ICM 49144

Team #49144

January 31, 2016

Abstract

The goal of an abstract is to give an overall description of the paper. We should illustrate these points:

1)A little background of the problem.
2)Our basic model
3)Our new model, namely what is the improvement of this model compared with those previous ones. Some descriptions about the weakness of previous models(which are what we are trying to solve).
4)The good results we get from our new model.
In general, the abstract must illustrate what we have done in this paper and the significance of these improvements(usually by showing those good results we have).

Contents

1 Introduction

Understanding process of information getting around society network has attracted increasing attention. Information flow in complex society network is of paramount importance when trying to solve the

1.1 literature review

In previous research, scientists usually base their model on the well know susceptible-infected-recovered (SIR) model. The advanced methodology of SIHR [1], the further developed SEIQR model[2], inspired other research of this kind, therefore we can find some relatively mature research like the one by $LinyuanLv$ and $Duan-BingChen$, which we based on to obtain our result. The basic idea of their's is to change the continuous SIR model into a discrete-time model to resemble the spreading of news, and introduce four similar situation of node to represent the attitude of people towards news to describe the probalistic spreading of news.

$--------------workinthemodelpart---------------$

Furthermore, there exist a couple of applications of the model or the variants of it, including the description of growth and decline of network service[2], predicting forest fire spreading[3]. But when taking into consideration the details of the real world society, we find there are still things left to be discussed.

The nodes, links, and the information itself differs from that of a pure network.

- We can introduce a memory effect to resemble the actual memory loss[4].
- The information itself, or rather the value of it, determines to which degree people tend to believe it. This is discussed in this paper.
- A link may be strong or weak, and the speed of information getting around differs.

The topology structure is distinguished in the following aspects.

- The change of link between nodes, namely the dynamic social network decribed in previous work[5]. It represents people making new contacts and lose touch with old friends.

- The abidance of $SixDegreesofSeparation$[6] is missed in previous work during simulation.
- The difference of structure entropy, which makes such much difference that the Small World Structure, Random Structure and Regular Structure distinguishes from each other a lot[7]. Furthermore, we witness a change of that when taking into account the historical trend of society, which we will discuss in our work.
- The connection between the nodes are not to be simply described as non-directional link. Some links are . Data could be collected by crawling the real world network[8], and we will give it a thorough consideration in this paper.

Despite of the previous research, there are essential points neglected such as the involvement of social media

$$somethinghere$$

, the change of topology structure and the historical explanation of it, the inherent value of information as mentioned above.

2 Assumptions and Justifications

This is a very important section. In this part, you should give all the basic assumption under which you build your model. Every model has its applicable range, so you must figure out what are those assumption at the beginning of building a model. Those assumption should be inspired and concordant with our daily intuition. The form of assumption and justification should be more or less like this:

- **A sentence of assumption** A sentence of justification. Usually, this is a narrative of an intuition from which you derive your assumption. You should depict this intuition and its connection with the assumption.

3 Notations

All the variables and constants used in this paper are listed in **Table 2** and **Table ??**.

Table 1 Symbol Table–Constants

Symbol	Definition	Units
	Constants	
$m_i(t)$	The number of times the i^{th} node has heard the news until time t	Local
	The confidence of i^{th} node towards the news until time t	
Unknown	The node has not yet heard the news, waiting for news	Local
Known	The node has heard the news at least once and has a probability to approve	Local
Approved	The node will spread the news to its neighbor and then shift to Unconcerned state	Local
Unconcerned	The node will not participate in any procedure	Local

Table 2 Symbol Table–Constants

Symbol	Definition	Units
	Constants	
λ	The probability a node will approved when it first hears the news	Local
λ_1	The inherent value of the news	Global
λ_2	An individual's bias towards news	Local
b	The degree of trust in a community (network)	Global
T		Global
c_i	The authority of the i^{th}node	Local
w_{ij}	The weight of edge e_{ij}	Local
t	Time-step	Global
v	The spreading speed of news from media	Global
a_M	The likelihood an individual get information from media	Local
a_F	The likelihood an individual get information from famous people	Local
a_N	The likelihood an individual get information from normal people	Local

4 Model

4.1 Basic Model

4.1.1 Overview

This is a fundamental model which will give us some insight into the network. Inspired by [**?**], we use N nodes to simulate N identical users in reality and each node stochastically links to other k nodes. When a given event occur at a node in this graph, we use a dynamic probabilistic information propagation model (DPIP), discussed in details as follows, to simulate the spreading of informations. This model only depicts the case that all users are the same, with no existence of famous people or huge media company.

$Ourmodelconsistingoffourstates, describethetransferringofstatestomakeitclear!!!!ieafterKnon$

4.1.2 Assumption and Justification

- **All nodes are identical** In this model, we do not consider famous people or huge media company, therefore, all nodes equally important in the sense that they all have k edges connecting with them.
- **Each node has four states** Since we use a node to represent a person, a node has four states corresponding with a person has four attitudes towards an information, 'Unknown', 'Known', 'Approved' and 'Unconcerned'.
- **Transferring from 'Known' to 'Approved' is a probabilistic event** Once a node heard a piece of news from its neighbor, it has a probability to transfer from 'Known' to 'Approved'. The more times it hears the same news, the higher probability it has to approve this news.

4.1.3 Methodology

Our model is a time-discrete network model. We have a set of laws which update the parameters of the model every time-step. Those laws are based on previous assumptions and observations. In this part, we formalize those laws in a mathematical way, and explain how they could update parameters from time t to time $t + 1$.

Step 1: **Initializing the network**

(a) Randomly constructing the network
Our network is a randomly connecting network. It has N nodes. Then we randomly connecting nodes with edges and make sure that, for each node, there is always k edges connecting with it.

(b) Initializing the states
To begin with, every nodes should be unknown to a piece of news except the node which originates this news. Therefore, we set every nodes in our network to the 'Unknown' state.

(c) Generating news

We randomly choose a node to be the news producer, transferring its state from 'Unknown' to 'Approved'. This indicates that this node approves this news and is willing to tell its neighbors. The other nodes remain in the 'Unknown' state.

Step 2: **Propagating information**

(a) Spreading news

We traverse all nodes and find out those nodes with state 'Approved'. For each 'Approved' node, it is willing to spreading this news to its neighbors. If the neighbor is unknown about this news, it shift to known states. If the neighbor is already in known state but not approved state, it will memorize this event and has higher probability to transfer to approved state in later time.

Formalizing this step mathematically, we denote $m_i(t)$ to be how many times the i^{th} node has heard this news from its neighbor until time t. For obvious reason, this $m_i(t)$ could also be explained as belief. For each 'Approved' node A in our network, we manipulate the states of all its neighbors. Suppose node B is one of its neighbor. If B is in the state 'Unknown', we set B to the state 'Known' and let

$$m_i(t+1) = 1.$$

If B is already in the state 'Known', we remain it at 'Known' and let

$$m_i(t+1) = m_i(t+1) + 1.(1)$$

If B is in the state 'Unconcerned', we do not make any change to B. After doing this to all neighbors of A, we change its state to 'Unconcerned'.

(b) Transferring from Known to Approved

Based on our observation, once an individual hears a piece of news, he will choose to believe in or not depending on how many times he has heard it. The more times it is, the more willingly he is to trust it. Therefore, we use a probability model to describe this phenomenon.

For each 'Known' node, if it hears the news from its neighbor at time t, it has a probability to approve this news. Denoting $P(m)$ as the probability of a node transferring from 'Known' to 'Approved', we have

$$P(m) = (\lambda - T)e^{-b(m-1)} + T,$$

where m is $m_i(t)$ and λ, T, b are network parameters.Once a node transfers to 'Approved', it will spread the news at next time-step, namely, at time $t+1$. We could also observe that the larger m is, the larger $P(m)$ is; this is in accordance with our previous observation. We need to mention that if a node does not hear news from its neighbors at time t, it will definitely not transfer to 'Approved', even though it may have heard the news once before $t-1$. This is in accordance with our intuition that one will not abruptly change his mind as long as he does not hear anything new.

(c) Iterating
In previous steps, we have updated states and $m_i(t+1)$ for each nodes. By iterating those previous two steps, we could simulate the propagation of information.

4.2 Improved Model

4.2.1 Overview

Our new model is a hierarchy, 3-layered, directed, weighted network, with belief cumulated and probabilistic transferring mechanism.

4.2.2 Modifications

In this section, we compare our new model with the previous one and illustrate the modifications explicitly.

- **Constructing the network as a directed graph.**In the basic model, we only construct a undirected graph; information could flow either from A to B or B to A. However, to consider a finer structure, we need to use a directed graph. Followers on Twitter and receivers of media are both good examples to support our choice.
- **The effect of famous people.** In the basic model, we treat every individual equally. Nevertheless, in reality, there are famous people who are more significant to the social network, i.e. a people may have a large number of followers on Twitter and hence are more influential. In this new model, we allow a fraction of nodes to have much more edges to represent those significant roles in network. There are a huge amount of edges directed from them to the normal users, meaning that they have lots of followers and news could flow from them to lots of people. These nodes also have four states, and they use the same probabilistic transferring mechanism as other nodes.

- **The effect of media.** In the basic model, there is no media or media companies. In the new model, we take them into account. We add one single node to represent the media, including newspapers, radios, . As for the structure of edges, there are edges from media node to every other nodes in the network, namely media permeates in everyone's life. There are also edges from every other nodes directing to media, meaning that media could gather information from all over the world.This media node also has four states, and it uses the same probabilistic transferring mechanism as other nodes.

- **Every node has a authority value.** We introduce the concept of authority. In real world, we tend to believe the informations telling by a reliable person or a authoritative media company. We use c_i to denote the degree of authority of i^{th} node. Once the i^{th} node is in state of 'Approved' and spread news to its neighbor j, the formula (1) is replaced by

$$m_j(t+1) = m_j(t) + c_i.(2)$$

- **Every edge has a weight value.** As the development of technology, we use different methods to receive news, such as newspaper, radio and internet. Hence, we can contact with more people; we have higher speed to spread news and we gradually change our source of informations. To simulate this evolution, we use a weighted edge graph. Every edge has a weight and this weight will influence how fast the belief $m_j(t)$ cumulates. Mathematically, formula (2) is replaced by

$$m_j(t+1) = m_j(t) + w_{ij} * c_i, (3)$$

where $w_{ij} \in [0,1]$. Although each edge has a weight, the meaning of weights for different types of edges are different. We illustrate this as follows.

  1. If the edge is connecting between two individuals(not including the media node), the weight represents the strength of social tie. With the help of technologies, we are able to communicate much more people compared with early stage, so it is reasonable to have more connections in our network model. However, in order to sustain the stability, we use a weight increasing with respect to time rather than increasing edges. From a probabilistic view, the expectation number of ones connections is $w_{ij}k$. If we increase w_{ij} by time, we could simulate the effect that one has more and more friends in his social network.

  2. If the edge is directing from media node to individual, the weight represents to what extent people are exposed to social media. In early time, only a few people were able to buy newspapers. As the invention of radio and TV, more and more people are able to gather information from all kinds of media, and nowadays, the majority of human

beings are exposed to media in daily life, whether they are willing or not. Thus, this weight is also a increasing parameter as time goes by.

3. if the edge is directing from individual to media, the weight represents to what extent media pay attention to this individual. In our model, media node could also gather information from individual, and, after it approves this news, the media node spread this news to other nodes. Since a media company will only believe credible people, we give more weights to famous people compared with normal people.

- **A node has an appetite of information source.** A node has different likelihood to gather informations from its neighbors depending on its neighbors' types. For example, in 1900s, people could only know news from newspaper. Now, some people are more likely to hear news from famous people's Twitter rather than from traditional media, even though they are exposed to traditional media everyday. We have used the weight of edge to depict the extent of exposure to media, now we use $[a_M, a_F, a_N]$ to depict people's willingness to get information from media, famous people and normal people respectively. Furthermore, we assume

$$a_M + a_F + a_N = 1$$

since they also represents the ratio of information people get from different source.

- **The edges from media have speed attribute.**In different periods, the speed of media is different, varying from a few days to a few hours. This will significant affect the speed of information propagation. Therefore, to observe this effect, we denote v as the speed and the delivering of message from media node to an individual needs $\frac{1}{v}$ time steps. We calculate v as the weighted average of speed of different media, discussing in later section.

- **Each information has different inherent value.**As a matter of fact, people are more willing to spread valuable news. In our model, λ depicts this willingness, since it is the probability a node transferring to 'Approved' when it first hears the news. We use $Model$ to evaluate the inherent value λ_1 of a piece of news and then let

$$\lambda = \lambda_1.(3)$$

- **People have different bias.** In different periods, different people have different affection towards different types of news. We use $Model$ to evaluate a node's bias λ_2 towards a piece of news. Then, combining with previous modification, we replace (3) by

$$\lambda = \lambda_1 \lambda_2$$

5 Results

1)All results we have

2)What criteria we should use to evaluate our model and why

3)Discussion and comparison of result and each model(Better if we could compare the result of our model and previous one, demonstrating our model is better in some sense).

4)If we use different parameter value, is there any difference? How should we choose our parameter value?

5)Can we use external data or proof(searching on Internet) to demonstrate our model is satisfactory?

6)Any more specific discussion about model(par example, is there any probability distribution similar to our result?and why? Can we use a statistical way to evaluate our result(variance etc). If we aggrandize our data size, will our result be more convincible? Is our model computational friendly and in what extent it is?)

6 Sensitivity Analysis

This is actually the concern about whether our model is statistical robust to some outliers. Those outliers could be change in data, could be change in assumptions, could be change in parameter value. We should try our best to test under different outlier conditions, what are the behaviors of our model and compare them in details.

7 Further Discussions

I think this part is more adaptable to the problem requirement, namely we could find what the problem expect us to discuss in this part. Nevertheless, except those specific topic, there are some general things we should discuss no matter what the problem is. They are,
1)Strengthness and weakness.

2)What else can we do in the future.

References

[1] Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J., and Cui, H., "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and its Applications*, Vol. 391, No. 7, 2012, pp. 2444–2453.

[2] Tanaka, M., Sakumoto, Y., Aida, M., and Kawashima, K., "Study on the growth and decline of SNSs by using the infectious recovery SIR model," *Information and Telecommunication Technologies (APSITT), 2015 10th Asia-Pacific Symposium on*, IEEE, 2015, pp. 1–3.

[3] Karafyllidis, I. and Thanailakis, A., "A model for predicting forest fire spreading using cellular automata," *Ecological Modelling*, Vol. 99, No. 1, 1997, pp. 87–97.

[4] Zhao, L., Wang, Q., Cheng, J., Chen, Y., Wang, J., and Huang, W., "Rumor spreading model with consideration of forgetting mechanism: A case of online blogging LiveJournal," *Physica A: Statistical Mechanics and its Applications*, Vol. 390, No. 13, 2011, pp. 2619–2625.

[5] Gerberry, D. J. and Milner, F. A., "An SEIQR model for childhood diseases," *Journal of mathematical biology*, Vol. 59, No. 4.

[6] Milgram, S., "The small world problem," *Psychology today*, Vol. 2, No. 1, 1967, pp. 60–67.

[7] Miritello, G., Moro, E., and Lara, R., "Dynamical strength of social ties in information spreading," *Physical Review E*, Vol. 83, No. 4, 2011, pp. 045102.

[8] Kwak, H., Lee, C., Park, H., and Moon, S., "What is Twitter, a Social Network or a News Media?" *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, ACM, New York, NY, USA, 2010, pp. 591–600.