Team Control Number

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

50007

Problem Chosen

D

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

Measuring the Evolution and Influence in Society's Information Networks

Summary

Basically, we build three models（BIS model, SIR model, and BA model） to explore both the flow and inherent value of the information and compare their relationships based on the consideration of 5 periods: the1870s, the 1920s, the 1970s, the 1990s, and the 2010s.

For task 1, the BIS model (based on epidemic model) is used to evaluate the flow of information and determine the filtering criteria to find out what qualifies the news.

For task 2, we examine our BIS model by calculating the information flow of the news "*Obama becoming the president of the United States*". The final result is higher than the filtering criteria during that period, which shows the validation of the BIS model. As for the capability of predicting, we run our model to predict the communication situation of 2015 and compare with the reality. And the experiment result shows that our model is not only reliable in formulating the filtering criteria line in different information periods, but also making predictions.

In this way, we further predict the information situation of the coming 2050 to solve task 3.

To solve task 4, we build the dynamic SIR model to evaluate how the public opinions change by information networks. The changing rate of susceptible nodes (whose interests would certainly be changed) exactly expresses how the opinion and the public interests be changed. Hence, we draw the complex networks by using the data from *Facebook*（*BA model*）to analyze the influential characteristics of propagation. Then use the dynamic SIR model to measure the changing of susceptible nodes, and get the problem solved.

In task 5, we detail the evaluation criteria for different aspects and also developed the *SIR* model by considering the node (individual) with emotional characteristics. In other words, the individual can choose the information according to its own interests and bias. Based on the threshold rule, we built a strategy to judge one's acceptance of a piece of information, and simulate the result of getting stable.

What's more, the strength and weakness of our models are discussed. In the end, we added our ideas about how the people's news consumption will change through information networks.

Key words: SIR model; BA model; Threshold rules; Social network;

Contents

1. Introduction

As is known to all, information is indispensable in our daily life. From the past till now, the content of the information varied and the flow of information changed as well. We're here trying to help the Institute of Communication Media (ICM) to explore the evolution and influence in society's information networks. We're assigned to analyze the relationships between the flow and the inherent value of the information based on the consideration of 5 periods: the 1870s, the 1920s, the 1970s, the 1990s, and the 2010s. To reach this, it requires us to focus on these tasks:

- Build models to explore the flow of information and filter what qualifies as news.
- Validate our model's reliability and make prediction of the current information communication situation and compare with the reality.
- Use our models to predict the communication networks' relationships and capabilities of 2050.
- Model how public interest and opinion can be changed through information networks today.
- Determine how information value, people's initial opinions, form of message, and the topology of information network worldwide could be used to spread information and influence public opinion.

1.1 Previous Models

Actually, many previous studies have given us ideas. According to the materials we searched, we find that the spreading of rumor is in many ways similar to the spreading of epidemic infection. After notified or infected, that the spreader can become the removed is similar as the infectious can recover after a few time. Thus many epidemic models have been used to describe the spread of information [1] [2] .For instance, a classical model for the spread of rumor was introduced by Daley and Kendal [3] which was the first one to propose the DK model of rumor spreading. Also, Maki and Thompson [4] focused on the analysis of the rumor spreading model, which was based on the mathematical theory. In this way, we develop our BIS model.

Also we searched for the dynamic model to find out how public opinions change [5] [6].Moreover, we developed the SIR model by considering the node (individual) with emotional characteristics.

1.2 Our Work

The development of our model was inspired by two ideas, which respectively to the two tiers, with respect to both the inherent value of the information itself and the flow of information through various media.

In order to analyze the relationships between the flow and the inherent value of the information by taking a historical perspective：

First, we build the model to explore the flow of information and filter what qualifies the news in order to understand the essence of news. Then we validate our model and make predictions of the current information communication situation and compare with the reality as well as make predictions for the coming 2050.Then we need to evaluate how the public opinions change by information networks. And finally we need to determine how information value, people's initial opinion, media and the information network strength could be used to influence public opinion and spread information.

2. Symbols and Definitions

Symbol	Definition
t	Denotes the time of spreading the news.
$x\ (t)$	Denotes the ratio of people who have already known the information. (X_0 describes the initial ratio)
r	Denotes the not-spreading rate
i	Denotes the propagating media (ranges from 1 to 5 which stands for newspaper, radio, TV, Internet and other digital news as the media respectively).
j	Denotes the period of time (ranges from 1 to 5 which stands for the 5 periods of time (1870s, 1920s, 1970s, 1990s, 2010s.)).
f_{ij}	Denotes the usage of media i to get the information during the period j.
h_{ij}	Denotes the speed of information propagating during the period j using the media i.
h_j	Denotes the information propagating rate during the period j.
h	Denotes the information propagating rate.

Table 1: Symbol description

3. Basic Assumptions and Justifications

- Rumor is a kind of social phenomenon that a remark spreads on a large scale in a short time through chains of communication and runs through the whole evolutionary history of mankind [7]. Hence, we assume that the rumor is a kind of information.
- The information can be transmitted by interpersonal connections.
- The individuals in the social connection networks can be defined as nodes in our SIR model.
- The information is independent from each other in our BA network model.
- During the same small time intervals, the amount of information of the same kind of node obeys the same distribution.

4. Our Models

4.1 Model for the Breadth of the Information Spreading (BIS)

In order to find out what qualifies the news. The first thing we need to concern is to explore the breadth of the information spreading.

Assume that every single who accepted the piece of information decided to spread the information, then we get the differential equations:

$$\begin{cases} \dfrac{dx}{dt}N = h \times N(1-x) \\ x(0) = x_0 \end{cases} \tag{1}$$

where $\frac{dx(t)}{dt}$ describes the change rate of the amount of people who know the information. Then we get the function (2):

$$x(t) = (x_0 - 1)e^{-ht} + 1 \tag{2}$$

$x(t)$ is close to 1 when t is becoming infinite, and this means the amount of people who know the information will grow exponentially. As a result, every individual will know this information. However, considering some of the information-receivers don't spread the information, hence, we optimize our model by considering r as the ratio of not-spreading rate:

$$\begin{cases} \dfrac{dx}{dt}N = hN(1-x) - Nrx = N(h-(h+r)x) \\ x(0) = x_0 \end{cases} \tag{3}$$

Then we get another function (4):

$$x(t) = \left(x_0 - \frac{h}{h+r}\right)e^{-(h+r)t} + \frac{h}{h+r} \tag{4}$$

This time, $x\ (t)$ is close to $\frac{h}{h+r}$ when t is becoming infinite, this means the information will get desalted as time goes by.

In this way, in a certain area, we can calculate the value of $\frac{h}{h+r}$ to preliminary judge whether the information is a piece of news.

4.2 the Dynamic Information Propagating Model (SIR)

Assumptions:

$s,\ i,\ r$ stands for the state of susceptible, infected and removed respectively. And $p_s,\ p_i,\ p_r$ stands for the possibility of the node be in the state of susceptible, infected and removed in the moment t. Then, $p_{jj'}$ stands for the possibility of the node j be in the state of mode j and be in the mode j' in the moment of $t+\Delta t$. For example, p_{ss} stands for the probability that node j is susceptible both in the moment t and the moment $t+\Delta t$.

We assume that the probability of a susceptible node being infected by an infective node during the period is p_1, and the probability of a susceptible node becoming a removed node is p_2. Also, we define $m=m\ (t)$ as the number of susceptible nodes among the neighbor nodes of node j.

Methology:

To get the p_{ss} of the node j:

$$p_{ss} = (1 - p_1 \Delta t)^m \tag{5}$$

If the node j has k edges, then we can get the distribution of m by considering m obeys the *binomial distribution*. Then we get the equation as follow:

$$P(X = k) = \binom{k}{m} \omega^m (1-\omega)^{k-m} \tag{6}$$

where X indicates the random variable stands for the number of susceptible nodes of j and ω stands for the probability of that node j has k infected edges at moment t.

Then, we can calculate the average transition probability of p_{ss}:

$$\overline{p_{ss}(k,t)} = \sum_{m=0}^{k} \binom{k}{m} (1 - p_1 \Delta t)^m \omega(k,t)^m (1 - (\omega(k,t))^{k-m} \tag{7}$$

As we know that the susceptible node would not spread all the time, and it will spread at a speed v and stop as the result of becoming a removed node. Then we can calculate p_{ii} and the average transition probability $\overline{p_{ii}(k,t)}$:

$$p_{ii} = (1 - p_2 \Delta t)^m (1 - v \Delta t) \tag{8}$$

$$\overline{p_{ii}(k,t)} = \sum_{m=0}^{k} \binom{k}{m} (1 - p_2 \Delta t)^m \omega'(k,t)^m (1 - (\omega'(k,t))^{k-m} (1 - v \Delta t) \tag{9}$$

As for the calculation of the total number of nodes $N(k, t)$, it is obvious that:

$$N(k, t) = I(k, t) + S(k, t) + R(k, t) \tag{10}$$

where $I\ (k,\ t)$, $S\ (k,\ t)$, $R\ (k,\ t)$ stands for the number of the infected, susceptible and removed nodes who share the same degree k.

Then the flow of information is depended on the change in I, S and R during the time [$t,\ t+\Delta t$]:

$$\begin{cases} S(k,t+\Delta t)=S(k,t)-S(k,t)(1-\overline{p_{ss}(k,t)}) \\ I(k,t+\Delta t)=I(k,t)+S(k,t)(1-\overline{p_{ss}(k,t)}) \\ R(k,t+\Delta t)=R(k,t)+I(k,t)(1-\overline{p_{ii}(k,t)}) \end{cases} \tag{11}$$

Moreover, the calculation for $\omega(k,t)$ is:

$$\omega(k,t)=\sum_{k'}p(k'k)\rho^{s}(k',t) \tag{12}$$

where $p\ (k'k)$ stands for the possibility of a node with the degree k who is linked with another node with the degree of k'.

Then the change the equation (11) into:

$$\frac{\rho^{s}(k,t+\Delta t)-\rho^{s}(k,t)}{\Delta t}=-\frac{\rho^{s}(k,t)}{\Delta t}\left(1-(1-p_1\Delta t\sum_{k'}\rho^{i}(k',t)p(k'k))^{k}\right) \tag{13}$$

Finally, when Δt becomes indefinitely small, we can get the following 3 differential equations:

$$\begin{cases} \dfrac{\partial\rho^{s}(k,t)}{\partial t}=-kp_1\rho^{s}(k,t)\sum_{k'}\rho^{i}(k',t)p(k'k) \\ \dfrac{\partial\rho^{i}(k,t)}{\partial t}=-kp_2\rho^{i}(k,t)\sum_{k'}\rho^{r}(k',t)p(k'k)-v\rho^{i}(k,t)+kp_1\rho^{s}(k,t)\sum_{k'}\rho^{i}(k',t)p(k'k) \\ \dfrac{\partial\rho^{r}(k,t)}{\partial t}=kp_2\rho^{i}(k,t)\sum_{k'}\rho^{r}(k',t)p(k'k)+v\rho^{i}(k,t) \end{cases} \tag{14}$$

4.3 BA Model [Barabási et al.1999] (BA)

Characteristics:

- **Increase:** start from that the amount of nodes is a small number m0, and at any time interval, add a new node with degree of m $(m_0 \geq m)$.
- **Preferential attachment:** when choosing the new node to attach, assume that the probability of the now node getting attached with the old node i depends on the connectivity of i,

$$\pi(k_i) = \frac{k_i}{\sum_k k_j} \tag{15}$$

after t time intervals, the model will produce a random network of $N=m_0+t$ nodes and m_t edges, as the model's evolution become an immutable state, its distribute obey power-law distribution:

$$p(k) \sim k^{-\gamma} \tag{16}$$

where γ is a positive number and not relative to m , which means the system has become a situation of statics.

Static Attribution:

- **Degree**: all the edges a node n has is k_n
- **The distance between n_i and n_j [$d(n_i, n_j)$]**:The average shortest path of the network is:

$$L = \frac{\sum_{i,j=1}^{N} d(n_i, n_j)}{\binom{N}{2}} \tag{17}$$

- **Clustering coefficient:**

For the node n, if it can be linked to $(N\text{-}1)$ nodes, then the maximum number of edges is:

$$m(n) = \binom{N-1}{2} \tag{18}$$

The clustering coefficient of node n is：

$$C(n) = \frac{a(n)}{m(n)} \tag{19}$$

where $a(n)$ is the actual edges of n.

Finally, we can get:

$$C = \sum_n C_n \tag{20}$$

5. Solutions

5.1 Task 1: Explore the flow of Information and filter the news

We use the BIS model for task 1 which requires us to calculate different h_j during the period of 1870s, 1920s, 1970s, 1990s, and 2010s. First we get our information of f_{ij} from Figure 1 and assign the value of h_{ij} according to our researches. Finally we put the value of f_{ij} and h_{ij} into the formula (21) to determine h_j as follow:

Initialization:

Assign appropriate initial value to h_1.

Calculation:

$$h_j = \sum_i f_i h_{ij} \tag{21}$$

Use the formula (21) to calculate each h_j (when one h_{ij} is very small comparing with other h_{ij}, it can be ignored). And the calculation result of h_j value for 5 periods are shown in Table 2.

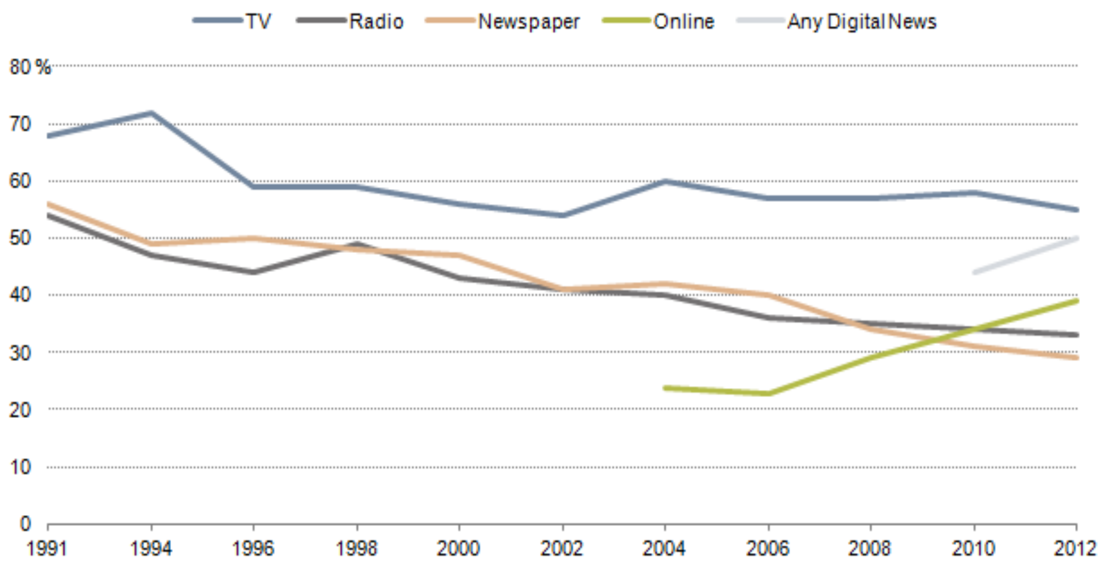

Figure 1: the Usage of Different Media

h_1	h_2	h_3	h_4	h_5
0.01	0.017	0.03	0.057	0.139

Table 2: h value for different periods

As for the not-spreading rate r, in order to filter what qualifies the news, we assigned the critical value r of becoming the news as a constant as 0.4. In other words, when people receive a piece of news, the tendency for not spreading the information is 0.4, which apparently conforms to our common sense.

Putting the value of h_j and r into the formula (4), we finally get the $x\ (t)$ as ratio of people who have already known the information in different periods (Figure 2):

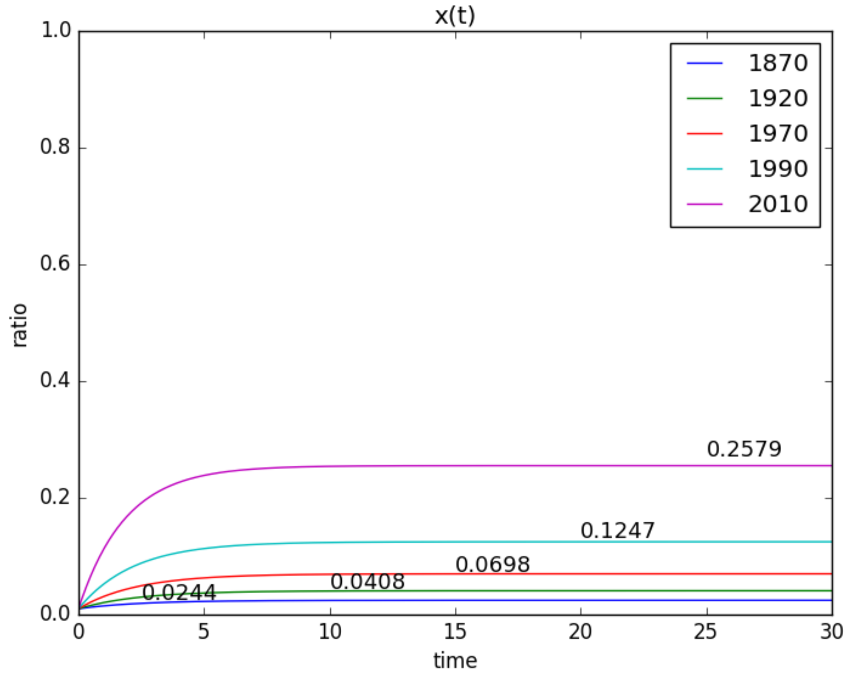

Figure 2: X (t) for different periods

Obviously, we can get the different filtering criteria to determine whether the information qualifies the news. The results are shown in Figure 3,which means as long as the spreading of a piece of news qualifies the filtering criteria line in different periods, we can determine that this information is a piece of news.

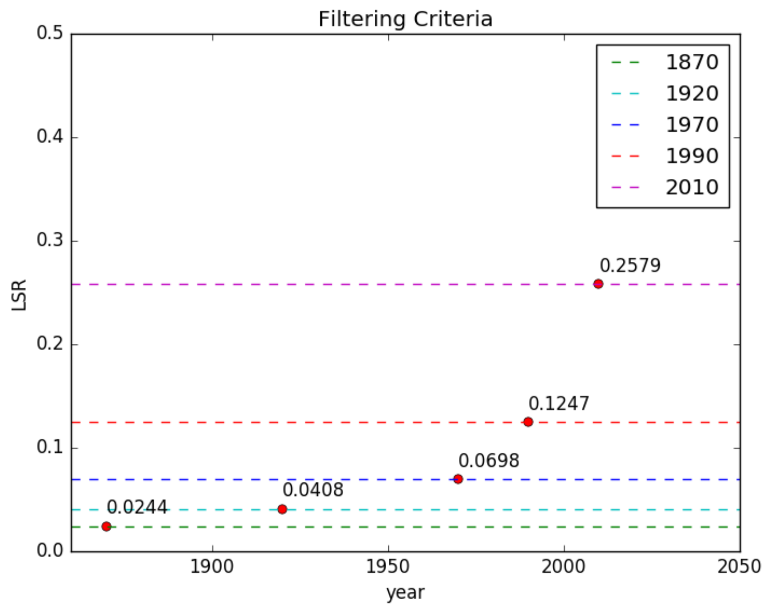

Figure 3: the Filtering Criteria for Different Periods

5.2 Task 2: Validation & Prediction

- **Validate the reliability**

We consider a piece of news of *Obama becoming the president* happened in 2008 to validate our model's reliability. The h value for this information is high, so we assume the spreading speed of TV and Internet becoming twice of the average speed. Then we get value of h', which is 0.278., and the r value of this model is 0.2

In this way, we can get the final ratio is higher than the filtering criteria:

$$\frac{h}{h+r}=\frac{0.278}{0.278+0.2}=0.58>0.2579$$

And the result apparently shows that this piece of information can be defined as news which means our model is workable.

- **Predict the Current Situation and compare with the reality**

As we use our model to predict the current situation, we find that the change of value h_j basically followed the exponential function (22):

$$h=2\times10^{-17}e^{0.0179t} \tag{22}$$

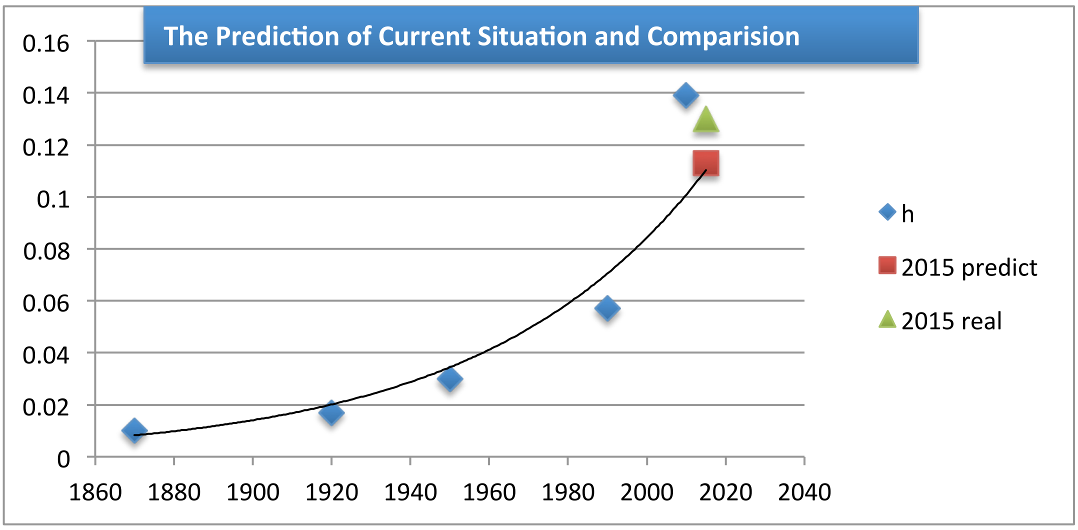

Figure 4: the h value of the prediction and the reality

Figure 4 illustrates the variation of h_j value through different time periods.

The red point shows the prediction of h_j values from the past periods, while the green one shows the reality.

Then fit the formula (22) to the data in Figure 4, we can get the curve in Figure 5 as follow, and obviously, the deviation is small which proves our model's validity.

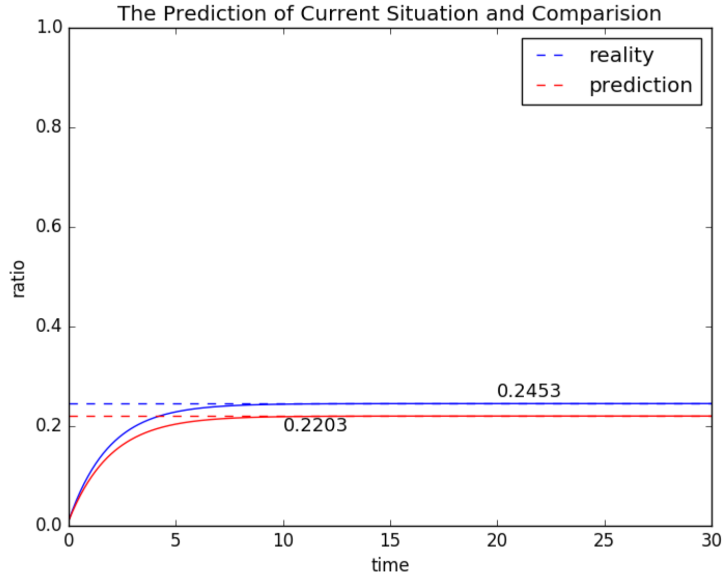

Figure 5: X (t) for the prediction and the reality

5.3 Task 3: Predict the Situation of 2050

We use the same model and method in task 2 to predict the communication networks relationships and capabilities of 2050 and the value of h is shown in Figure 6:

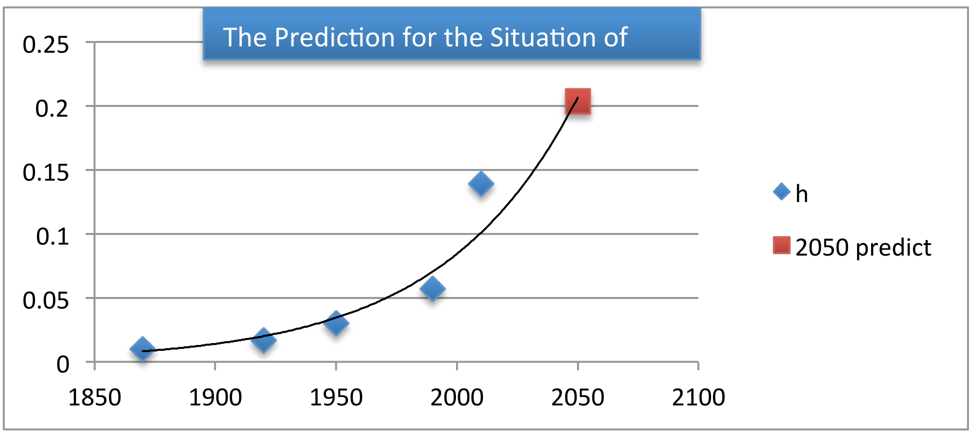

Figure 6: the Prediction for the Situation of 2050

Then fit the formula (22) to the data in Figure 6, we can get the curve in Figure 7, which clearly shows the spreading ratio of information in 2050.

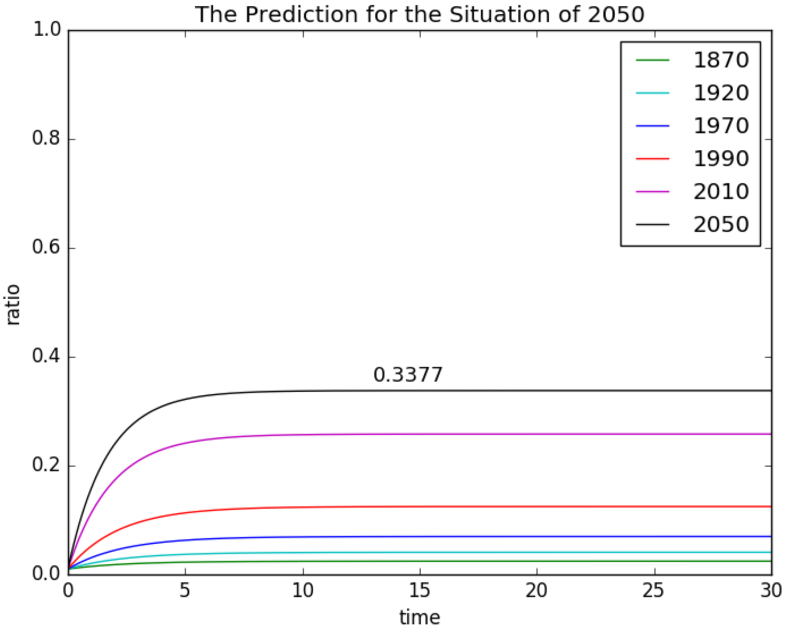

Figure 7: the Prediction of the Situation of 2050

And finally we extract the limit spreading ratio as the news filtering criteria in 2050 (Figure 8). Apparently, as the ratio of spreading the news increases due to people's usage of different media, the filtering criteria line to justify whether a piece of information is news increases as well.

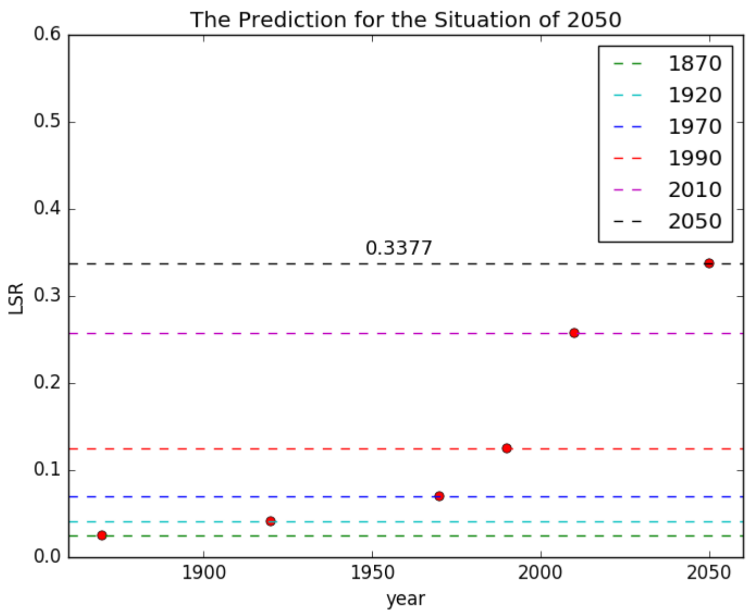

Figure 8: the Criteria Filtering Line for 2050

5.4 Task 4: Explain how public opinions change

Model 2 (SIR) is used for solving the problem of task 4. Also, we build the model 3 (BA network model) to evaluate the recent information influence.

We consider the public interest and opinion's changing as the result of information's spreading. The one who get and spread the information is influenced by the spreading opinion. And we can get our data base from the social network *Facebook.*

Step 1. Fit the data into the equations in our dynamic SIR model. Using *Python* programing software, we can depicts the information networks. And the result is shown in Figure 9:

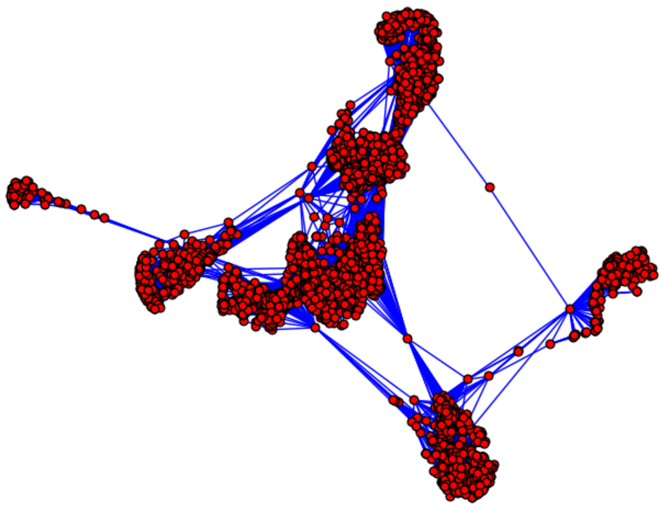

Figure 9: the Information network of SIR model

The total nodes of the database is *4039*, and the total edges are *88234*. The average degree is *78.92*, which is high. The highest degree is *538*, and the average clustering coefficient is *0.2056*, which illustrates that the relationships between the nodes are close.

Step 2. We can figure out the tendency of the change in ρ^i, ρ^s and ρ^r as time changes. To measure how the public's interest and opinion changes, we decide to take the change of susceptible nodes into consideration. And the result is shown in Figure 10:

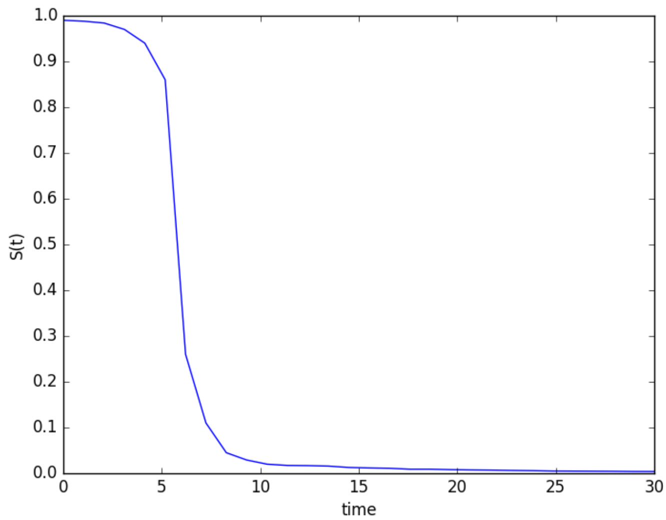

Figure 10: the Change of Susceptible Nodes

Figure 10 indicates that when a new opinion comes out, the dropping trend of the population who won't change their opinions or interests is getting less and less rapidly. In the very beginning of propagating the information, the susceptible nodes declining rate is extraordinary high *(around time 5).*

5.5 Task 5: the factors determine the spread information and influence public opinion

In task 4, we used the SIR model which classified all the nodes into three different categories (the susceptible, infected, and removed condition). However, for the requirements in task 5 to determine how information value, people's initial opinions, propagating media, and the topology of the information network be used in propagating information and influencing public opinions.

For this task, we optimized our SIR model by considering the changing process from susceptible condition to infected condition. In other words, one individual may only get the information as a *"receiver"* not an *"acceptor"* by its own judgment.

- **The Information Value**

We assume that the information value is determined by the benefits that the information bring to an individual. And the benefits can be divided into direct benefits and indirect benefit.

- **People's initial opinion and bias**
  - **People's interest index**

We use o_i to evaluate the initial opinions and bias degree towards an opinion O of the person i, and the individual has a threshold of accepting or adopting an opinion of a_i, if $o_i > a_i$, then the individual will accept the opinion O, so the probability that he will accept to it is "*1*". Otherwise, it will be "*0*".

When $o_i < a_i$, the first time i hear about the opinion, he will certainly not accept it, by the second time, he still would not…at the Nth times he heard it, he may take into consideration if that opinion is good. That means the relationship between o_i and a_i is not as clear as before.

- **People's Probability of Accepting an Opinion**

In order to evaluate an individual's probability of accepting an opinion, we use the *Threshold Rule*. We measure people's acceptance of an opinion is binominal (*"0"* for *"reject"*, *"1"*for *"accept"*).

From these two aspects, we can better understand people's initial opinion and bias.

- **the Topology or Strength of the Information Network**

We can use the BA model to validate the topology of the information network by formulating the current network situation and calculate the whole connection of each node to determine the strength of the information network.

- **Form of the message or its source**

We think that the media of propagating the message weighs a lot from the analysis results we got from task 1.In this way, we can calculate the information spreading through different media to determine different media form's part of influencing the public opinion.

Moreover, usually, the value of o_i differs obviously. As time passes by, the information would pass every individual many times, and the distribution of value o_i will turn from relatively distributed to relatively concentrated, and finally reaches a steady condition.

This indicates the public opinions may be leaded by the spreading of information, and come to a steady condition, however, this doesn't mean that every single's opinions would be the same.

6. Strength and Weakness

6.1 Strength

- Reasonable：the deviation between the prediction of our model and the reality of the current situation is small, which proves the rationality and correctness of our model
- In the BIS model, the criteria filtering line is numerical and easy to calculate.

6.2 Weakness

- In the BIS model, we simply define the value of x_0 in different situations as a constant, which may influence our models accuracy.
- We just figure out what influences public opinions but still needs to be validated quantitatively.

7. Our Ideas about Information Propagating

From the analysis we did, it's quite obvious that the social networks changed in the past times thanks to the innovation of technology from spreading the news orally to this smart phone era. And the growth of social networks change people's way of getting information and make people closer to each other.

Due to its fast propagating rate, social networks has become the major media that people get information and people tend to use the social networks to explore the world.

8. References

[1] L. J. Zhao, J.J. Wang, Y. C. Chen et al., "SIHR rumor spreading model in social networks," Physica A, vol. 391,no.7, pp. 2444-2453,2012.

[2] K. Kawachi, "Deterministic models for rumor transmission," Nonlinear Analysis: Real World Applications, vol.9, no.5, pp.1989-2028, 2008.

[3] D. J. Daley and D.G, Kendall, "Epidemics and rumors," Nature, vol.204, article 1118, 1964.

[4] D. P. Maki and M. Thompson, Mathematical Models and Applications, Prentice-Hall, Englewood Cliffs, NJ，USA，1973.

[5] DUNCAN J.WATTS, "Influential, Networks, and Public Opinion Formation."

[6] Zhang YanChao,Liu Yun,Zhang HaiFeng,etc.(Key Laboratory of Communi cation and Information Systems,Beijing Jiaotong University,Beijing Munici pal Commission of Education,Beijing 100044,China)
[7] K. Kawachi, M. Seki, H. Yoshida, Y.Otake, K. Karashina, and H.Ueda, “A rumor transmission model with various contact interactions,” Journal of Theoretical Biology, vol. 253, no.1, pp.55-60, 2008.
[8] http://snap.stanford.edu/infopath/