Credibility Network for Online Social Network

Marialaura Previti First Year Ph.D. course activities

Tutor: Prof. Ing. Vincenza Carchiolo

Ph. D. Course in Systems, Energy, Computer and Telecomunications Engineering - XXXII Cycle

ISSUE

Social networks have become popular over last decade, because people use it to keep in touch with friends, family and colleagues or, more generally, with interesting people, chat with them and exchange news and contents in real time. For a part of population they become substitutes of traditional mass media, because posts spread faster than printed paper and television news, so also traditional mass media recently use it to release instant news and government agencies publish on it official announcements. In addition to these authoritative sources, there are many others that enter contents of questionable truthfulness.

The absence of an authority that is responsible for assessing the truthfulness of the post allows users' improper behaviors, hence it is necessary **to model users activities in social network** and **to identify methodologies that allow the detection of hoaxes and defamatory contents automatically**.

PROPOSAL

An online social network is a structure made up of social actors and links among them that represent social interactions.

Social networks can be modeled as a graph where each pair of persons, represented by nodes, is connected with an edge if they interact each other, hence the graph is an **acquaintance network**. In particular, I focus on **undirected scale-free networks**, because in typical social networks there are few nodes with many contacts, whereas most have a medium-low number of contacts.

I use a **Susceptible-Infective-Removed (SIR) Model**, typically exploited to describe how disease spread over a contacts network, to describe how news spread over an online social network.

In this model, I consider a population consisting of N individuals which, with respect to the news, are divided into:

- **ignorants**, i.e. people who are unaware about the news;
- **spreaders**, i.e. people who are already aware about the news and intend to share it with others;
- **stiflers**, i.e. people who are already aware about the news, but have no interest in spreading it.

As in real social relationships occur, I introduce the credibility between two nodes that indicates to what extents a node trust another one. To represent this, I created a **directed network**, called **credibility network**, with the same nodes of acquaintance network and, for each undirected edge of acquaintance network, two edges of the credibility network, one for each direction, because the end nodes of an edge do not trust each other the same way.

To measure the credibility, we use a parameter **Cr**, representing the **direct credibility**, whose value ranges in [0,1], in particular:

- if it is **close to 0**, the source node believes that the target node provides mostly **false news**;
- if it is **close to 1**, the source node believes that the target node provides mostly **true news**.

At start time, since we do not have any information about past node activities, all the credibilities are set to 0.5, that is the **neutral credibility**, i.e. the confidence that each individual assigns to other individuals he has never been related before.

Post Sharing-Based Credibility Network For Social Network

In this work, the abovementioned **acquaintance network** is a **multislice network**, where each slice represents the diffusion of a news that only involves active nodes at the time of its publication, hence **not all the nodes are present in every slice**.

In each slice, the news spreading is performed if an individual behave as spreader and make visible a news to his neighborhood for a certain amount of time and if in this time interval some neighbors are active and decide to repost a news.

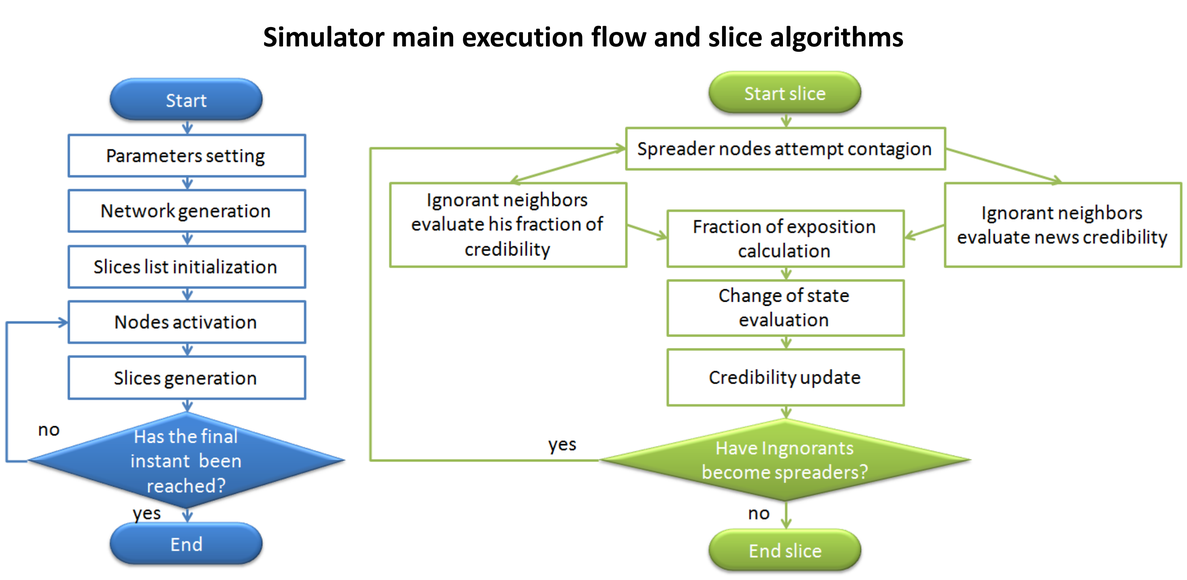
Among many factors affecting credibility, I consider only:

- the **credibility of posts published**;
- the **neighborhood activity**;
- the **propensity of each individual to repost true news**.

If an individual decides to repost a news is likely he considers the news is true and the spreader neighbors are reliable, instead if he does not repost it, the vice versa is true. Furthermore, some individuals are more careful in news contents evaluation, while others repost news without checking their sources or generate false news deliberately.

To evaluate the proposed model, I implemented a simulator, the purpose of which is to verify whether an individual's credibility is directly proportional to his ability to repost true news.

Simulator main execution flow and slice algorithms



Conclusion: individuals with high attitude to repost true news receive high credibility whereas those who post false news receive low credibility.

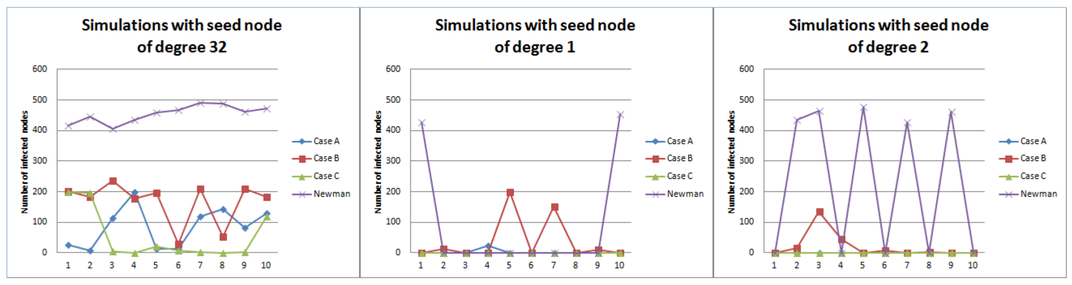
Introducing Credibility To Model News Spreading

In this work, I perform an analysis of online social networks users behaviors to explain why it is necessary to introduce a parameter representing **direct credibility** in **transmissibility formula of Newman's epidemic spreading model**, in order to use it to model news diffusion in online social network.

$$T_{ij} = 1 - e^{-\beta_{ij}\tau_i C r_{ji}}$$

I generated a network of 1000 nodes with the features outlined in proposal section and I ran groups of 40 simulations where a chosen node seed node spreads a news over the network. No credibility updates are performed. The 40 simulations are arranged as follows:

- 10 simulations with the credibility of all edges set to 0.5 (**case A**);
- 10 simulations with the credibility of incoming edges in the seed node set to 1 and others set to 0.5 (**case B**);
- 10 simulations with the credibility of incoming edges in the seed node set to 0.1 and others set to 0.5 (**case C**);
- 10 simulations with the **Newman model**.



Various attempts were made with nodes of different degrees, in order to display the ability to propagate a news of well-inserted nodes and semi-isolated nodes.

For **nodes with high degree**, with **a medium-high credibility** (**cases A and B**), the seed node finds the way to propagate the news to neighbors and since all the other nodes have neutral credibility, after the first level of contagion the news can reach almost the same number of nodes, instead, in the case of **low credibility** (**case C**), propagating the news becomes harder for the seed node since it is able to persuade a smaller number of neighbors to the repost it, but if and only if another hub node is reached by the news, this will be propagated as in the previous two cases thanks to the neutral credibility of its neighbors.

For **nodes with low degree**, credibility and contact rate play a significant role, in particular if **credibility is medium-low** (**cases A and C**) it will be very difficult to convince the few neighbors to propagate the news whereas if **credibility is high** (**case B**) it all depends on the contact rate; despite the high credibility indeed, news cannot be propagated if it is not displayed by neighbors.

Conclusion: The news propagation does not depend only on the first spreader's adjacent nodes, but also on neighbors of neighbors who propagate the news, hence, without any credibility updates of all the nodes involved in spreading processes, news will continue to propagate over the network.

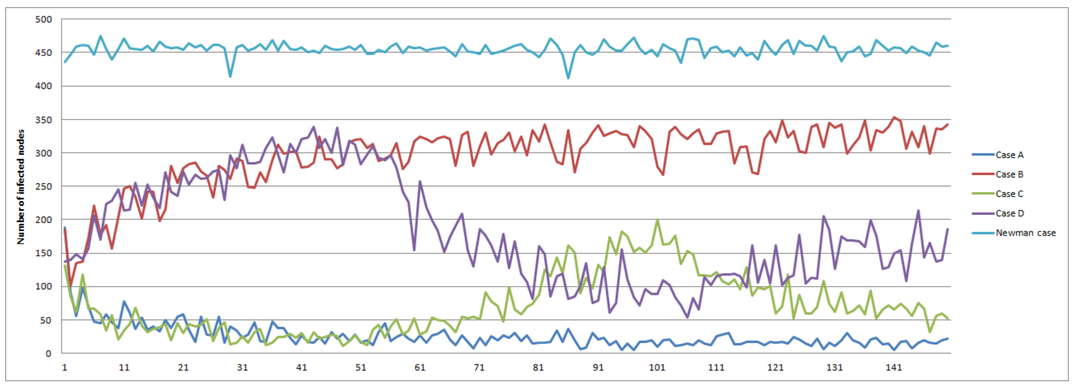
A Trust-Based News Spreading Model

In this work, I explain how is possible that news spread by an individual will change his credibility, and, in turn, his credibility will change his ability to spread the news over time. Starting from previous work modified transmissibility formula, I introduce a mechanism for updating local credibilities in credibility network that take into account the freshness of news, giving progressively greater weight to recent news and lower weight to old ones, in order to allow individuals to quickly modify their neighbors' trust, if they chance their behaviors over time.

$$Cr_{ji} = \frac{\sum_{x=0}^{n}(x+1)C_{news_x}}{\sum_{x=0}^{n}(x+1)}$$

I generated a network of 1000 nodes with the features outlined in proposal section, in which a well-connected seed node behaves as a spreader of 150 news. I performed five groups of 10 different simulations where news are so divided:

- all news are false (**case A**);
- all news are true (**case B**);
- 50 news are false, other 50 are true and the last 50 are false (**case C**);
- 50 news are true, other 50 are false and the last 50 are true (**case D**);
- all news are propagated with **Newman transmissibility formula**.



Starting with neutral credibility, all simulations have a similar starting point. In **case A and B**, the change of credibility affects the ability to propagate news after just few inserted news. In **case C and D**, behavioral modification affects the ability to propagate the news after only few news whose truthfulness is different from the previous one and, as in real social network, losing credibility is easier than achieving it again.

Conclusion: Results shown that the model shapes social media users real behavior for what concern credibility dynamics and news propagation.

FUTURE WORKS

I will refine the model distinguishing also the case in which some nodes are inactive in one or more temporal steps during which the news is visible on social network from the case in which nodes read that news but decide to no repost it. I also aim at validating the model on larger networks to apply it on real social media data.

REFERENCE

Newman, M.E.: Spread of epidemic disease on networks. Physical review E 66(1) (2002) 016128