11th International Symposium on Intelligent Distributed Computing
11 – 13 October 2017
Belgrade, Serbia

Post sharing-based credibility network for social network

V. Carchiolo, A. Longheu, M. Malgeri, G. Mangioni, **M. Previti**

Dipartimento di Ingegneria Elettrica, Elettronica ed Informatica (DIEEI)
Università degli Studi di Catania, Italy

Outline

- Goals and motivations
- Our model
- Simulator
- Results and conclusion

Goals and motivations

- We propose a general model that takes into account both **posts and users' credibility**, using a duplex network of acquaintances and credibility among users.
- **Social networks** are intensively and extensively used to exchange news and contents in real time. The lack of a global authority for assessing posts truthfulness however allows malicious to exhibit unfair behaviors, therefore identifying **methodologies to detect hoaxes and defamatory contents** automatically is more and more required.

Outline

- Goals and motivations
- Our model
- Simulator
- Results and conclusion

Epidemic model: SIR

- We 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.

Duplex network layers

Our network consists of two layers:

- **acquaintance network A**
- **credibility network C**

Acquaintance network

- The acquaintance network is a multislice network

$$A = (N, E, S) \qquad S = \{S_1, S_2, \ldots, S_H\}$$

- N is the nodes set;
- E is the edges set;
- S is a family of subgraphs (or slices) of A
- each **slice S_x** represents the diffusion that a news x had on the acquaintance network.

Nodes interaction

- An individual i behaves as spreader, inserting a new post at time t;
- he makes it visible to his neighborhood U(i) up to a time $t + \Delta t$;
- if an individual $j \in U(i)$ is active and repost the news in this time window, he established a contact e_x;

Credibility network

- The credibility network is a directed network.
- At start time, this network consists of all nodes of the acquaintance network.
- At **each undirected edge** of the acquaintance network **correspond two directed edges** of the credibility network, one for each direction.
- Each edge is associated with a **direct credibility**, i.e. the confidence that a node has toward his neighboring node.

$$C_{D_{j->i}} \in [0,1]$$

- The starting credibility is 0.5, because there is no information about reputation of the nodes yet.

Factors affecting credibility

- **Credibility of published posts**

$$d_{news}^{S_x} \in [0,1]$$

- **Neighborhood activity**

$$d_{U(i)}^{S_x} \in [0,1]$$

- **Propensity to repost true news**

$$pt_j \in [0,1]$$

Fraction of exposition to the news

$$P_j^{S_x}(t) = \frac{\alpha d_{news}^{S_x} + \beta d_{U(j)}^{S_x}(t)}{\alpha + \beta}$$

- Where α and β are weighed parameters.

Neighborhood influence

$$P_j^{S_x}(t) = \frac{\alpha d_{news}^{S_x} + \beta d_{U(j)}^{S_x}(t)}{\alpha + \beta}$$

$$d_{U(j)}^{S_x}(t) = \frac{\sum_{i=1}^{N} a_{ij}^{S_x} X_i^{S_x}(t) C_{D_{j\to i}}(t)}{\sum_{i=1}^{N} a_{ij}^{S_x} C_{D_{j\to i}}(t)}$$

- Where $a_{ij}^{S_x}$ are the elements of adjacency matrix of slice Sx, $A_{ij}^{S_x}$.

State of node *i* at time *t*

$$d_{U(j)}^{S_x}(t) = \frac{\sum_{i=1}^{N} a_{ij}^{S_x} X_i^{S_x}(t) C_{D_{j\to i}}(t)}{\sum_{i=1}^{N} a_{ij}^{S_x} C_{D_{j\to i}}(t)} \qquad P_j^{S_x}(t) = \frac{\alpha d_{news}^{S_x} + \beta d_{U(j)}^{S_x}(t)}{\alpha + \beta}$$

$$X_i^{S_x}(t) = \begin{cases} 1, P_i^{S_x}(t-1) > pt_i \\ 0, P_i^{S_x}(t-1) \leq pt_i \end{cases}$$

Direct and local credibility

$$C_{D_{j\to i}}(t) = \frac{1}{n_i}\sum_{x=1}^{n_i} P_j^{S_x}(t)$$

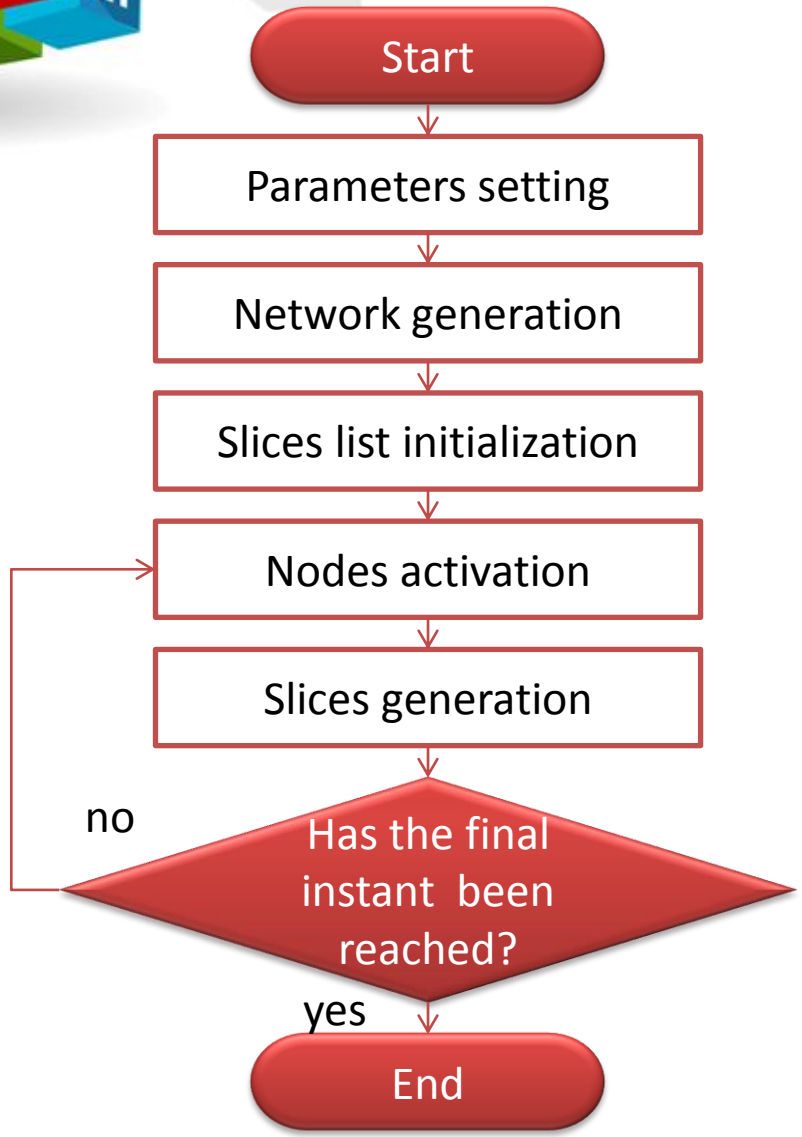
- Where $n_i \in [0, H]$ is the number of spread news from node i.

$$C_L(i) = \frac{1}{k}\sum_{n=1}^{k} C_{D_{i\to j}}(n)$$

Outline

- Goals and motivations
- Our model
- Simulator
- Results and conclusion

Simulator: Main execution flow



Simulator: initial step

- **Parameters setting** (number of network nodes, initial and final instants of observing time window, propensity to repost true news of each node);
- **Duplex network generation** (scale-free network);
- **Empty slices list generation**;

Simulator: simulation step

For each time slot:

- **Turn on or off of each node** (nodes are not active every time);
- **Some active nodes generate new news** and simulator generate relative slices with seed node as spreader $(X_i^{S_x} = 1)$ and other active nodes as ignorant $(X_i^{S_x} = 0)$;

News spreading process

Start slice

Spreader nodes attempt contagion

Ignorant neighbors evaluate his fraction of credibility

Ignorant neighbors evaluate news credibility

Fraction of exposition calculation

Change of state evaluation

Credibility update

Have Ingnorants become spreaders?

yes

no

End slice

Simulator: news spreading step

In each slice:

- Seed node tries his **neighbors contagion**;
- Neighbors use **fraction of exposition formula** to decide if repost or not the news;
- The **credibility** of edges with nodes that have changed status **are updated**;
- New infected nodes try their neighbors contagion with the same mechanism since no new contagion are performed.

Outline

- Goals and motivations
- Our model
- Simulator
- Results and conclusion

Simulation results

- In a good model, it would be desirable that the **individual's propensity to post true news (*pt)*** is as close as possible to the **credibility of the corresponding node (*C_L*)**.
- Five simulations were carried out with **100 nodes** and an observation time window of **100 time slots**.
- We calculated the difference of the two abovementioned values for each node and we counted the corresponding nodes by inserting them in bands of 10% of the difference.
- The nodes that did not share and repost any news during the observation window were excluded from counting (0-post nodes).

Simulation results

	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5
No. News	532	476	498	512	458
0-post nodes	2	2	3	4	7
0-10%	26	24	31	25	22
10-20%	25	20	21	24	24
20-30%	14	20	16	17	17
30-40%	16	21	13	14	16
40-50%	12	12	14	9	9
50-60%	5	1	1	6	4
60-70%	0	0	0	1	1
70-80%	0	0	0	0	0
80-90%	0	0	0	0	0
90-100%	0	0	0	0	0

In each simulation over 60% of the nodes have the aforementioned difference between credibility and propensity to share and repost true news less than 30%.

Conclusion

- We proposed a model for the assessment of posts truthfulness in social networks.
- To this purpose, **we created a credibility network starting from an acquaintance network using both the news contents and the fraction of each node neighbor reposts**. The latter parameter is, in turn, influenced by credibility acquired in last time slot, hence in our model the history of the network is also taken into account.
- Some preliminary results have shown that **individual with high attitude to repost true news receive high credibility whereas those who post false news receive low credibility**, therefore the proposed model agree with real behaviors.
- In future work, we will validate our model on larger networks and larger time windows in order to apply it on real social network.

Thank You For Your Attention!