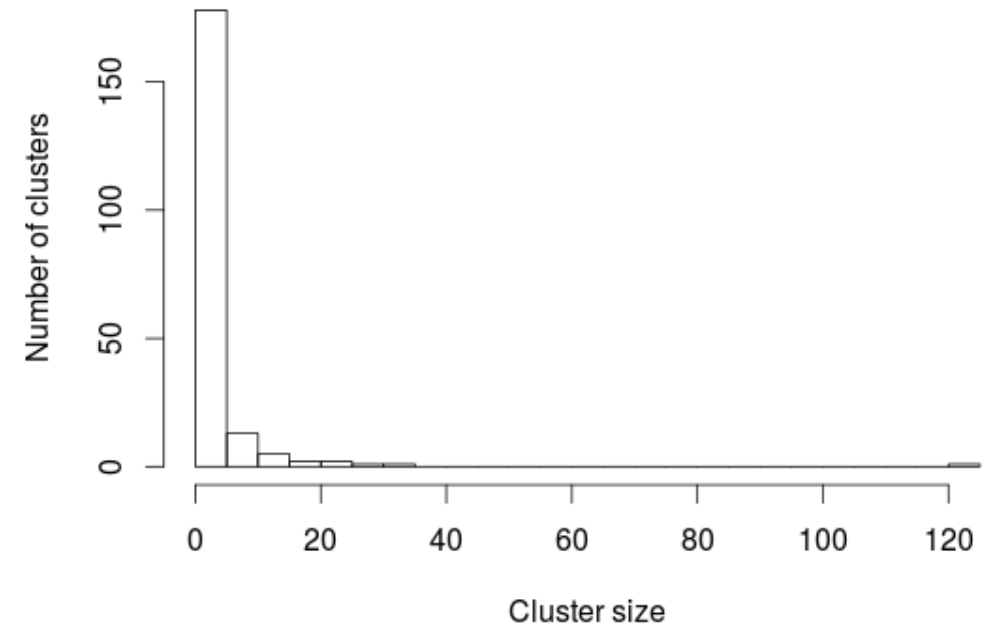


Combining Clusters(Contd..)

Cluster Dendrogram



Hierarchical clustering performance

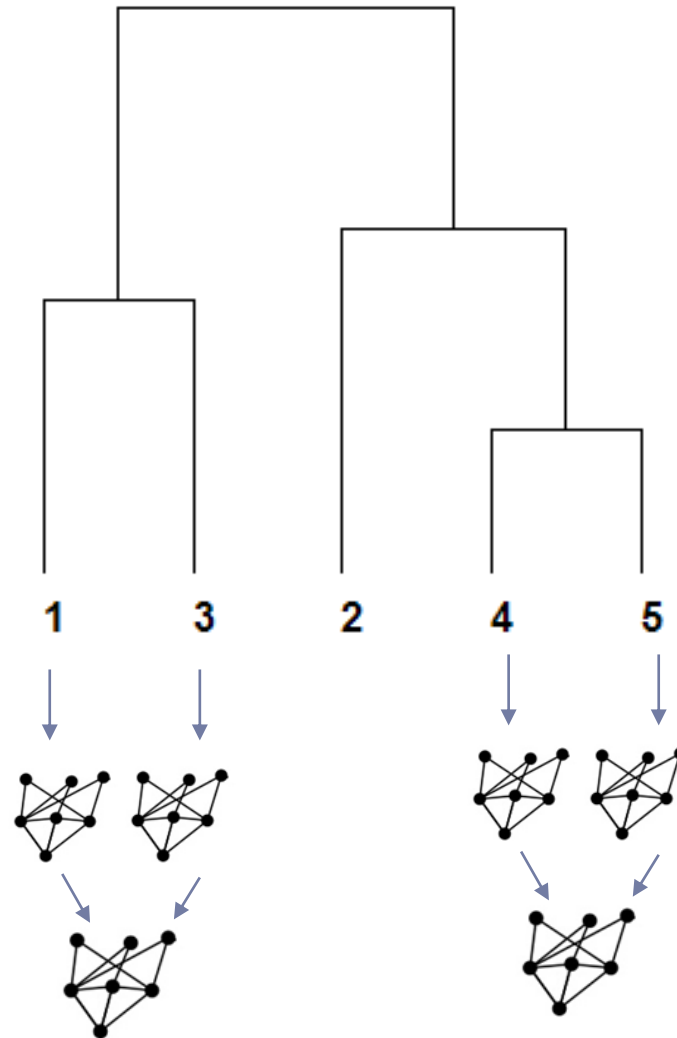


Sparsity of words creates small clusters

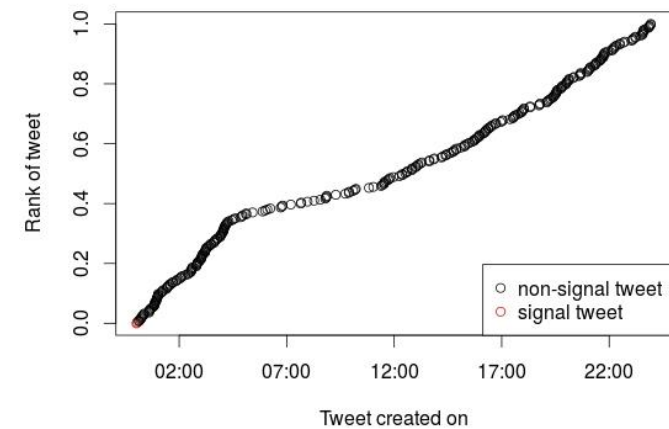
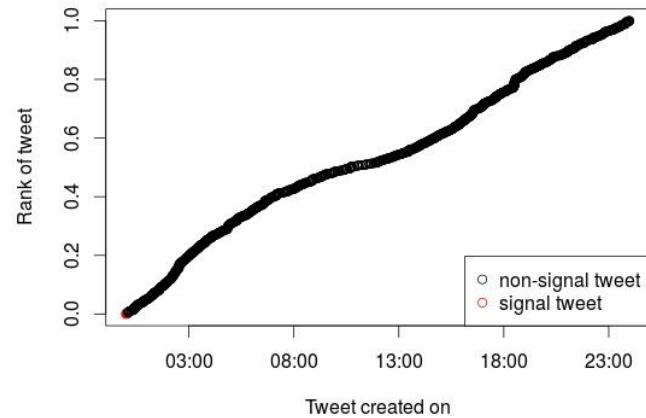
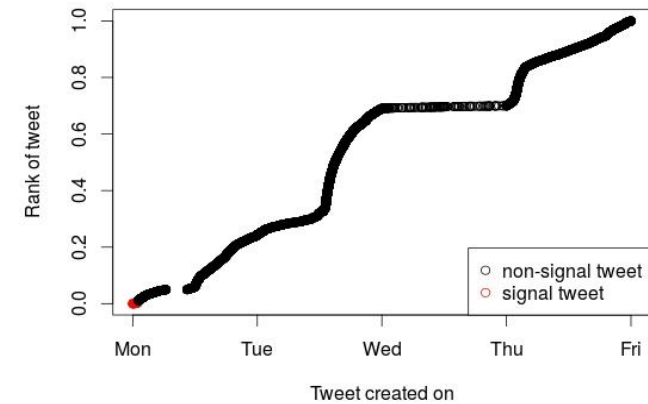
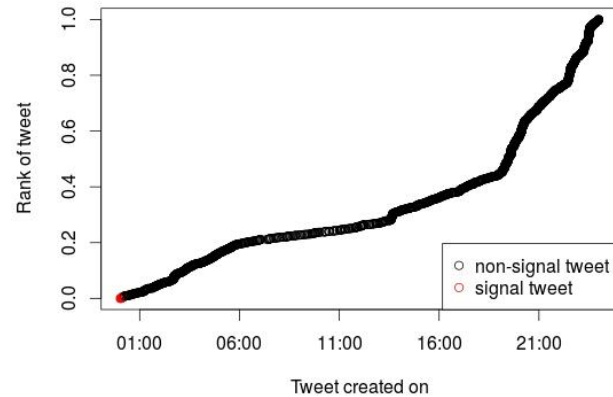


Need to merge clusters using some other measure as well – POS tags

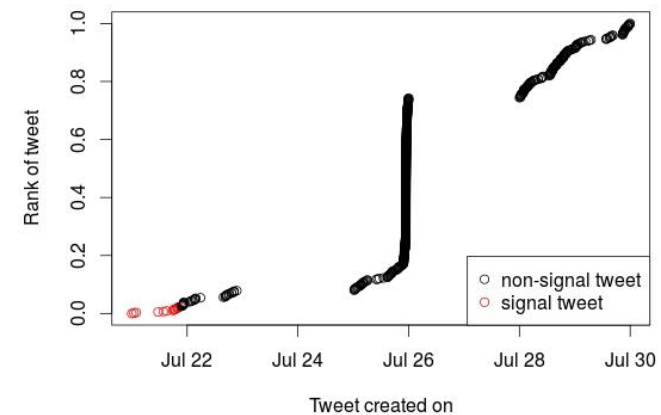
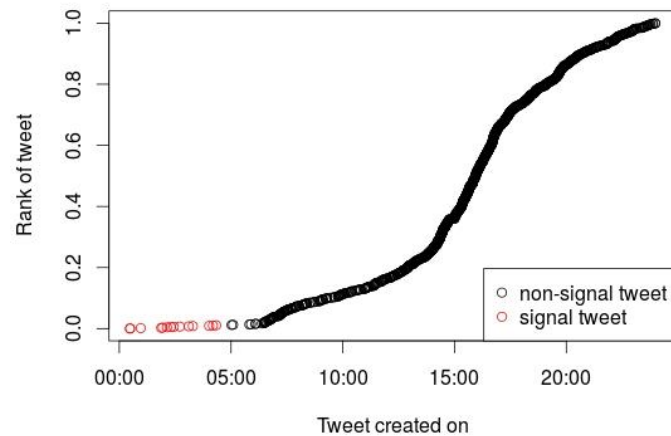
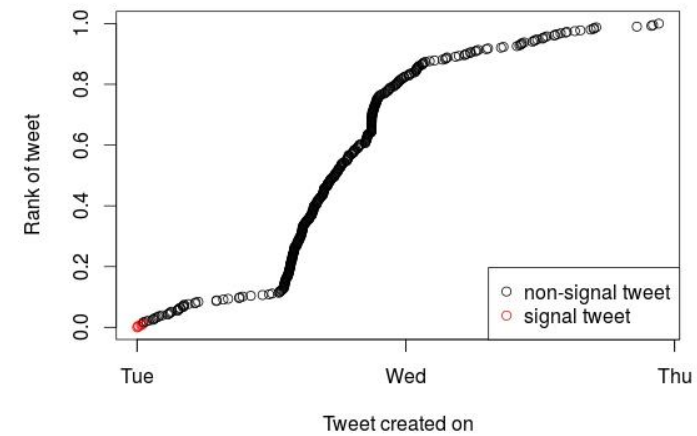
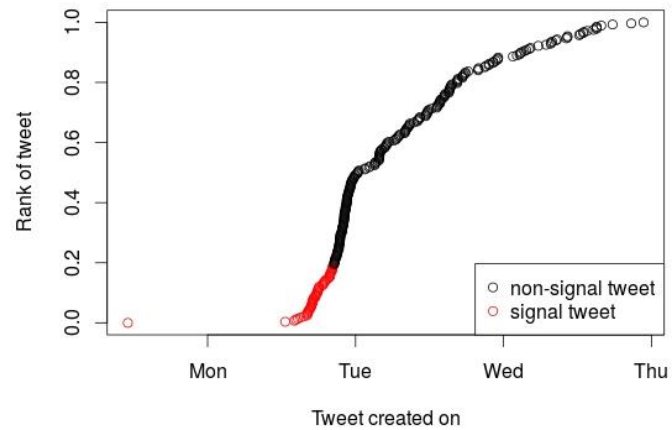
Combining Clusters(Contd..)



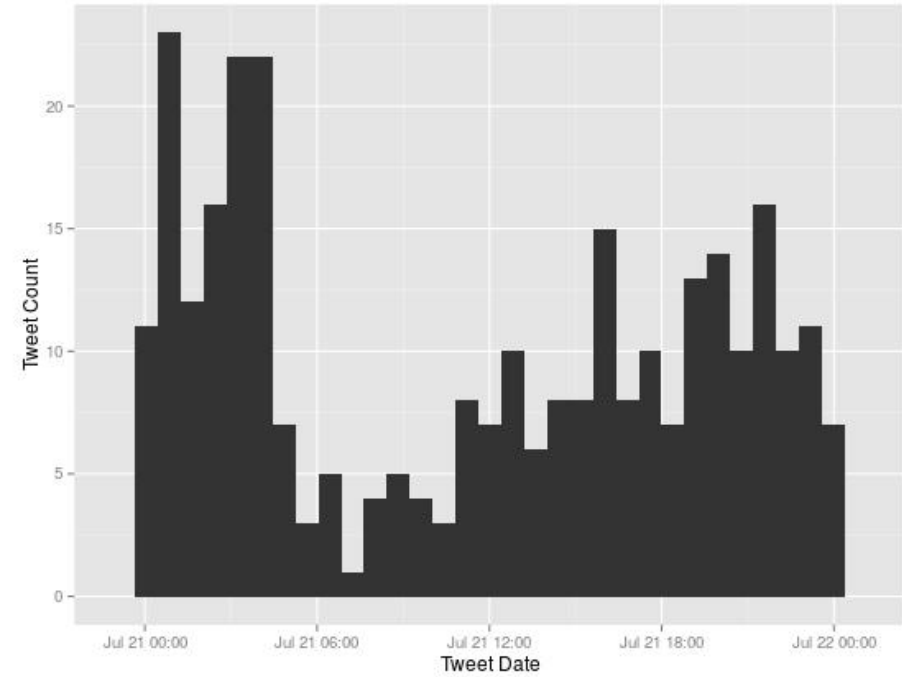
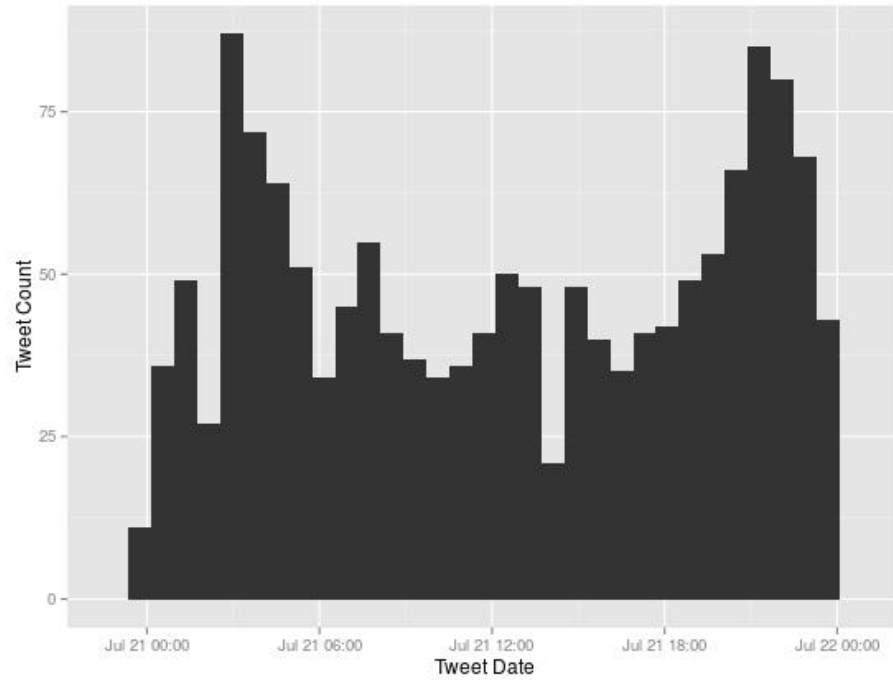
Temporal Analysis – Non-Rumor



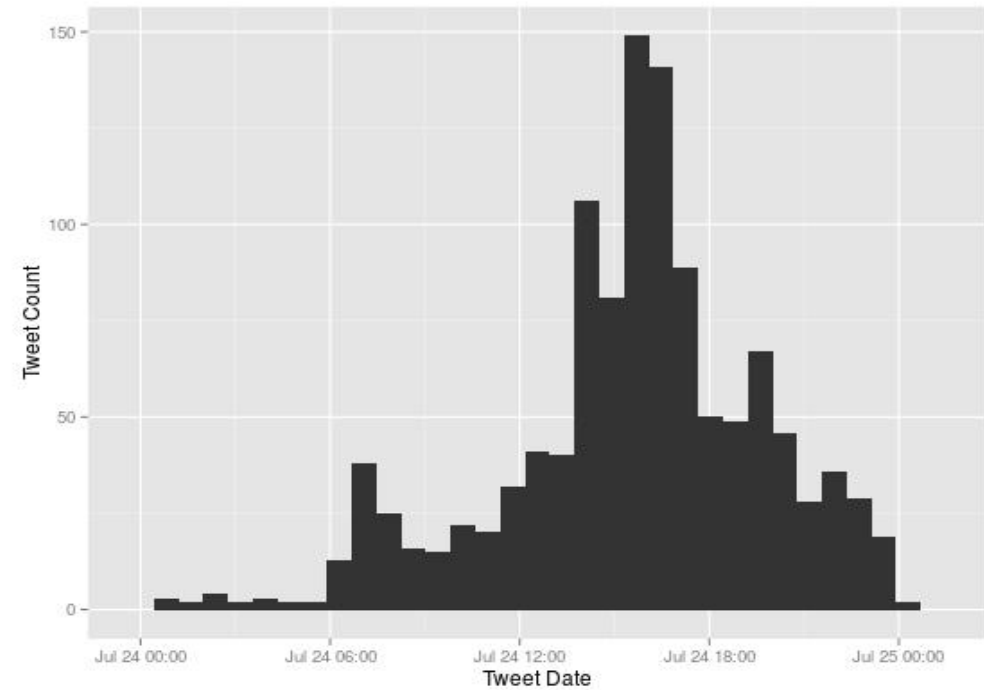
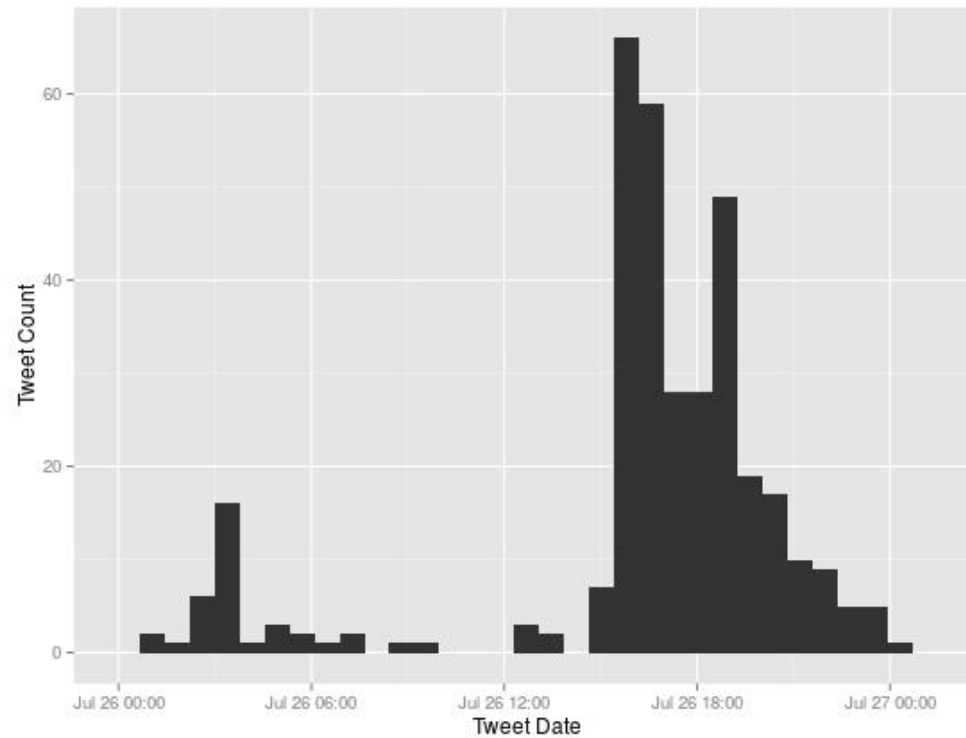
Temporal Analysis – Rumor

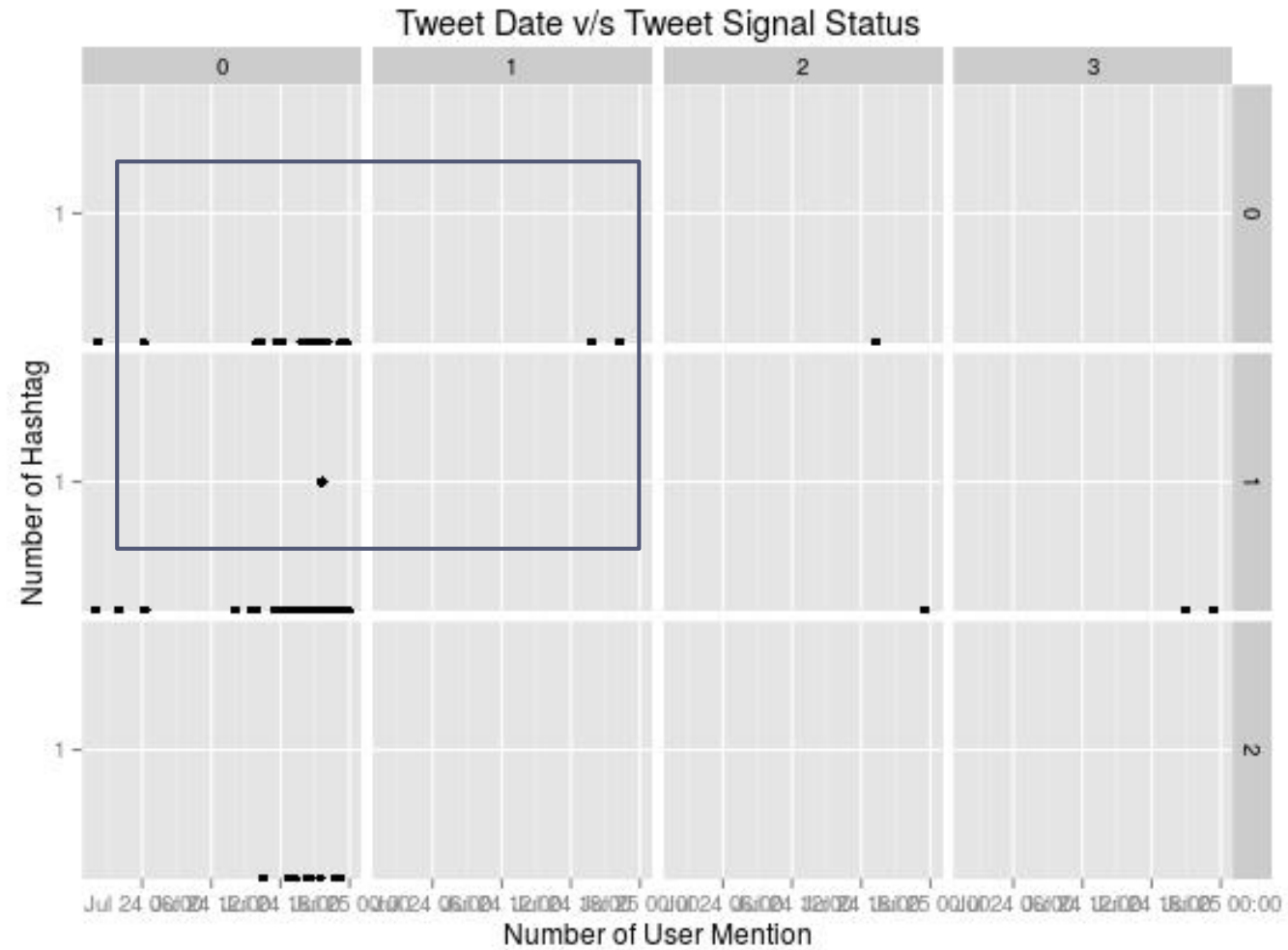


Burst Analysis – Non-Rumor

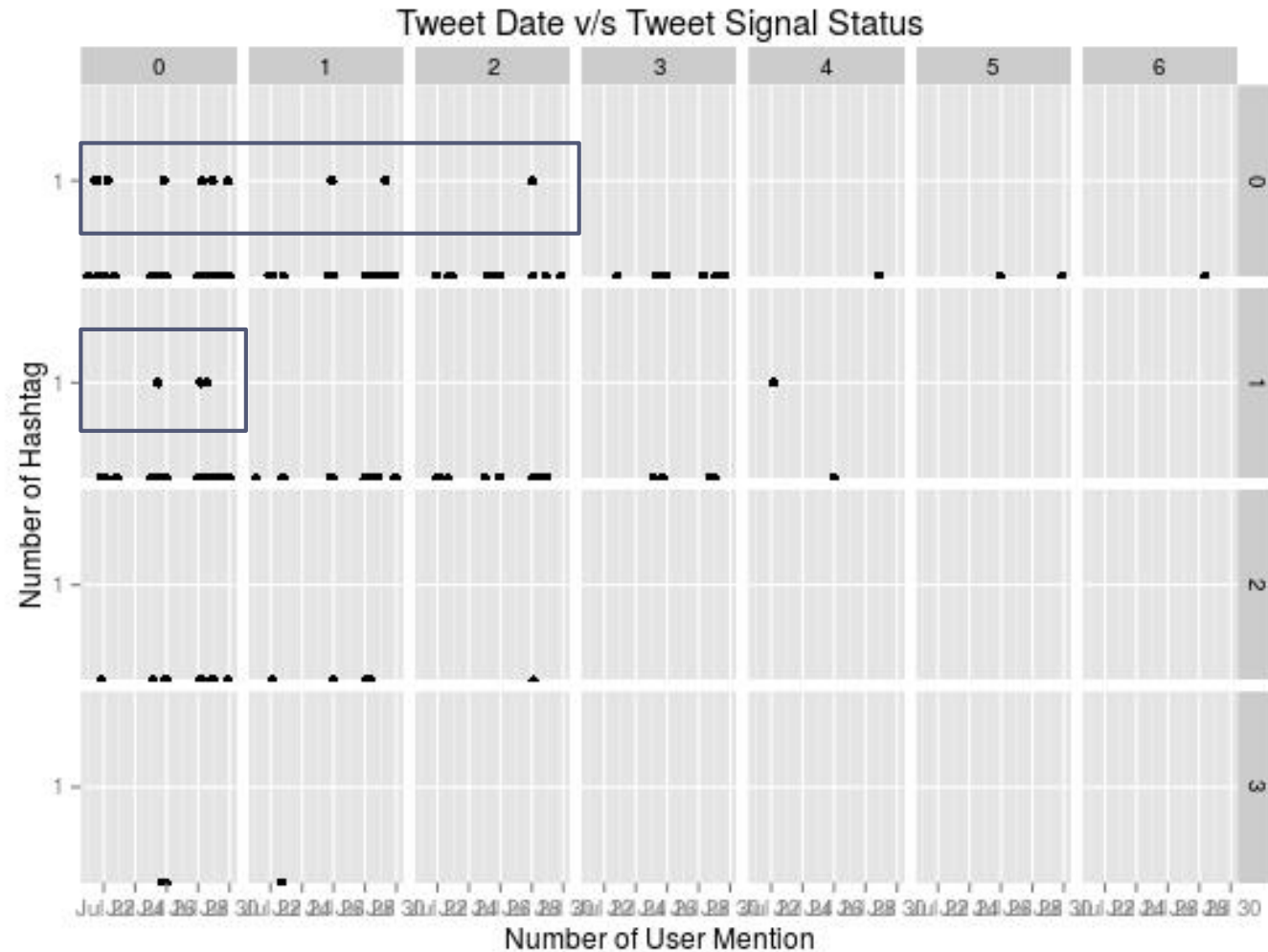


Burst Analysis – Rumor





Data Analysis – Rumor



Future Work

- ▶ Linking to semantic web
- ▶ Design of classifier using temporal as well as non-temporal properties
- ▶ Extend system to other media platforms (social and traditional)
- ▶ Predict the impact of rumors
- ▶ Strategies for dampening the effects of rumors

THANK YOU!



References

- [1] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 1589–1599. Association for Computational Linguistics, 2011.
- [2] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web, pages 1395–1405. International World Wide Web Conferences Steering Committee, 2015.
- [3] Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. Part-of-speech tagging for twitter: Annotation, features, and experiments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pages 42–47. Association for Computational Linguistics, 2011.
- [4] Tom White. Hadoop: The definitive guide. " O'Reilly Media, Inc.", 2012.