# Structural Evolution of Information based Online Social Networked Systems


**Kaj-Kolja Kleineberg[1*], Despoina Antonakaki[2**],**
Marián Boguñá[1], MÁ Serrano[1], Sotiris Ioannidis[2], Paraskevi Fragopoulou[2]
[1]Universitat de Barcelona – [2]FORTH ICS – *kkl@ffn.ub.edu – **despoina@ics.forth.gr


## Abstract


Information based online social networked systems like Twitter are reshaping the flow of information in today's world. The functionality of these systems is essentially different from friendship based networks like Facebook. We plan to investigate the structural evolution of information based social networked systems. To this end, we will gather data from Twitter and extract the topological features of the graph. Based on our findings, we will develop a model to identify the main mechanisms responsible for emergence of the observed topological features. The comparison of the topological evolution of information based and friendship based networks is expected to yield insights in the relationship between functionality and the mechanisms underlying the evolution of the respective system.


## Data from Twitter network

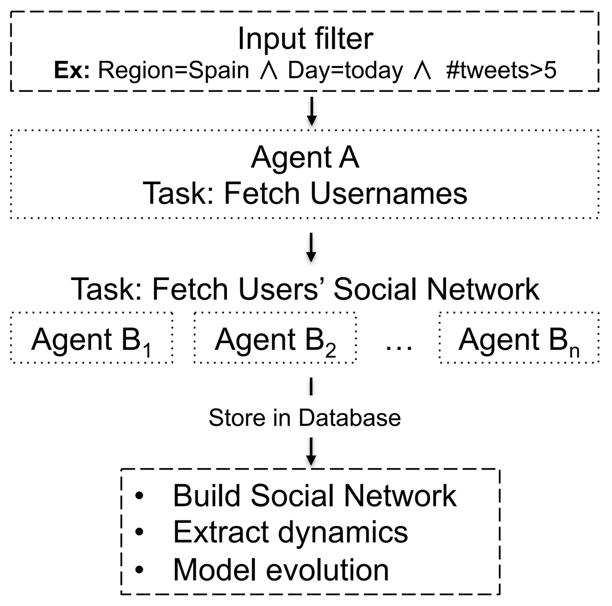


Fig. 1: Multi-agent data collection strategy

Our data collection strategy is to use multiple agents to query a portion of Twitter social network.
Initially, an agent queries Twitter for users that satisfy a preset filter (i.e. Region, Trends, tweeting patterns).
Then numerous (10-20) agents download asynchronously the followings and followers list of these users. The throughput of our approach is estimated to reach million users per day.
Querying the followers/followings list of the same users daily, will give significant insights of the evolution of Twitter's Social Network.

## Unique co-occurrences of posted URLs

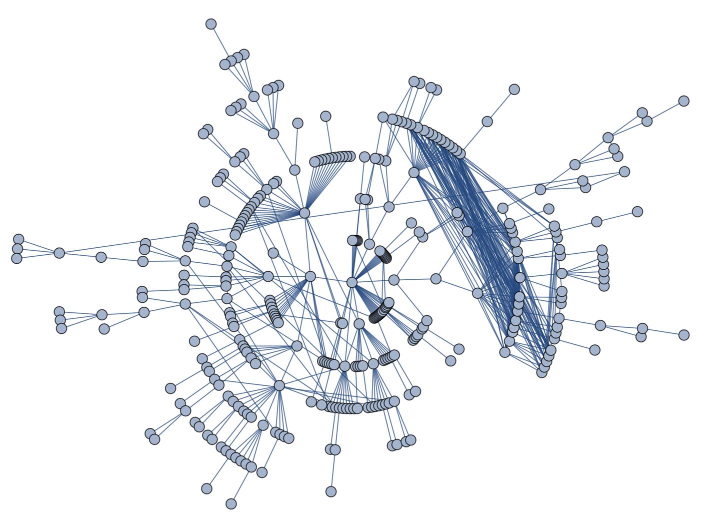

Fig. 2: Each node here is a user. Two users are connected if they share a common URL.

Fig. 2 shows the leading components of the one-mode projection of the bipartite networks of Twitter users and posted URLs where nodes are only connected if they both posted a unique URL, which means it was posted by one else in the dataset.
The complete network consists of 3779 nodes from which 336 form the largest connected component.
We notice that although most of users belong to small (<5 nodes) subclusters (not shown in the graph), there is a big cluster of connected users that resembles a following/follower visualization.

## Interesting social structures in Twitter

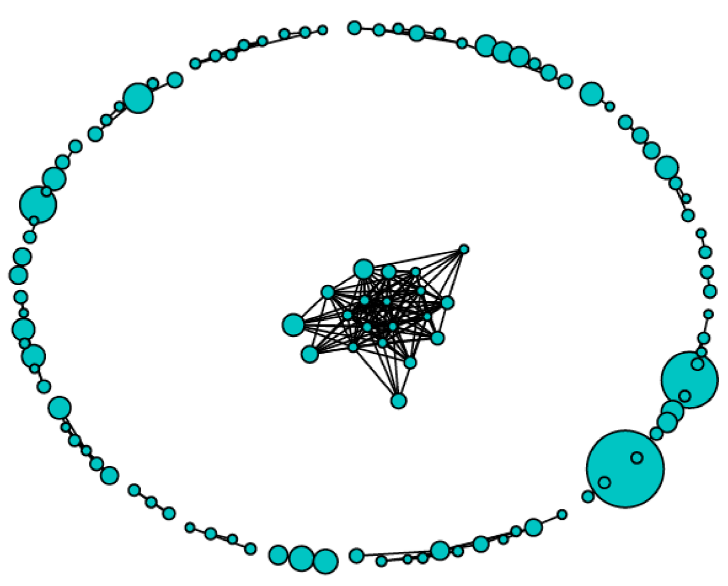

Fig. 3: Each node is a URL. The size of the node is proportional to the number of users that have posted this URL. Two URLs A,B are connected if the group of people that posted the URL A and the group of people that posted the URL B share the same elements with a factor of 50% or more. This indicates that the two URLs are semantically related.

Visualization of Twitter data can reveal interesting structures and provide indication of abnormal social behaviors. In Fig. 3 we plot URLs as nodes. Two URLs are connected iff the groups of people that posted them is similar.
The cluster of URLs in the center of the graph reveals a set of different URLs that have been posted by the same group of people.
The size of the cluster is unexpectedly high given the rest of URLs in the graph. Indeed a closer inspection of these URLs proved that they belong to the Get More Followers campaigns. This is a special class of spam that lure users to give access to their accounts for the benefit of acquiring more followers.
Similar visualizations can reveal not only suspicious activities but interesting social patterns as well.

## Connectivity structure of directed networks

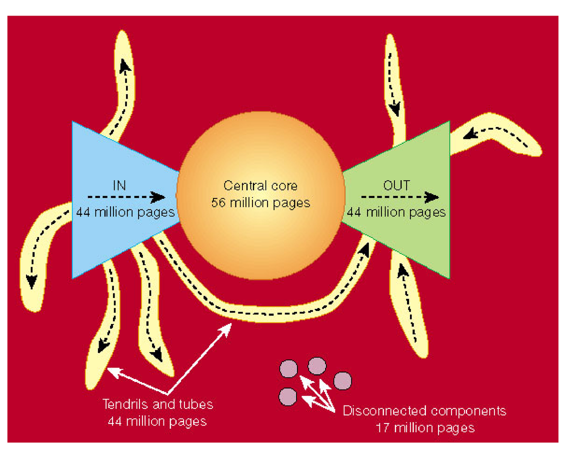


Fig. 4: Connectivity structure of directed networks. Figure taken from [1].

The connectivity structure of directed networks exhibits a "bow tie" structure [2]. Namely there are four main components: (1) a main core (popular/easily visited pages), (2) new pages linking to the core but not reachable (3) an 'out' cluster reached from core but not linking to it and (4) other 'tendrils' and 'tubes', connected only to in or out parts. Whether Twitter's users structure resembles a similar model is an open question.

## Connection preferences

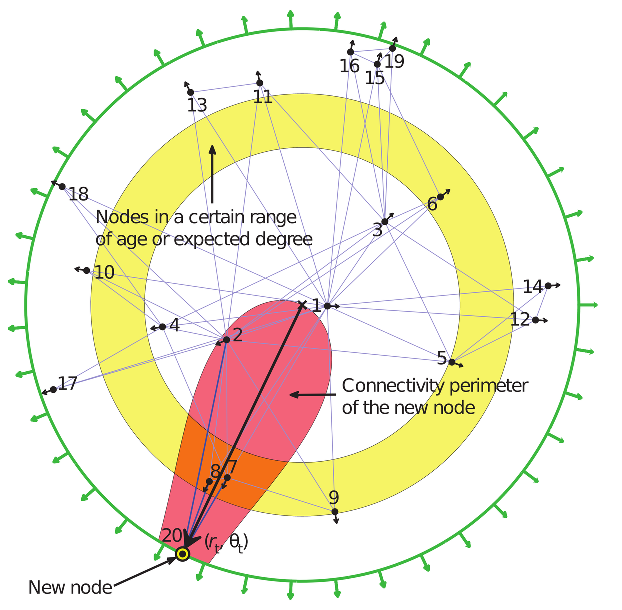


Fig. 5: Popularity versus Similarity in Growing Networks. Image taken from [3].

The idea that nodes optimize popularity and similarity of their connections [3] has proven to be successful to understand the topological features of many real world complex networks. However, this optimization requires that each node is aware of the properties of each connection candidate.
If the network is used primarily to obtain information, the information about the properties of the nodes will be transmitted on the network itself. The dynamics of information propagation (see next block) then allows individuals to obtain knowledge about the properties of other nodes. The previous optimization then can be generalized to the case of limited knowledge which acts as a constraint on the pool of potential connection candidates.

## Propagation dynamics

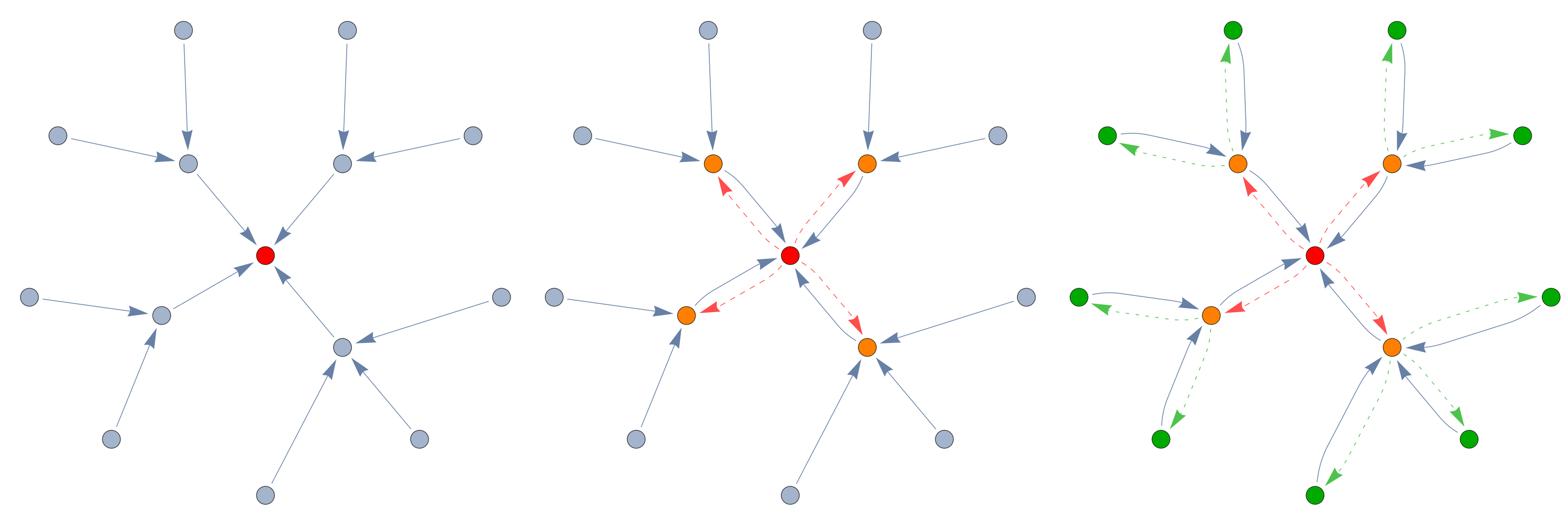

Fig. 6: Dynamics of information propagation.

The propagation of information emitted from one node is shown in Fig. 6. Depending on the topological distance from the emitting node, the probability of obtaining the message reduces. Interestingly, nodes can reversely follow the path of the information to obtain additional knowledge about the sender.

## Modeling framework

The modeling of such system naturally leads to the inclusion of memory effects. The individuals have to remember which nodes they are aware of and what their properties are. It could be convenient to model such system by including an additional layer, which we denote "meta layer", where directed weighted links represent the level of knowledge nodes have about each other. See Fig. 7.

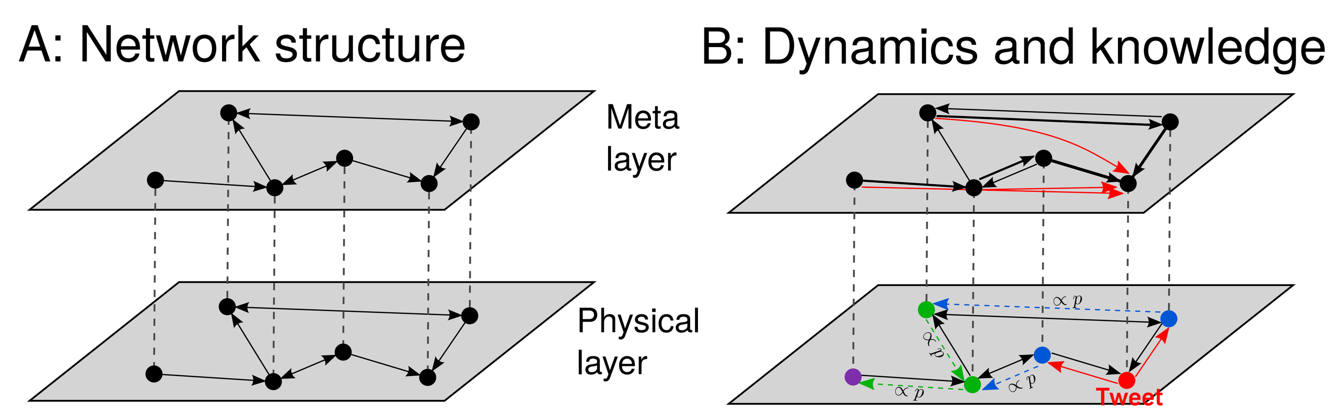


Fig. 7: Modeling of network awareness as additional meta layer.

## Research questions

We are planning to address the following questions:

1. How is the structural evolution of the Twitter network? How does the connectivity structure evolve?
2. Which are the underlying mechanisms for the topological evolution of information based social networked systems?
3. What are the main differences to friendship based networks [4]?

## Literature

### References


[1] The web is a bow tie. *Nature*, 405(6783):113–113, 05 2000.

[2] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Comput. Netw.*, 33(1-6):309–320, June 2000.

[3] F. Papadopoulos, M. Kitsak, M. Serrano, M. Boguñá, and D. Krioukov. Popularity versus Similarity in Growing Networks. *Nature*, 489:537–540, Sep 2012.

[4] Kaj-Kolja Kleineberg and Marián Boguñá. Evolution of the digital society reveals balance between viral and mass media influence. *Phys. Rev. X*, 4:031046, Sep 2014.