New Methods for Ranking Influence in Social Networks

Luis F. Chiroque
IMDEA Networks Institute

Hypothesis & Problem

The basic hypothesis is that past dynamics on social networks can be used to predict the most influential users in a future.
In this work, propagation dynamics on social networks are studied in order to identify the most influential users.
We use the activity in a social network during a period to predict the influential users in a different period.

Methodology

- For this purpose, diffusion data has been collected during 4 weeks from a microblogging OSN (online social network) called *Tumblr*. Then, the propagation graph has been built and studied using the first 2 weeks data (period T_1). Subsequently, this graph has been used to predict the influencers during the last 2 weeks (period T_2). A ranking of influential nodes is obtained for T_2, set as the *ground truth*. The aim is to predict this ranking using the data from T_1.
- Based on the average spread of users' posts, rankings obtained with several techniques are tested and compared. These techniques include classical centrality measures used in the literature, the T_1 ranking itself, and new alternatives based on effective degree using local (network) information.

Definitions

- **Influence (*definition*)**
  A user is more influential than other when the former has a greater average propagation-cascade size.
- We build a ***weighted directed graph*** G(V,E) from the union of all cascades caused by the messages posted.
- A real value is assigned to each vertex in V, which is derived from its activity and its properties in the cascades graph.
- **Classic Centrality Metrics:**

Degree	Betweeneess	Closeness
μ-PCI*	PageRank	HITS (auth & hub)

- **Our metrics:**
  - **Effective degree**

$$\hat{k}_v = \sum_{w \in N_{out}(v)} weight\left((v,w)\right)$$

  - **Ego-Additive Effective Degree**

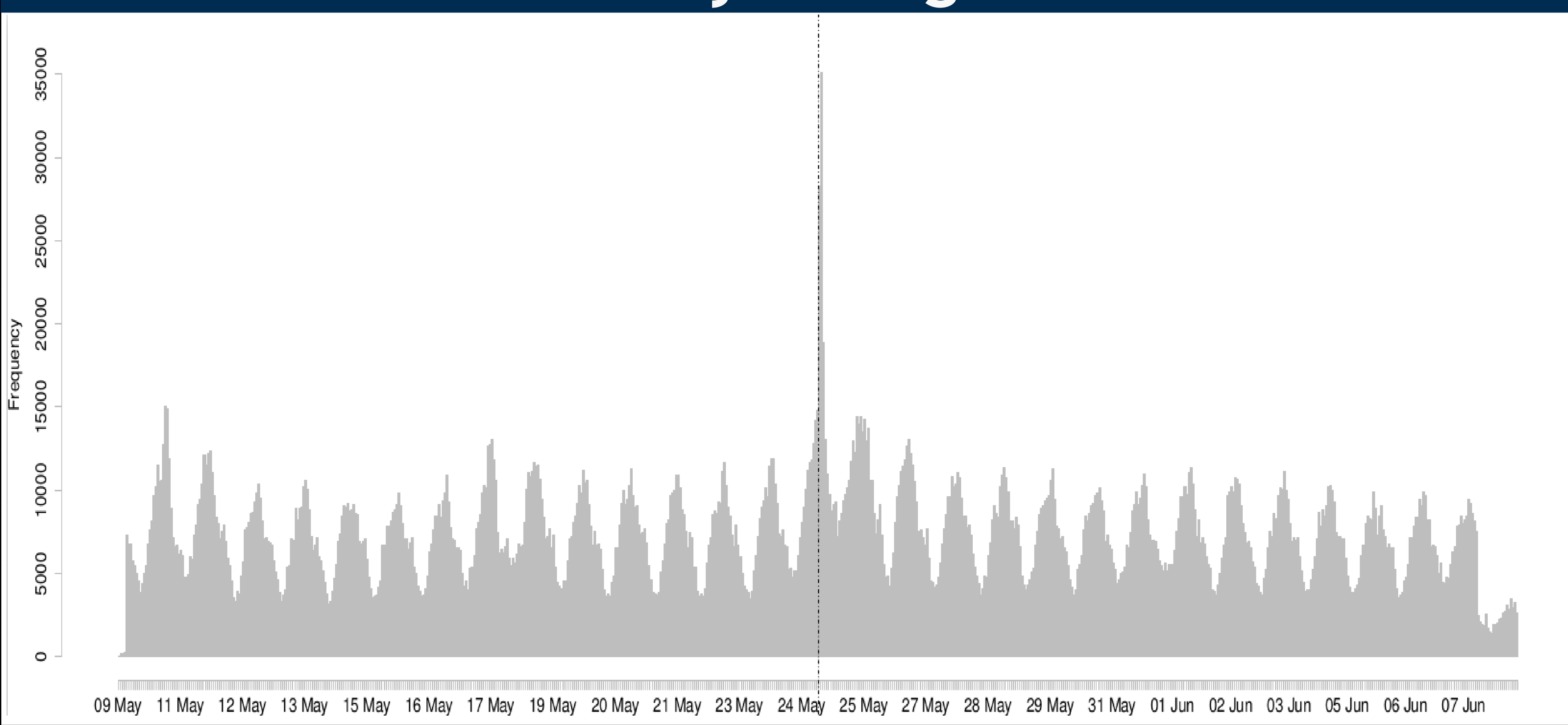
$$\text{EgoAED}_v = \hat{k}_v \sum_{w \in N_{out}(v)} \hat{k}_w$$

* this metric has no weighted version.

Data Set

event	duration	vertices	edges	Content-generator users
2014 UEFA Champions League final	4 weeks	17,756	205,011	872

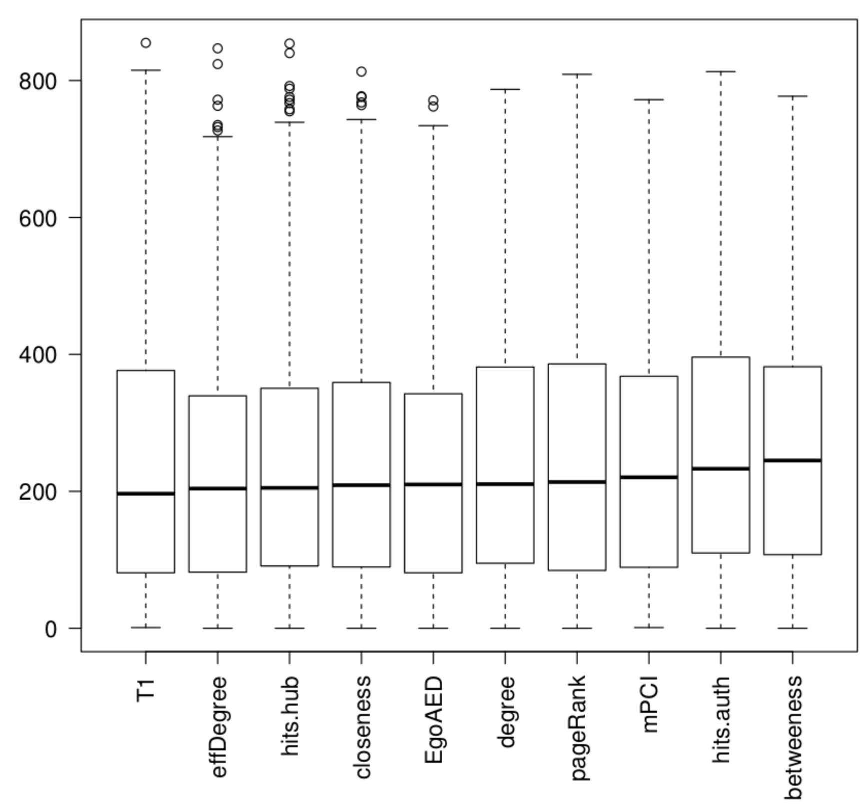
Activity Histogram



Global Ranking Results

- Absolute mean error

$$absErr = \frac{\sum_{i=1}^{N} |D|}{N}$$



- Spearman's rank correlation coefficient

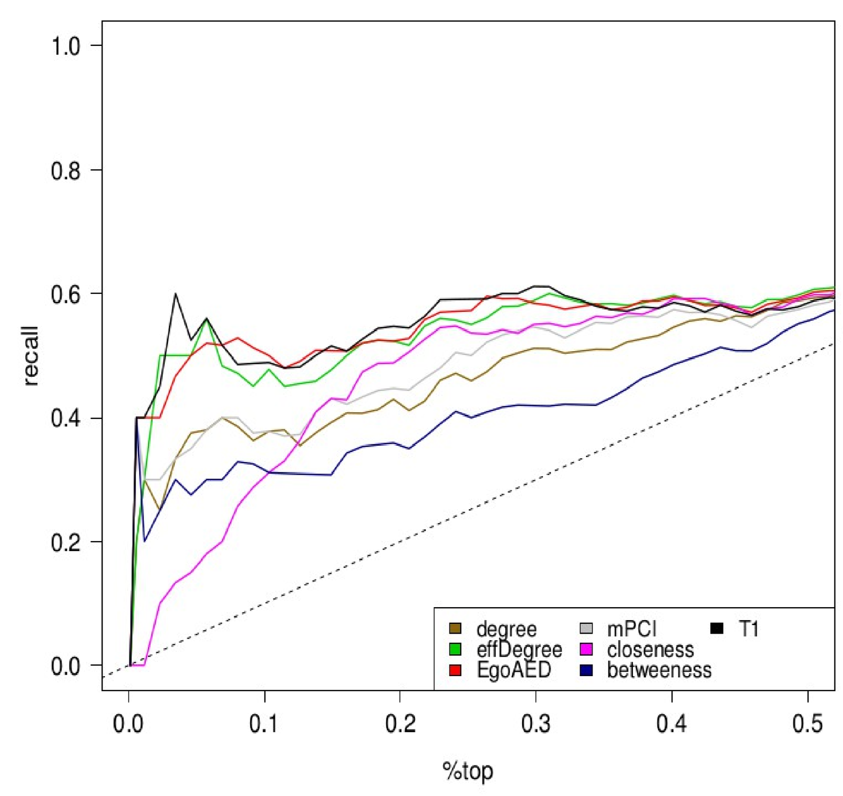
$$\rho = 1 - \frac{6 \sum D^2}{N\left(N^2 - 1\right)}$$

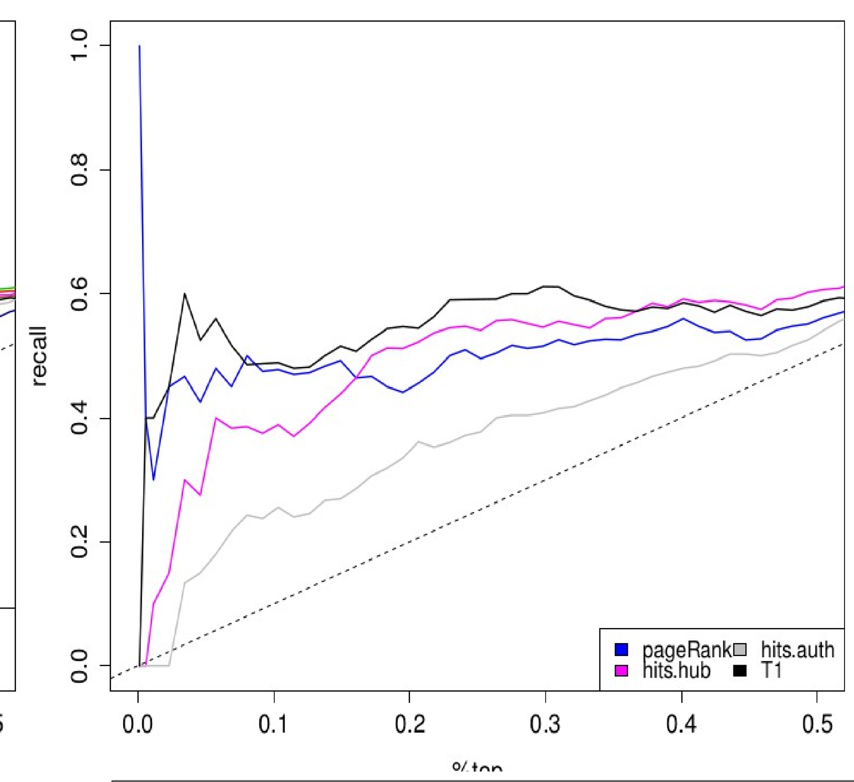
Metric	Spearman's ρ	absErr
Degree	0.1324	248.82
Betweenness	0.1346	266.42
Closeness	0.0935	245.43
μ-PCI	0.1104	246.32
PageRank	0.0935	251.03
HITS (auth)	0.1004	267.33
HITS (hub)	**0.1584**	**240.87**
Eff. Degree	0.1332	236.10
EgoAED	**0.1515**	**234.84**
T1	0.0891	247.89

Partial (top) Ranking Results

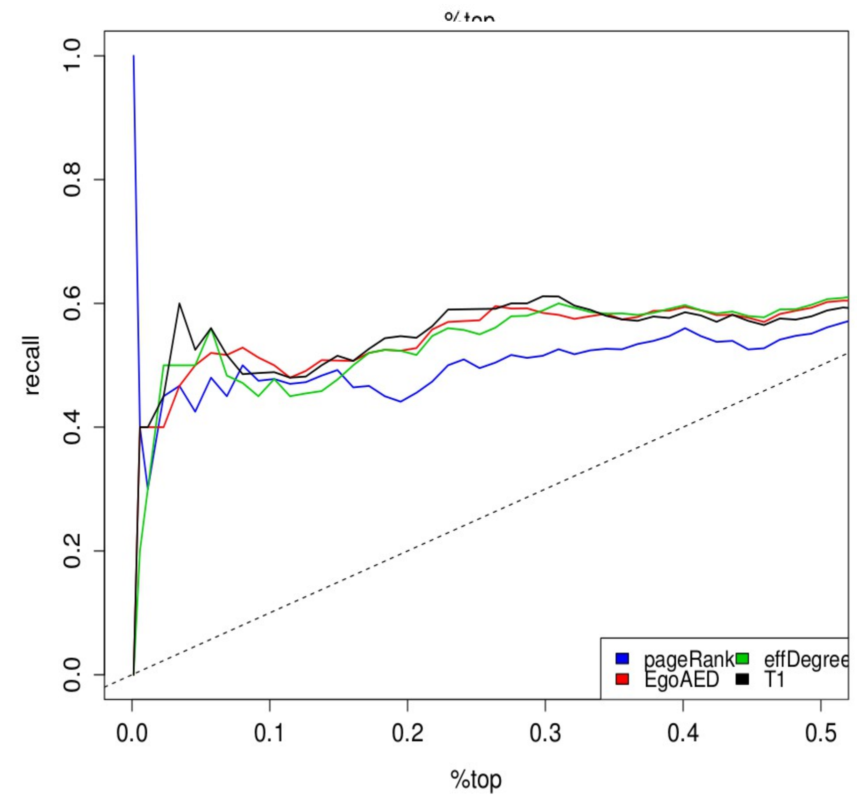
- Top-m recall

$$top_m \text{recall} = \frac{|\text{top}_m \text{nodes}(\text{refRank}) \cap \text{top}_m \text{nodes}(f)|}{m}$$





- Whilst all methods perform similarly when considering whole global ranking, differences among them appear when ranking the top influencers. For those, in general, the methods proposed here outperform the classical centrality measures.



Acknowledgments

This work has been funded by the Regional Government of Madrid (CM) under project Cloud4BigData (S2013/ICE-2894) cofunded by FSE & FEDER.

References

[1] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Identifying influencers on twitter. In Fourth ACM International Conference on Web Seach and Data Mining (WSDM), 2011.
[2] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. ICWSM, 10(10-17):30, 2010.
[3] Yi Chang, Lei Tang, Yoshiyuki Inagaki, and Yan Liu. What is tumblr: A statistical overview and comparison. SIGKDD Explor. Newsl., 16(1):21-29, September 2014.
[4] Sysomos Inc. Replies and retweets on twitter. http://www.sysomos.com/insidetwitter/engagement/, 2010.
[5] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNAKDD '13, pages 8:1–8:9, New York, NY, USA, 2013. ACM.
[6] Mark Newman. Networks: An Introduction. Oxford University Press, Inc., New York, NY, USA, 2010.