Imperial College London

Rumour Propagation and Identification of their Sources in Social Networks

Robin DUPONT - CID 01277640

Introduction

Online social networks have a gradually increasing importance in our daily lives. A huge part of human communications is made through social networks. If most of the information that are exchanged are inoffensive, some of them might be harmful. This is the case of the rumours, since the origin of a such piece of information is unknow. Hence, the aim of detecting the source of rumours appears clearly.

Why detect rumours sources ?

- Check the source of a piece of information. (useful for everybody)
- Identify potential influencers. (useful for the companies)
- Find peoples responsible for psychological harassment. (useful for the police)

Aims

Provide a detection algorithm making use of time information provided by a finite amount of monitoring node, when several rumours are propagated by a single source.

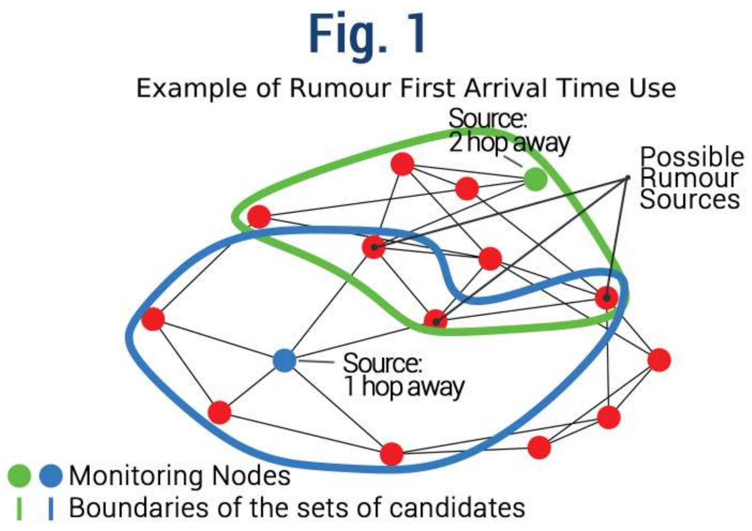
Methods

Around **10%** of the nodes in the graph are **monitoring nodes**. They report the precise time of their infection for each rumour.

• 1st Step: Reduce the number of candidates

Based on the earliest arrival, it is possible to deduce a maximum distance for the source. If the first rumour arrives at step k at a monitoring node, then the source cannot be more than k hops away from the monitoring node. Thus, a set of candidates per monitoring node can be established. By intersecting the different sets, a smaller set of candidates can be derived. An example is shown on **Fig. 1**

Fig. 1



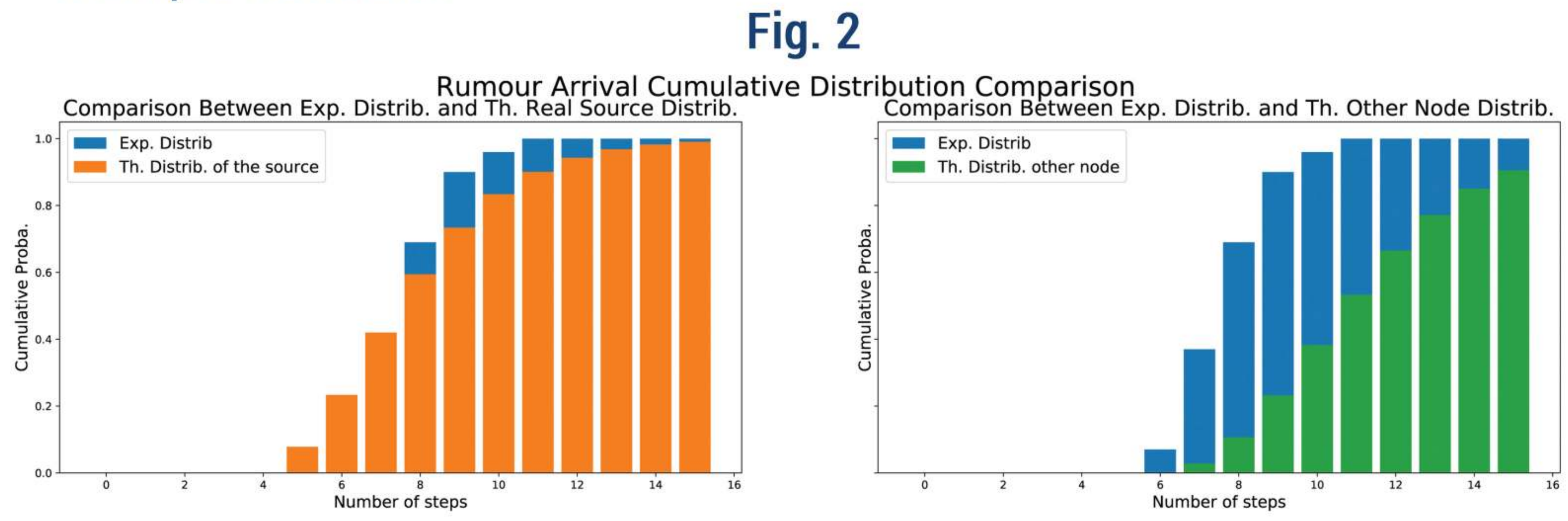
• 2nd Step: Rank the candidates

The ranking of the candidates is based on the similarity between the cumulative distribution of the rumours arrival steps and the theoretical derivation of this distribution. The following formula is used:

$$P_k^n = \sum_{i=0}^{n-k} \binom{n+i-1}{i} p^k (1-p)^i$$

P_k^n is the probability that a node is infected by step k if it is n hops away from the source, and p is the probability of rumour propagation. The L_2 norm is used. The possible sources are ranked according to the similarity of their distributions with the experimental one which is derived from the time of arrival of the multiple rumours.

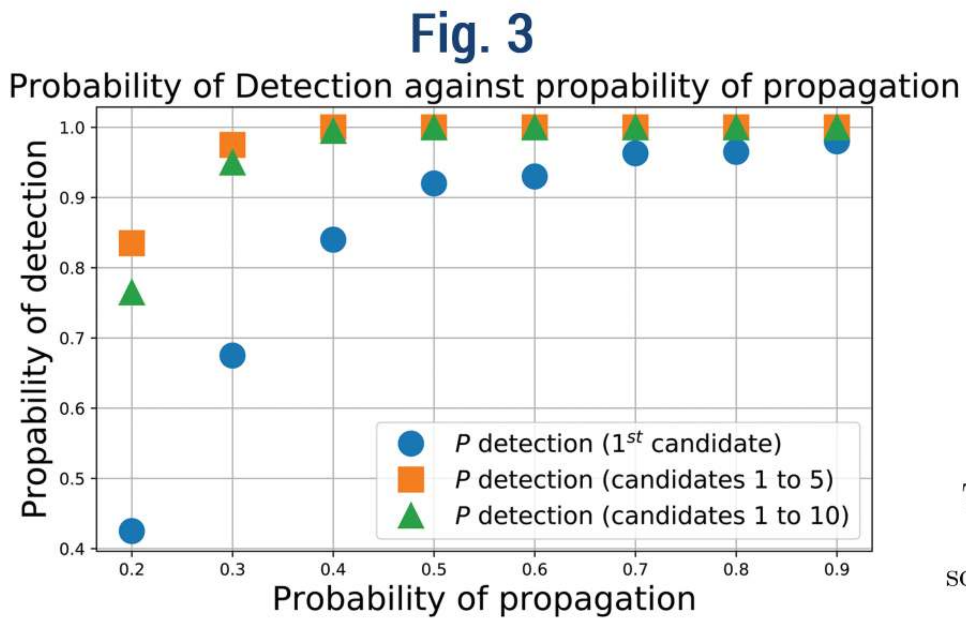
Fig. 2



Results

Simulations have been realised on **small world networks** [1]. The parameters were the following: 250 nodes, 25 monitoring nodes, probability of rewiring: 0.3. The results are averaged over 250 trials. It is possible to identify the rumour sources **more than 80% of the time** when the probability of propagation p is at least 0.4, and there are 100% chance that the rumour sources is in the first 10 and even the first 5 candidates (see **Fig. 3**). Moreover, the 1st candidate is always close to the real source in the network: **no more than 2 hops** away in the worst case (see **Table 1**). The average distance between 2 nodes for these networks is 4.8, which is slightly lower than the experimental value of 6 for the real world [2].

Fig. 3



p	avg. rank	avg. d to real source.
0.2	8.29	1.345
0.3	1.265	0.605
0.4	0.19	0.27
0.5	0.055	0.105
0.6	0.007	0.105
0.7	0.005	0.075
0.8	0.002	0.69
0.9	0.002	0.053

Table 1: Average rank of the real source and distance from the 1st candidate to the real source, for different probability of propagation.

In order to keep a good detection probability, the percentage of monitoring node should **not be lower than 5%** of the total number of nodes (see **Fig. 4**). The algorithm works best on small world graphs (see **Fig. 5**).

Fig. 4

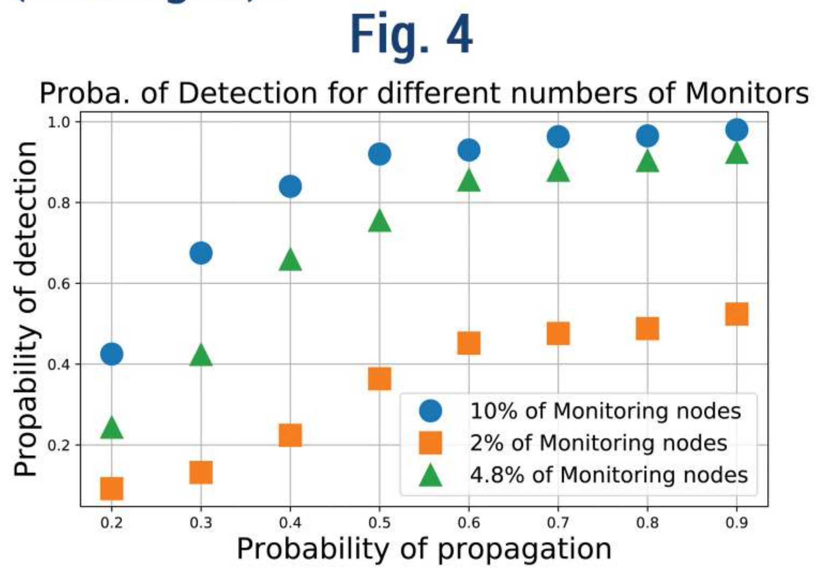
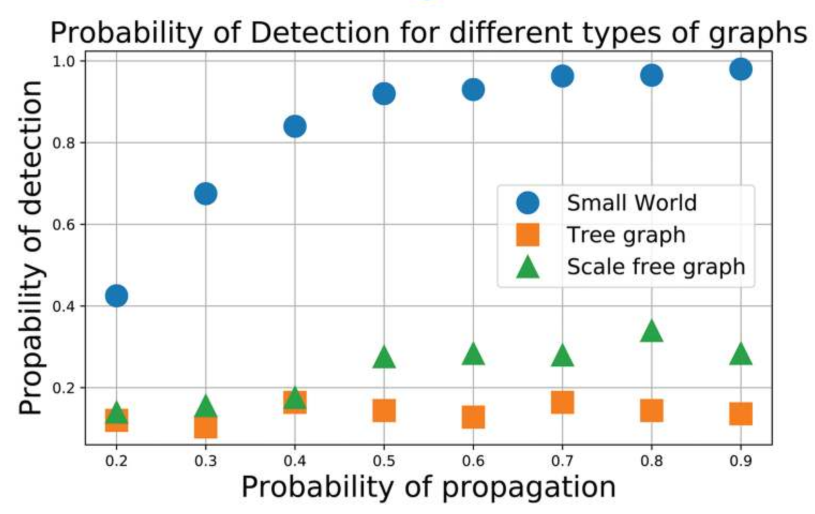


Fig. 5



Conclusion

- The developed algorithm allows detection of the source with good probability provided that the rumours spreads fast enough, typically when p is greater or equal to 0.4.
- It can be assumed that the real source of rumour is almost always in the first 10 and even first candidates.
- The first candidate, even though it is not always the real source, is close, topologically speaking, from the real source.
- The algorithm yields significantly more accurate results on Small Workd Graphs.

Futur Work:

The propagation simulation could be done with triadic closure [3] instead of only propability, which would model better the real human behaviour.

References

[1] D. J. Watts and S. H. Strogatz. «Collective dynamics of 'small-world' networks». *Nature*, 1998.
[2] S. Milgram. «The Small World Problem». *Psychology Today*, 1967.
[3] D. Easley and J. Kleinberg. Nertworks, Crowds, and Markets: Reasoning about a Highly Connected World. (pp 48-50). Cambridge University Press, 2010.