ICM 49144

Team #49144

February 1, 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).

Understanding the process of information getting around in society network has attracted increasing attention. Information flow in complex society's network is of paramount importance when trying to solve the problem of the spreading of news, whether for analyzing historic events or for future prediction. In the light of the development of social media and social ties, and based on the previous raised models, we bring up a model here so as to represent the society's information spreading process in the network, as well as the historical trend of information network's development.

The requirement is to build a model to simulate the , with which we have to obtain the following results: build a virtual network to model the spreading of information, and validate it with data from past and now; use it for prediction in year 2050; use the controllable and deterministic factors of information, network topology, links and nodes, to influence and predict the spreading speed and the number of influenced individuals. With more taken into consideration, we obtained more freedom in the model, by far improved the simulation result of real world event, and also coincide with the historical development of social network.

1 Introduction

News is packaged information about current events happening somewhere else. [$Wiki$]It flows from individual to individual, from region to region. News could move through many different media, based on word of mouth, printing, broadcasting, and electronic communication. Thanks to the development of media and our tech-connected social network, the flow of information becomes such easy and wide-ranging as it is today. Meanwhile, the contents of news and people's bias have also changed hugely. Therefore, it is intriguing to investigate how social network structure and information value could influence the spread of information and public opinion. In this paper, we bring up a network model to simulate the information spreading process and its influence upon public interests, as well as predicting how the network will develop in the future.

We organize our paper in this following structure.

- **Introduction** includes background, restatement of problem and literature review.
- **Model** includes illustrations of two models. We first build a basic model based on previous studies, and then devise a new model incorporating our observations.
- **Result** includes simulation result for each task and explanations of some phenomena.
- **Sensitivity Analysis**
- **Further Discussion** includes what we could have improved and what may have potential meaning and worth detailed experiment.
- **Strengths and Weakness**

1.1 Restatement of Problem and Solutions

We are required to build a mathematical model to analyze the relationship between speed/flow of information vs inherent value of information. More specifically, we have five tasks; we categorize them into two sub-problems:

- Build a model to simulate the spread of information, and analyze the interaction between network structure and information flows.This includes (a),(d) and (e)
- Build a model to simulate the development of social network, and use data from the past to predict the future. This includes (b) and (c).

In this paper, we present a network model inspired by variant of SIR model [1]. We use nodes to represent individuals and edges to represent communication links between them. Given parameters, which depict structure of the network, we could simulate the spreading of news by updating network variables. Furthermore, we could change those parameters so as to mimic the development of our social network. This method solves first sub-problem satisfactory. As for the second sub-problem, we draw an analogy between the development of social network and news spreading procedure. We then derive some reasonable trending curves which accord with historical data. Moreover, our model could be used to explain diffusion of innovations and data explosion.

$liangdianliangdianliangdianliangdianliangdian$

1.2 literature review

In an very early stage, researchers believed that there are many similarities between spreading of news and spreading of epidemics. Thus, it is quite common to analyze spreading of news by an epidemic model, such as susceptible-infected-recovered (SIR) model. There are also studies based on advanced methodology, i.e. SIHR [2]. However, SIR is a analytical model; this lead to an inevitable problem that we can not take too many considerations into account. Therefore, more and more studies are focused on a cellular automata variant of SIR[1]. This new class of model would enable us to consider a much finer network structure.

In this cellular automata variant of SIR, each node has four states, namely 'Unknown', 'Known', 'Approved' and 'Exhausted'. When a node receives a message from its neighbor, it transfers to 'Known' state and remains. Every time it receives this news, it has a probability to shift to 'Approved' and spread message to its neighbors. Thereafter, this node will remain at 'Exhausted' and do nothing. Since this is a discrete-time network model, one can investigate further by manipulating network structure.

In recent years, researchers focus heavily on the effect of network structure. Liu et al. propose a dynamic network which could rewire its links during the procedure[3]. Ou et al. discuss the effect of heterogeneous connections upon news spreading[4]. Kwak et al. crawl real data from Twitter and build a network to simulate the reality[5].

However, despite the fact that there are lots of studies on different structure, few people have compared social network in different periods. In the light of previous works, we devise our new network model. On the one hand, our model takes many important factors into consideration, such as probabilistic states transition, memory effect, directed information flow, heterogeneous structure, news value and public bias, etc. On the other hand, we incorporate a few network structure parameters which would vary depending on different periods. Thus, we could simulate and predict the development of social network for decades.

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 1** and **Table 2**.

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 The confidence of i^{th} node towards the news until time t	Local
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 celebrities	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 [1], 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 celebrities or huge media company.

$Ourmodelconsistingoffourstates, describethetransferringofstatestomakeitclear!!!!ieafterKnon$

4.1.2 Assumption and Justification

- **All nodes are identical** In this model, we do not consider celebrities 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 celebrities.** In the basic model, we treat every individual equally. Nevertheless, in reality, there are celebrities 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 celebrities 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 celebrities' 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, celebrities 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 Result

6.1 Verification of the initial model

In this part we define build a virtual network to model the spreading of information, and validate it with data from past and now; use it for prediction in year 2050; use the controllable and deterministic factors of information, network topology, links and nodes, to influence and predict the spreading speed and the number of influenced individuals.

6.2 Validate the Improved Model with Data

The data we collect[] here provides a standardized reference for our model. By measuring the

$$-speeddefinedas\mathbf{v}andflowdefinedas\mathbf{n}-$$

against the actual data, we can evaluate the accuracy of the model. Here we explain the criteria we use to evaluate our model.

6.3 Future Prediction

After validate our model with the data, we can change the parameters accordingly and describe

7 Result

Generally speaking, we are utilizing a social network model to simulate the information spreading process in real world society, during which a relationship between information value and the In the following part, we define the the number of time steps required to obtain the steady result as **T**, the final number of approved individuals as **n**, and the number of approved ones divided by the steady time as ν: $\nu = \frac{n}{T}$. n and ν could be viewed as the breadth and speed of the information flow.

7.1 Requirement A

To give a quantificationally description of news' character, we can evaluate the three properties of news[6]: dissemination, timeliness and value. The value could be represented by λ_1(see datails in Model Section and the Appendix), while the dissemination could be described by n as the number of individuals approving it, and timeliness by ν as how fast the news get around.

7.2 Requirement B

In this part we give a time-based description of the parameters: historical change take place when the

7.3 Requirement C

The formula we derived from the previous data enables the prediction of social network capacity and relationship.

7.4 Requirement D

7.5 Requirement E

The above mentioned parameter λ_1,λ_1, b could correspondingly describe the information value, peoples initial opinion and bias, and strength of the information network; we can use the seed of information to resemble the source of the message; and accordingly we can vary the network structure to simulate the topology of information network.Spreading of information and influence on public opinion is measured by ν,n, and the [1] λ_1

8 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.

9 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.

[1] Δ

10 Further Discussion

In the measurement against inherent value of news, there do exist some defective part, for example the Analytic Hierarchy Process (AHP) could be used to give a quantized measure of news value. However, due to lack of space and time, we are not expounding it in details.

On the other hand, we realized that our discussion on historical change of media is not convincible enough. Without enough time and professinal insight, we found it hard to obtain the first-hand data for the media. The flawed model we used can, to some degree reflect the fact for social network, but still requires some modification on the media related parameters. The topology structure here is another point to be discussed.

The memory decay and new edge establishment is deliberately neglected here.

The numerical description of topology could possibly be acquired. Here we restrict the discussion within the categories of topology but did not give a more theoretical description.

References

[1] Lü, L., Chen, D.-B., and Zhou, T., "The small world yields the most effective information spreading," *New Journal of Physics*, Vol. 13, No. 12, 2011, pp. 123005.

[2] 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.

[3] Liu, C. and Zhang, Z.-K., "Information spreading on dynamic social networks," *Communications in Nonlinear Science and Numerical Simulation*, Vol. 19, No. 4, 2014, pp. 896–904.

[4] Ou, C., Jin, X., Wang, Y., Chen, W., and Cheng, X., "The Impact of Heterogeneous Spreading Abilities of Network Ties on Information Spreading," *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, IEEE, 2015, pp. 454–460.

[5] 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.

[6] Stephens, M., *A history of news*, Oxford University Press, USA, 2007.