

For office use only  
T1 \_\_\_\_\_  
T2 \_\_\_\_\_  
T3 \_\_\_\_\_  
T4 \_\_\_\_\_

Team Control Number

**48039**

Problem Chosen

**D**

For office use only  
F1 \_\_\_\_\_  
F2 \_\_\_\_\_  
F3 \_\_\_\_\_  
F4 \_\_\_\_\_

### **2018 Mathematical Contest in Modeling (MCM/ICM) Summary Sheet**

#### **From Gossip to Social Networks — Measuring the Evolution and Influence in Society's Information Networks Summary**

For the sake of *Measuring the Evolution and Influence in Society's Information Networks*, our idea is to set up a comprehensive model with plentiful properties of nodes which are consistent with the current states. Therefore, the concept of *Complex Network* is introduced to depict information dissemination vividly. What's more, the *Twitter's Network Dataset* is utilized to verify most of the models we build.

About Problem A, in order to *Explore the Information Flow* and *Filter out Valuable Information*, we establish the *Information Flow and Filter Model*. Having considered the diversity in information media, the *Flexible Network Structure* is explored to suit for the various types of information flow. Meanwhile, to gain a better information filter result, we create a filter consisting of 6 evaluation indexes. After theoretical preparation, we use the *Twitter's Network Dataset* to establish Internet Information Flow and Filter Model for simulation. Finally, the coverage rate of information dissemination is equal to 96.7% , after reaching the steady state of information dissemination.

As to Problem B & C, based on the Information Dissemination Model, our assignment is to use the data to verify the *Reliability of the Model* and *Predict the Future Data*. We establish the *Network Evolution Model* to imitate the dynamic network structure in a single medium. The *Average Euclidean Distance* of the *real data* and the *predicted data* is 1.784 , which stands for a good forecast result. After that, we take the model as the basis for the prediction of the Society's Information Network in 2050. Without considering the emergence of new technologies, the cell phone usage is 51.34%, which is the highest in the 4 kinds of media. However, after making bold assumptions that the *Quantum Communication* would be invented in 2035, the share of the information medium will change dramatically, and the *Quantum Communication* would be the main medium.

When solving the Problem D & E, we are required to analyze the Public Opinion Dissemination. We divide the network nodes into 3 types: *Individuals, Opinion Leaders and Internet Marketer Group*. And then we establish the *Public Opinion Dissemination Model* to simulate the real situation of the public opinion dissemination, and draw the contrast curves of public opinion dissemination coverage rate before and after *Taking "PUSH" Operation*. Finally, we screen out the effect factors of public opinion dissemination, and use the example of *President Obama's Nearly Opened Twitter* to verify the stability of the model.

The accuracy and stability of our model are proved by simulation using the *Twitter's Network Dataset*. Furthermore, the Information Dissemination Network Model can adapt to different types of media. By slightly adjusting the coefficients, we can apply this network model in various fields of information dissemination.

**Keywords:** *Information Dissemination Model; Dynamic Network Structure; Twitter's Network Dataset; Simulation & Verification*

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Problem Restatement and Analysis . . . . .	1
<b>2</b>	<b>Assumption</b>	<b>2</b>
<b>3</b>	<b>Symbol Description</b>	<b>2</b>
<b>4</b>	<b>Problem A : Information Flow and Filter Model</b>	<b>2</b>
4.1	Information Flow Model . . . . .	3
4.1.1	Theoretical Model of Information Flow . . . . .	3
4.1.2	Example of Information Flow Network Model (Internet) . . . . .	5
4.2	Information Filter Model . . . . .	6
4.3	Information Flow and Filter Model Simulation . . . . .	7
4.3.1	Simulation assumptions and data sources . . . . .	7
4.3.2	Simulation process . . . . .	7
4.3.3	Simulation results . . . . .	8
<b>5</b>	<b>Problem B : Forecast Information Network's Structure &amp; Validate its Reliability</b>	<b>9</b>
5.1	Short Term Prediction Model of Information Dissemination . . . . .	9
5.1.1	Model preparation . . . . .	9
5.1.2	Realistic network model construction . . . . .	9
5.2	Model Simulation Process . . . . .	11
5.3	Forecast & Validation Results . . . . .	12
5.3.1	Forecast Results . . . . .	12
5.3.2	Validation Results . . . . .	13
<b>6</b>	<b>Problem C : Long Term Network Evolution Model</b>	<b>13</b>
6.1	Model building . . . . .	13
6.2	The bold assumption of future communication technology . . . . .	14
6.3	Model solving . . . . .	14
<b>7</b>	<b>Problem D : Public Opinion Dissemination Model</b>	<b>16</b>
<b>8</b>	<b>Problem E : Analyze the Factors that Influence the Public Opinion Dissemination</b>	<b>17</b>
8.1	Influence Factors of Public Opinion Disseminate . . . . .	17
8.2	Analysis One Historical Event based on Public Opinion Dissemination Model . . . . .	18
<b>9</b>	<b>Strengths and Weaknesses</b>	<b>19</b>
<b>References</b>		<b>20</b>

## 1 Introduction

### 1.1 Background

In the past, because of the limited approaches and speed of transmission, information which derives the necessary data and knowledge is dominant in dissemination network. People need such information to arrange their lives and production. However, with the evolution of communication speed and method, widely widespread information becomes more multiple.

The rapid development of information network makes information available for meeting mankind's various needs, among which there are not only indispensable daily needs like transportation and education needs, but also information, such as entertainment news and humorous videos that serve for amusement and delight. For all these information spread, they can be classified into two types: information that people receive actively or passively. The apparently inherent value of these information differs. Though information network itself won't decide which to transmit, users' choices and the progress of the information network work together to have an effect on information's inherent value.

Consequently, it is necessary to find out the relationship between flow of information and inherent value of information. Thus we can learn how to make the transmission of certain information become more large-scale and faster. Meanwhile, realizing this relationship, makes it becomes easier to improve the quality of information. It also helps us make today's superfluous information flow practical and efficient and, more importantly, enables us to understand the maturation of humans' culture of information and communication.

### 1.2 Problem Restatement and Analysis

For Problem A, in order to explore the flow of information, we are required to establish an information dissemination network to probe into specific process. Meanwhile, we have to screen out the content which can become news. Hence, an evaluation model needs to be created and used to carry out the simulation test.

In regard with Problem B, the forecast about nowadays' information dissemination situation needs to be built up and verified with real data. This requires us to establish a prediction model, together with an evaluation index system which can describe the attributes of network varies in different media, which share similar dissemination structure. Thus we can take the evolution of information network structure into consideration to conclude our results.

It's closely followed by Problem C that asks for the forecast about the communication networks' relationships and attributes around the year 2050. Considering the lengthy time interval between now and 2050, we can analyze the changes of network's media at first, and then evaluate the efficiency of network in different time.

Problem D requires us to explore how public interest and opinion can be change by applying the theories and concepts of information influence on networks. We take into account that the public opinion can not be quantified, we need to look for evaluation indicators as a measure.

As for Problem E, the elements mentioned in the problem will be used to analyze how they exert an influence upon the flow of information and public opinion. Meanwhile, it should be known that those elements influence the information network directly, and then affect the information spread in it. Thus, we adjust the parameter of network to simulate those elements' influence. Finally, we can validate it with analyzing an instance.

## 2 Assumption

- (1) Information will be spread stably except for special situation.
- (2) Information network and evolution network are universal and suitable for each kind of media.
- (3) The information network is affected only by the factors that have been recognized.
- (4) In the process of information dissemination, we assume that the network edge is directed.
- (5) It is assumed that only the neighboring nodes will be affected.
- (6) Information can only be transmitted from the source to the spread point, and it will not produce cross dissemination.

## 3 Symbol Description

In the section, we use some symbols for constructing the model as follows.

Symbol	Description
$T_n$	The discrete time series of the node checking information.
$p_{in}$	The probability of Node $i$ to check information at the moment of $T_n$ .
$q_{in}$	The probability that information is able to continue to spread
$\lambda_i$	The probability of node's retransmission or comments after receiving the message.
$N_n$	The number of nodes that have received information at the moment $T_n$
$Q_n$	The number of nodes that have received information at the moment $T_n$
$\gamma_i$	The new attention degree of a certain node in the unit time $\Delta t$
$D_i$	The degree of Node $i$
$C_i$	The clustering of retransmission Node $i$
$L_{ij}$	The shortest path length of Node $i$ and Node $j$

P.S: Other symbol instructions will be given in the paper.

## 4 Problem A : Information Flow and Filter Model

After analyzing the problem, we found two key points. One is *describing the flow of information*, the other is *finding a way to filter out the information that can be used as news*.

The concept of *information flow* in an information theoretical context is the transfer of information from a variable  $x$  to a variable  $y$  in a given process<sup>[3]</sup>. When we chat face to face, when we use Facebook or Twitter, when we write letters to friends in the distance, we are all in the transmission of information. There are many ways to spread information. Under these circumstances, the change of the information flow is very complicated. So we build a well time-applicability model including various transmission routes to describe the flow of information. And the part of the Information Flow Model graph is drawn in Figure 1 as follows.

About the other key point, the Information filtering system is a system that removes redundant or unwanted information from an information stream using (semi)automated or computerized methods prior to presentation to a human user<sup>[5]</sup>. Any online or offline information is

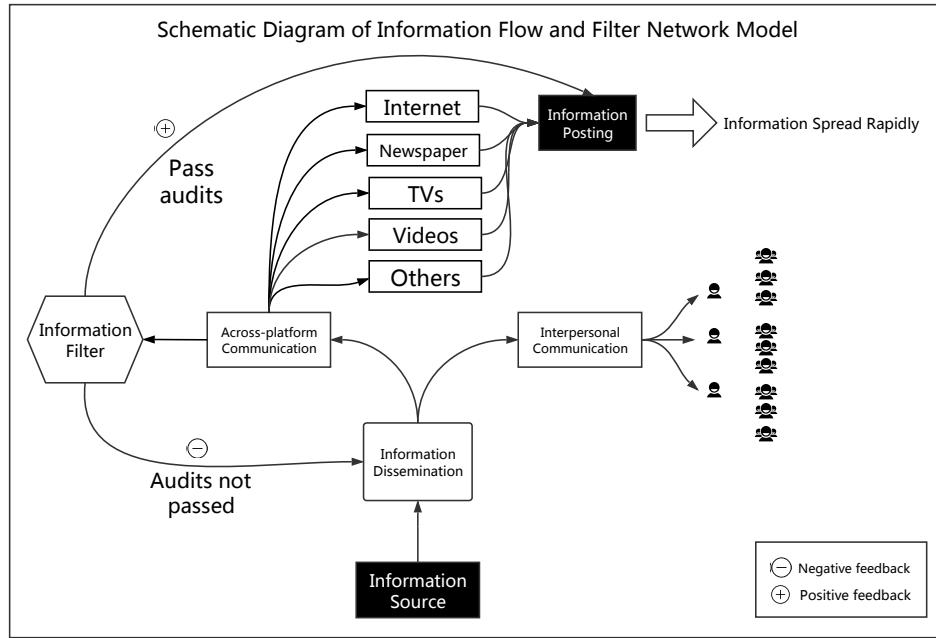


Figure 1: The schematic diagram of information flow and filter model. The information source spreads the information. The information dissemination has two ways to spread it. One is the *interpersonal communication*, the other is the *across-platform communication*. The former method is oral communication, and not restricted by the information filters. The latter includes internet, newspaper, etc.. They are constrained by the filter, which generates different feedback for different information.

part of larger and more complex ecologies, in which the false information and the truth are mixed together. In order to dig out useful information, we need to create a review mechanism. Part of the Information Filter Model graph is drawn in Figure 1 above. In the Figure 1, we stipulate that the message filter has two kinds of effects caused by the review of information: when the message is not audited, it will inhibit communication of information sources and slowing the spreading speed; when a message passes the audit, the spreading speed will be enhanced.

## 4.1 Information Flow Model

### 4.1.1 Theoretical Model of Information Flow

#### Basic theory of society's information network

The definition of the Information Flow Network is

$$G = (V, E) \quad \forall v, u \in V, u \leftarrow v, \forall e = (u, v) \in E \quad (1)$$

where  $v$  and  $u$  are the nodes of  $G$ , and  $e$  stands for the edge. Meanwhile, the  $e$  can also represent that  $u$  can receive information from  $v$ . There is information interaction between  $u$  and  $v$ . We also define  $|V|$  as the number of nodes and  $|E|$  as the total number of edges.

## Dynamic transmission mechanism

Information source is usually a point or several points from the beginning and gradually spreads to the entire network. Assume that the information is propagated from the information source point  $A$ ,  $A_j$  is the  $j_{th}$  node that receive information from  $A_i$  ( $j = i + 1, i \geq 0$ ) and the propagation velocity  $V = 1$ . The number of  $A$ 's neighboring nodes  $N(t)$ .  $A_j$  obtaining the information is in the sphere  $D$  with  $A_i$  as the origin and time  $t$  as the radius. After  $A_1$  gains the information, namely, the sub information source, they will become independent source of information, and spread the message to the other nodes independently.

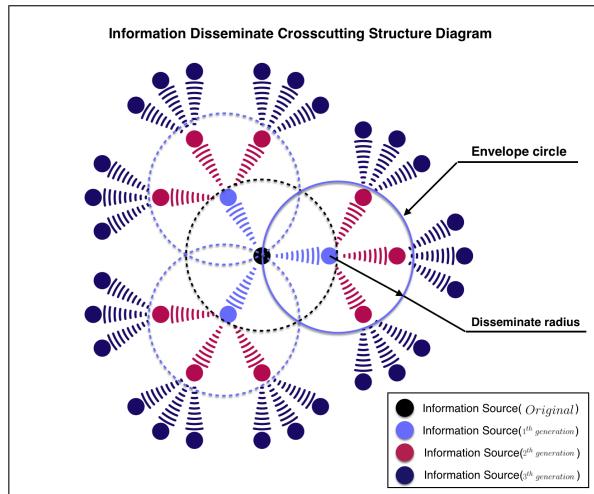


Figure 2: The information disseminate crosscutting structure diagram. Information dissemination is 3D , the nodes spread at the next time may become an independent source of information.

## Transformation of the states of nodes

In order to study the state changes of the network nodes quantitatively, and analyze the transformation of the network structure with time. We divide the nodes into 4 classes:

- (1) The nodes in the  $S$  state are the nodes which haven't received information yet.
- (2) The nodes in the  $I$  state are the nodes which have received information and can continue to spread it.
- (3) The nodes in the  $IR$  state are the nodes which have received information and do not spread information at the current time.
- (4) The nodes in the  $R$  state are the nodes which have received information and can't continue to spread it.

*Assume that nodes can spread information only one time, that is, after spreading a message, the  $I$  state of the nodes turn into the  $R$  state.*

## The attributes of each part of the network

- Network structure attribute: the characteristic attributes of different networks are embodied in the different number of nodes and different number of edges. Node degree is different, resulting in the transmission efficiency of the node is different, the speed of the network is different.
- Node attribute: the degree of nodes is different, resulting in the probability of the node to receive the message is different, the probability of the spread of the message is also different.
- Attribute of the information itself: the attribute of information itself is mainly reflected in the importance of the information and the correlation degree of the information.

#### 4.1.2 Example of Information Flow Network Model (Internet)

In order to better explain the theoretical model we have constructed, we build an example model, which takes the Internet as a single media of dissemination.

##### Construction of realistic network model

After analyzing the theoretical model, we will set up the concrete social's information flow model in Internet.

We establish a social network, which is a *directed* graph

$$G_I = (V_I, E_I) \quad \forall v_I, u_I \in V_I, \forall e_I = (u_I, v_I) \in E_I \quad (2)$$

where  $v_I$  and  $u_I$  are the nodes of  $G_I$ , representing the *people or group in society network*. And  $e_I$  stands for the edge, representing the *transmission route between people or groups*. Meanwhile, the  $e_I$  can also represent that  $u_I$  can receive information from  $v_I$ , the same as  $u_I \leftarrow v_I$ . We also define  $|V_I|$  as the number of people or groups and  $|E_I|$  as the total number of transmission route.

The discrete time series of the node checking information is  $T_n, n = 1, 2, 3, \dots$ , the probability of *Node i* to check information at the moment of  $T_n$  is  $p_{in}, i = 1, 2, 3, \dots, |V_I|$ ;  $p_{in} \in [0, 1]$  is an independent random variable.

Assuming that at the moment of  $T_n$ , the node has received a message, then the probability that information is able to continue to spread is  $q_{in}$

$$q_{in} = p_{in} \times [(1 - (1 - \lambda_i)^m)/e^{\alpha+\beta}] \quad (3)$$

where

- $\lambda_i$  is influenced by the information attribute factors and the subjective factors of nodes, standing for the probability of node's retransmission or comments after receiving the message.
- $m$  represents the *Node i* has accepted the cumulative number of information at the moment of  $T_n$ .
- $\alpha$  is the cumulative credibility of information sources.  $\beta$  is the cumulative number of strong relationship of information sources.

In the model we have established, the node is human. Under this circumstance, if we need to study node attribute, we have to take into account different types of people. The biggest difference among the different groups is the age structure. Different nodes have different age structures, so there are different node attributes. The influence has been drawn in Figure ?? as follows.

We find that not only the attributes of the nodes can affect the propagation of information, but also the properties of the network structure will play a very important role. Different times have different modes of transmission, in the 1870s, newspapers were delivered by trains and stories were passed by telegraph; in the 1970s, when televisions were in most homes; in the 1990s, when households began connecting to the early internet; in the 2010s, when we can carry a connection to the world on our phones. The main difference between different modes of transmission is the spread speed of the media, which can affect the structure of the network. Therefore, we have to take the transmission modes into account.

In this example, the information source is people or groups, spreading the information. The information dissemination has two ways to spread it. One is the *interpersonal communication*, the other is the *across-platform communication*. The former method is oral communication. The latter includes different ways.

## 4.2 Information Filter Model

The presence of incorrect and misleading content on the Web can have serious consequences for people who increasingly rely on the internet as their information source for topics such as health, politics, and financial advice<sup>[6]</sup>. Under these circumstances, we have to establish a model of information filter to screen for real, valuable messages.

In order to solve this problem, we set up an information value evaluation model. The schematic diagram of evaluation index classification is Figure 3 as follows.

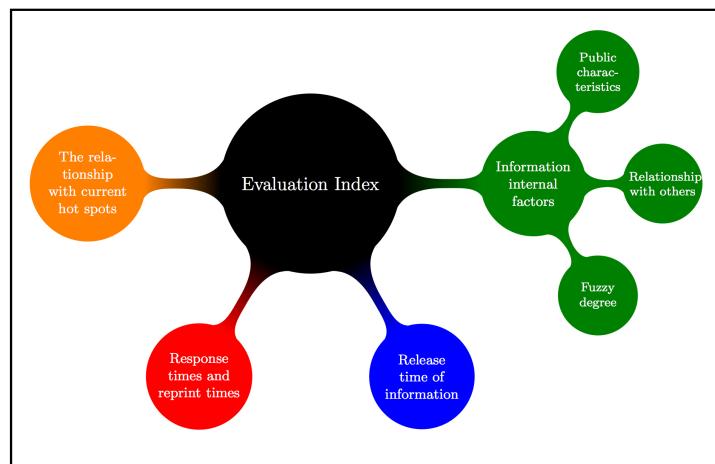


Figure 3: The schematic diagram of evaluation index classification. The center node has 4 screening indexes, each index has different subindexes.

After selecting good evaluation indexes, we will quantify the various evaluation indexes to measure the information and screen out the valuable information. The screening index calculation formula is as follows.

$$\omega = e^{-a(t-t_s)} * b_i * b_j * b_k * b_h \quad (4)$$

where the  $b_i, b_j, b_k$  are the independent random variable,  $b_i$  is the relationship degree with

current hot spots, the  $b_j$  is the degree of public concern about the information, and  $b_k$  is the response degree. What's more,  $b_h$  represents the information internal factor index.  $t$  is the current time,  $t_s$  is the release time.  $a$  stands for the time factor.

The formula above has 4 independent random variables. Because of the absence of specific screening data, we can not make a further analysis of the model. We can only give the preliminary model above.

### 4.3 Information Flow and Filter Model Simulation

In the above, we have set up a good information flow and filter model. To verify the reliability of the model, we simulate it.

#### 4.3.1 Simulation assumptions and data sources

Assumptions are as follows.

- Assuming that the same information is unilaterally circulated, the network model we built is directed graph.
- Our simulation is based on the Twitter's social network, and the data we analyze is the Twitter user's follows data<sup>[4]</sup>.
- Taking into account the convenience of access to the Internet and the rapid transmission of information, we assume that once the users have received a message, they have already obtained the information successfully.
- Taking into account the *freedom of speech* and *information security* in the United States, in this simulation, we do not consider the feedback role of information filter in information dissemination.
- We define  $\lambda_i = 1$ .

#### 4.3.2 Simulation process

The process of information flow and filter model simulation as follows.

- (1) Basic preparation and data processing.
- (2) According to the discrete time series  $T_n$ , we evolve the process of information dissemination.
- (3) Determine the  $N_n$  as the number of nodes that have received information at the moment  $T_n$ , and the  $Q_n$  as the number of nodes that have received information at this time.
- (4) We use Twitter 2012's social network data. The data consist of 81306 nodes and 1768149 edges. Through the selection of different information sources, we begin to simulate the process of information dissemination, and use the  $Q_n$  to measure the range of information dissemination at different times.
- (5) Summary and graphical data results.

### 4.3.3 Simulation results

After taking simulation, we can obtain some results as follows.

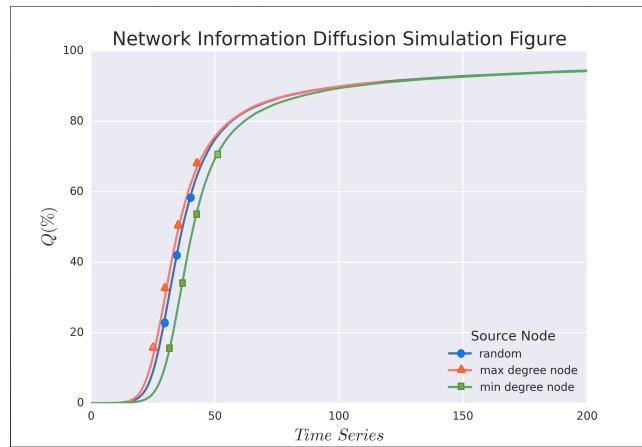


Figure 4: The network information diffusion simulation figure.  $Q$  is the diffusion ratio. Three curves are included in the graph, respectively, taking the maximum degree nodes, the minimum degree nodes and a node arbitrarily as the information source.

From the Figure 4, maximum degree node's spread is faster than the other two types of nodes at the beginning of the information spread. Minimum node's diffusion rate is the slowest. Meanwhile, the coverage ratio of the three sources is increasing exponentially in the medium term (about 20 times). After the evolution of about 100 times, the coverage ratio of the three sources tends to grow slowly, eventually approaching 100%.

After analyzing the dissemination characteristics of different information sources, we select the maximum degree node as the information source, analyze it, and make the data we got visualization. The results are in Figure 5 as follows.

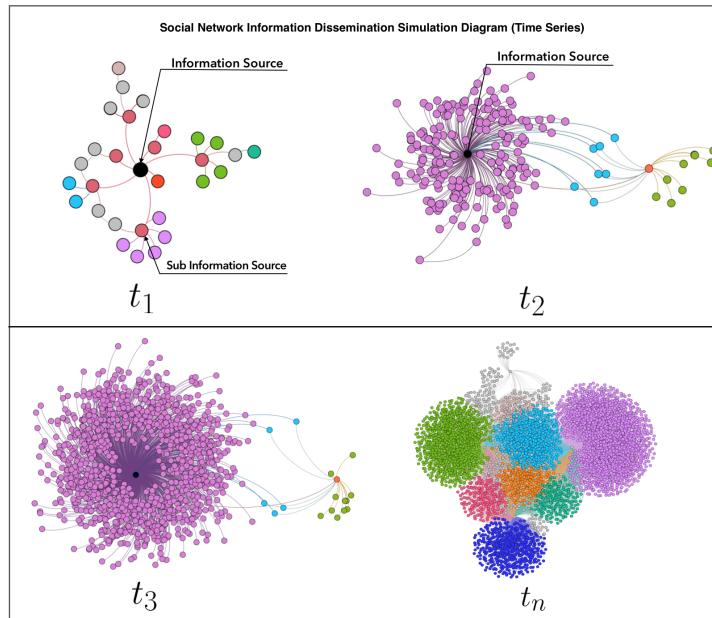


Figure 5: The social network information dissemination simulation diagram. The subgraph at the top left of the graph is the network structure diagram after the first transmission. The other three subgraph graphs are arranged according to the time of transmission.

From the Figure 5, at the beginning of the communication, each node is clearly visible. Since we do not consider the cross transmission of the same message, the message source can only continue to spread the message to the new nodes. This mode of communication will make the information show explosive growth, until all the nodes in the region have been informed. The last subgraph includes many different clusters of different colors and each cluster contains a huge number of nodes. This is the state of the network after the information disseminate reaches steady state.

After making information flow and filter model simulation, we verify the reliability of the network model, which has good adaptability and robustness.

## 5 Problem B : Forecast Information Network's Structure & Validate its Reliability

In order to verify our model's reliability, we have to build a short term prediction model to predict the information communication situation for today by using data from the past, and compare them with today's reality.

### 5.1 Short Term Prediction Model of Information Dissemination

In this chapter, we first describe the theoretical model of the building, and then build a realistic model of information dissemination.

#### 5.1.1 Model preparation

##### Node distinction

- Isolated node: there is no connection with other nodes at a certain time, and the next moment it may be linked to other nodes in a certain probability.
- Isolated cluster: there is no connection with other clusters at some point in the next moment, which may be related to other clusters at a certain probability. And there are interconnected nodes in the cluster.
- Large scale network cluster: cluster network is composed of many nodes or clusters, and there are direct or indirect links between nodes or clusters in the network.

*Assuming there is no absolute isolated node or cluster, any node or cluster is directly or indirectly connected with other nodes or clusters.*

#### 5.1.2 Realistic network model construction

Next, we build the realistic network model by steps.

##### (1) Constructing a dissemination network

We define that there are  $N_r$  nodes and  $M_r$  directed edge in the network. Meanwhile, we classify the nodes according to the node's own attribute and introduce the influence factor  $\gamma_i$ ,

which would change with the change of the node itself. The  $\gamma_i$  represents the new attention degree of a certain node in the unit time  $\Delta t$ .

$$\gamma_i = \frac{\Delta D_i}{\Delta t} \quad (5)$$

where  $D_i$  stands for the degree of Node  $i$ .

## (2) Adding the new people or group to the network

In each of the identified time series, the new node has two different ways to join the network.

One is establishing the connection relationship with the existing node freely. The probability of being associated with an already existing node is  $p_i$ ,

$$p_i = \frac{D_i + \gamma_i}{\sum_{i=0}^{N_r} (D_i + \gamma_i)} \quad (6)$$

The other is receiving the invitation, in accordance with the principle of common neighbor selection node to establish a connection. After the connection is established, according to the principle of preferential connection, the node will connect with the other nodes.

The probability of connecting with the node  $i$  is  $q_i$ ,

$$q_i = \lambda \beta_i + (1 - \lambda) \frac{D_i + \lambda_i}{\sum_{i=0}^{N_r} (D_i + \lambda_i)} \quad (7)$$

where  $\lambda \in [0, 1]$  stands for whether the node is invited. If not,  $\lambda$  taken 0,  $\beta_i \in [0, 1]$  represents the probability of common adjacent nodes to promote the evolution.

$$\beta_i = \frac{g_{ij}}{G} \quad (8)$$

where  $g_{ij}$  stands for the number of having same neighbor between Node  $i$  and Node  $j$ ;  $G$  is the total neighbor number of Node  $i$ .

## (3) Adding the new relationship to the network

For the existing nodes in the network, the probability of the link between the two nodes is  $O_i$ ,

$$O_i = (\beta_i + \frac{D_i + \lambda_i}{\sum_{i=0}^{N_r} (D_i + \lambda_i)}) / 2 \quad (9)$$

where  $\beta_i$  is the probability of common adjacent nodes to promote the evolution.

## (4) Determine the upper limit of the relationship

Taking into account the actual factors, people's energy is limited. So there is an upper

bound on the degree of the node. Before a degree does not reach the ceiling  $I$ , the broken link probability between nodes is very small.

We assume that  $k$  is the number of other nodes, which are followed by a certain node. Then the probability of breaking one link is  $Z_k$ ,

$$Z_k = 1 - \lambda\beta_k + (1 - \lambda) \frac{D_k + \lambda_k}{\sum_{k=0}^{N_r} (D_k + \lambda_k)} = 1 - q_k \quad (10)$$

### (5) Calculation of model evaluation index

In order to network structure, we ought to select several evaluation index. After choosing, the calculation formulas of model evaluation index are as follows:

$$AD = \frac{1}{N} \sum_i K_i \quad (11)$$

where the  $AD$  stands for the average degree,  $K$  is the degree of nodes.

$$L = \frac{1}{n(n-1)} \sum_{i+1} d_{ij} \quad (12)$$

where the  $L$  is the average path,  $n$  is the number of nodes, and  $d_{ij}$  is the shortest distance.

$$C_i = \frac{2l_i}{k_i(k_i-1)} \sum_{i+1} d_{ij} \quad (13)$$

where  $C_i$  represents for the average clustering,  $K_i$  is the degree of Node  $i$  and  $l_i$  is the actual variable between the adjacent points of the Node  $i$ .

## 5.2 Model Simulation Process

When the initial number of nodes is much smaller than the size of the final network, we are convinced that the number of initial nodes has no influence on the evolution of the network. So we chose the initial number of nodes is 3, and these 3 nodes constitute a complete graph. Meanwhile, We define a time period. In this period of time, we will add several nodes to the network randomly (up to no more than 10). The probability of being invited to joining the network and joined the network spontaneously is both 50 percent.

After joining the network, the newly added nodes are connected with the nodes which already exist in the network according to the probability of  $p_1$  and  $p_2$ . In each time period, for the nodes which have already existed in the network, connect two unlinked nodes in accordance with the probability of  $q_1$ , break the edge in accordance with the probability of  $q_2$ .

In order to find the answer, we need to continue fitting and determine the value of  $p_1, p_2, q_1, q_2$ . For simplicity, we assign fixed values to the  $p_1$  and  $p_2$  after a small range of tests. Furthermore, in the subsequent simulation, the  $p_1$  and  $p_2$  will be considered as fixed values.

We select the reference index is the average degree  $AD$  of the network, and then use the ***least squares method*** to take regression analysis, determining the parameters which have the smallest error .

The specific process of model simulation is

- (1) In the evolution of the network, when the network evolute to a state where has the same size with the existing referenced network, we mark this time as an observation point. At the end of evolution, we would compare the measured values of these observation points with the actual indicators of the existing reference network, and calculate the *square sum of errors*.
- (2) Erase data, and repeat the first step of the experiment 100 times.
- (3) After determining the parameters, we would make five simulation in each network size comparison point. And then, make the calculation of the each index to obtain the average value and get the final simulation data.

### 5.3 Forecast & Validation Results

After the simulation of the model, we have obtained the data, and then compare it with the actual data. The results are as follows.

#### 5.3.1 Forecast Results

Several results of forecast are listed in the Table 1 as follows.

Table 1: The Forecast Date of Information Network's Characterization (2010-2014)

year	simulate average degree	simulate average clustering	simulated ASPL
2001	13.01939492	0.419213291	4.801293321
2002	16.19395812	0.429192399	4.56234903
2003	19.94393111	0.435029123	4.423443242
2004	21.49283818	0.441919291	4.39919291
2005	22.89182849	0.450203023	4.311232131
2006	24.19294829	0.458049292	4.19191011
2007	24.94812912	0.462838322	4.16991911
2008	25.39181938	0.46478866	4.156568507
2009	25.59182389	0.467667754	4.151586061
2010	25.89182383	0.47086898	4.14725192
2011	26.10592911	0.472843919	4.141046295
2012	26.34929111	0.473818239	4.139822779
2013	26.72394393	0.47581239	4.132812418
2014	27.12939219	0.478923844	4.110359
2015	28.12939321	0.481192393	4.098181283

*ASPL: average shortest path length*

From the data we obtain above, the *Average Euclidean Distance* is 1.784 .

From Table 1 above, we find that under the condition of having errors, the growth of forecast data is relatively slow. Therefore, we can conclude that in the relatively stable society, the network structure is in the occurrence of slow evolution. Meanwhile, the network model we established can adapt to the change of time.

### 5.3.2 Validation Results

The forecast data and the real data comparison chart is drawn in the Figure6 as follows.

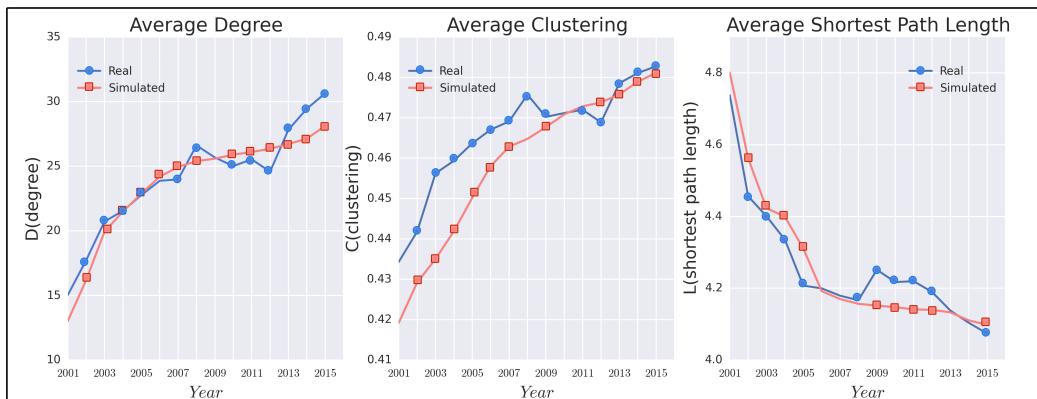


Figure 6: Simulation results and actual data comparison chart. We compare the *Average degree*, *Average clustering* and *Average shortest path length*.

From Figure 6 above, and compare with the real data, with we can draw the conclusion: Under the condition of having errors, we believe that the simulation data and the real data fitting degree are very high, the changing trends of various indicators are the same. Therefore, as far as we are considered, the model is reliable and can predict the change of network structure in a short time.

## 6 Problem C : Long Term Network Evolution Model

The change of information propagation model is divided into the slow change of the short time and the significant change of the long time. If the new revolutionary means of communication would be invited, the information dissemination model will be a complete change. We also need to set up an evolution model to predict the long time changes after the short information transmission is predicted.

### 6.1 Model building

In order to measure the change of information media with time, we need to find an index. After analyzing the historical data and the network structure, we define the *network dissemination efficiency* as the main index of the model, meanwhile, analyze it.

Through the analysis of the data, we find that in the emergence of new communication media, the usage rate of old communication media will be gradually reduced, but only a few communication media's usage rate tend to be 0, such as the telegraph. The remaining communication media will exist in the society according to a certain proportion, and affect our society in accordance with a certain proportion.

In this question, we use the  $E$  to represent the dissemination efficiency of the network.

$$E = \sum a_k E_k \quad (14)$$

where the  $E_k$  is the dissemination efficiency of the network of the  $k^{th}$  medium.

## 6.2 The bold assumption of future communication technology

*Innovative ideas require bold assumptions.* Our analysis of the evolution of information media from 1870s to 2010s found that the emergence of new media will greatly change the pattern of information dissemination.

The following Figure 7 is the timeline of the main information dissemination media changes.

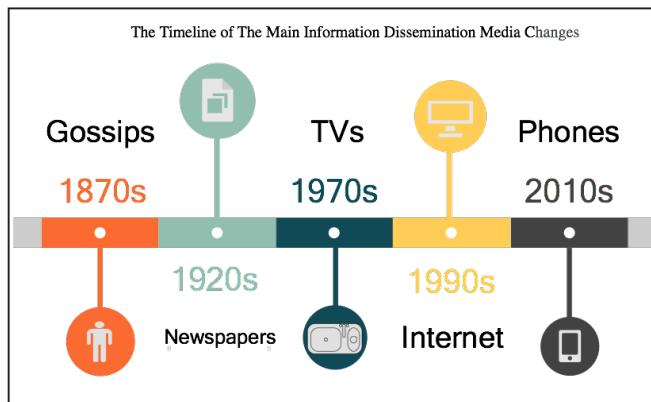


Figure 7: The timeline of the main information dissemination media changes

Under these circumstances, we are fully convinced that new information media is likely to be invented before the 2050. After consulting the relevant information<sup>[11]</sup>, we found that *Quantum Communication* is likely to go on the stage of history.

The definition of *Quantum Communication* is the high performance communication using the quantum effect under the condition of physical limit. Quantum communication has many advantages as follows:

- Quantum cryptography provides a new method for absolutely secure communication.
- Quantum Communication with good anti-interference does not depend on the transmission medium feature.

Because quantum communication has so many advantages and its research is also smooth, we boldly assume: before 2050, the quantum communication will come into being. Meanwhile, taking into account the process of quantum communication and the difficulty of commercial use, we assume that in 2035, the quantum communication will be officially entered the commercial market.

## 6.3 Model solving

We collect the data of the proportion of various media from 1990 to 2015, and bring them into our model to solve. The result are as follows.

The left figure above is the fitting result of date from 1990 to 2015<sup>[10]</sup> of the network efficiency. The right pie chart is the proportion of different information dissemination media (around 2050), from which the phone and the internet will become the most popular information dissemination media. Meanwhile, the radio, TVs and newspaper will gradually die out.

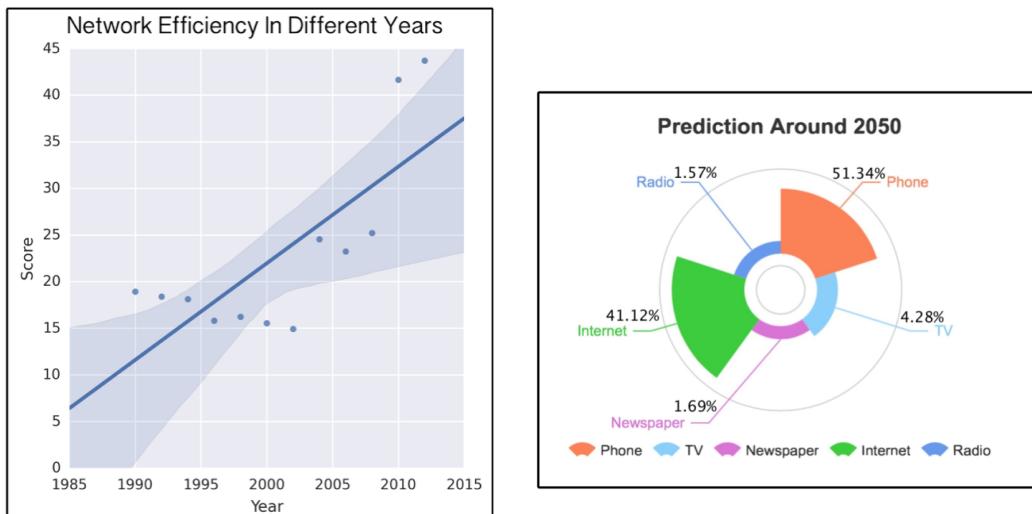


Figure 8: The Network Efficiency in Different Years (1990-2015)

Figure 9: The Proportion of Different Information Dissemination Media Pie Chart (around 2050).

The life line graph is the proportion of different information dissemination media around 2050 (*assume there is no new information dissemination technology appearing*). Left of the Dashed Line shows the real data collected online, while right display the predicted data.

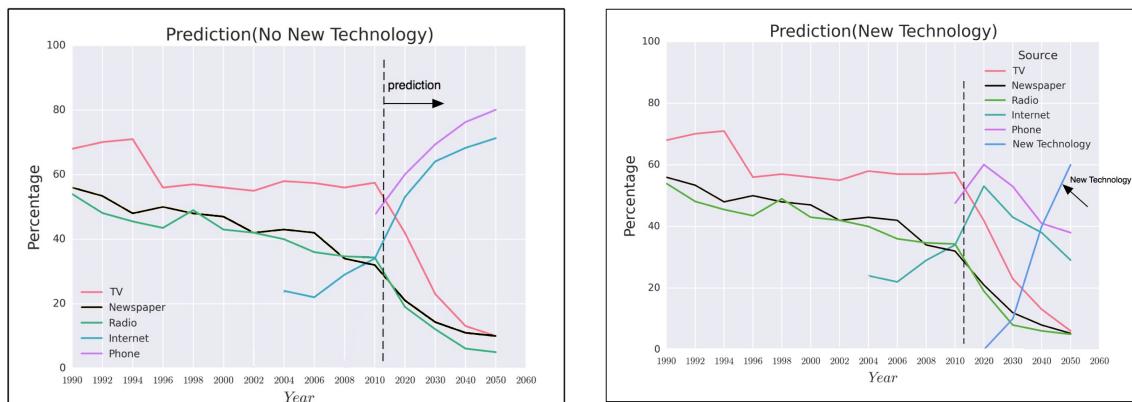


Figure 10: The Forecast Curve of Different Media Using Proportion (No New Technology)

Figure 11: The Forecast Curve of Different Media Using Proportion (Invent New Technology)

The right line graph is the proportion of different information dissemination media around 2050 (*assume there has new information dissemination technology appearing*). Left part chart of the Dashed Line shows the real data collected online, while the right one displays the predicted data. The line which is pointed represents the time that the new technology turning up in the future which tends to influence the society.

We use the data obtained from the ITU to predict the condition of information dissemination media in 2050, as the same time, we also get the data verify the reliability of the model. Not only did we obtain the data based on the prediction of the emergence of new technology, but also through access to relevant information, we can boldly gain the date based on have a new technology coming into being. The model is comprehensive and accurate.

## 7 Problem D : Public Opinion Dissemination Model

As we all know, information is not a constant when it spreads. Information will be changed with the spread. At the same time of information dissemination, the public opinion will come into being. In this chapter, we will study the mechanism of public opinion dissemination, and establish a model.

When constructing the model of public opinion, we believe that different people will have different influence on public opinion. The difference between the first model and this question is: when we construct the first model, we assume that the contribution rate of each node is the same, and the noise of information is not considered in the process of dissemination. However, in this chapter, we construct the network model of public opinion which has different nodes.

We can divide the network nodes into the three different types.

- Individuals: People or group.
- Opinion Leader: Determine the nature of the event in the network, and predict the future trend of the event.
- Internet Marketer Group: Internet Marketer Group is a certain scale of the people. They have a certain influence on the event of large-scale nodes in the network. This kind of influence is not related to the distance.

In a network of public opinion, the schematic map of different typical nodes is as follows.

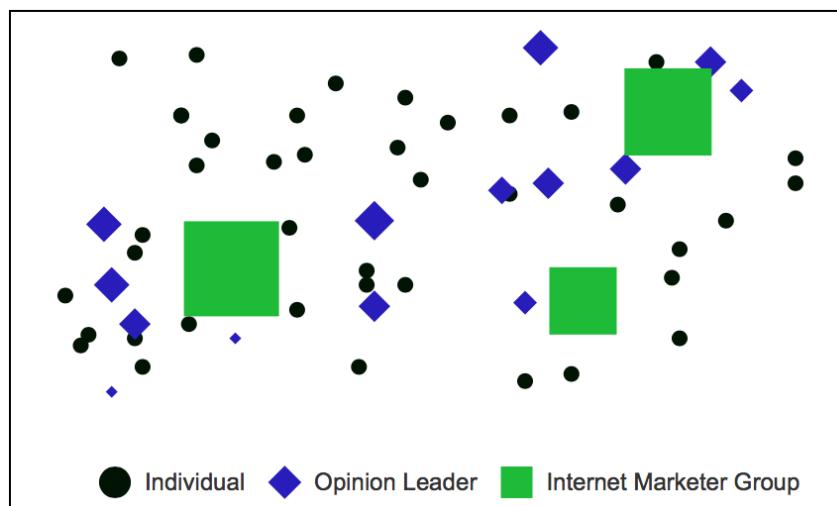


Figure 12: Schematic diagram of different typical nodes distribution. The number of **Individuals** is the biggest, while the number of **Internet Marketer Groups** is smallest, but its influence is very big. Opinion leaders have a large influence and their quantity is not fixed.

After analyzing the classification of the network nodes, we use the Twitter's data we find to simulate.

On the basis of the network established in the first question, we select nodes with maximum degree, and take them as the **Internet Marketer Group**. Meanwhile, we choose the nodes which have the second highest degree as the **Opinion Leader**. We also define the other remaining nodes as the **Individuals**.

After defining the nature of the network nodes, we use the *Information Network Flow Model* to simulate this chapter's model.

After a period of time, we take the operation of *Push* to network to improve the activity of the network information source, that is, to improve the success rate of information dissemination.

The results are shown in Figure 13 as follows.

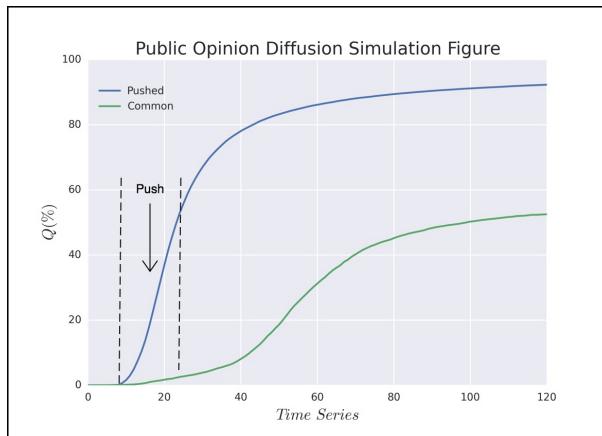


Figure 13: Public opinion diffusion simulation figure.

In Figure 13, after taking operation of *Push*, the message dissemination speed and the efficiency is greatly improved, showing the logistic growth, from which we can find that the operation of the network has a significant impact on the network. *Internet Marketer Group* and *Opinion Leaders* play an important role in the network of public opinion. Therefore, our *Public Opinion Dissemination Model* is reliable and stable.

## 8 Problem E : Analyze the Factors that Influence the Public Opinion Dissemination

In the first question, we set up a model of public opinion dissemination. In this chapter, we will analyze the factors that influence the spread of public opinion. At the same time, a real society is analyzed.

### 8.1 Influence Factors of Public Opinion Disseminate

There are many factors that affect the spread of public opinion, and we choose the following 4 factors as the main factors.

#### (1) Information value

The formula for calculating the value of information has been given in the first question. Information value can influence not only the speed of information dissemination, but also the public opinion. value of information is high, the probability of people receiving information and the degree of acceptance will be improved.

#### (2) People's initial opinions and prejudices

Psychologists believe that humans are "cognitive misers"<sup>[7]</sup>. People always simplify the cognitive process as much as possible. It is very difficult for people to change their ideas once they have produced a view or bias towards an event.

### (3) Form and source of information

The form and source of the information can impact the value of the message, which would also affect the public opinion. The evaluation formula of information's source and form have been given in the previous paper.

### (4) Network topology structure and strength

We will use several indexes of network topology to measure the efficiency of information dissemination. The topology or strength of the information network in a region, country, or worldwide is depended on the three index: Average Degree  $AD$ , Average Path Length  $L$ , and Average Path Length  $L$ .

Taking into account the indexes of the public opinion Dissemination are difficult to be quantified, for simplicity, we do not consider quantitative analysis of the spread of public opinion model, we only evaluate it subjectively.

## 8.2 Analysis One Historical Event based on Public Opinion Dissemination Model

In order to express more vivid description of the model we build, we look for a historical event, and analyze the historical events on the basis of our model.

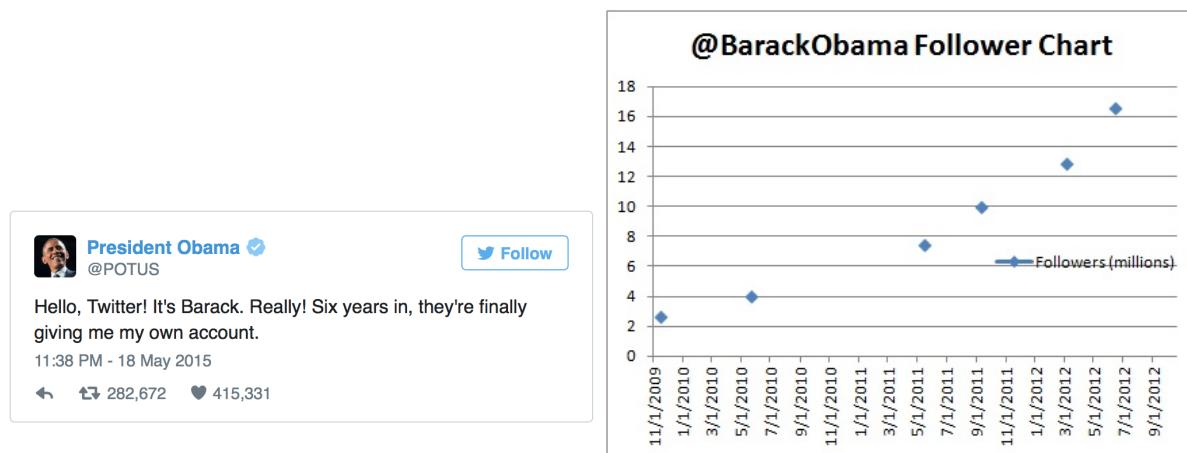


Figure 14: President Obama gets his own Twitter account. Figure 15: Graph of President Obama's follower growth.<sup>[8]</sup>

"Hello, Twitter! It's Barack. Really! Six years in, they're finally giving me my own account.". After President Obama opened his Facebook account, he opened his Twitter account on May 18, 2011.

From Figure 15, President Obama's followers has been growing rapidly over time. In our model, President Obama is one of the *Opinion Leader*, the news that he has released is a hot and authoritative topic of concern.

The analysis results of this case are very consistent with our expectations, therefore, this example verifies the accuracy of our model.

## 9 Strengths and Weaknesses

### Strengths

- (1) The information network model established by us has good universality, which is suitable for all kinds of information dissemination media, and has a good time-adaptability.
- (2) In our paper, we have simulated the most of the models we have built, and have been optimized to improve the stability and reliability of the model.
- (3) There are many different factors that affect the complicated problem. We classify the influencing factors and put forward the quantitative indexes. At the same time, we use of examples to test the quantitative indicators which are screened out. Only in this way, can we select out the well-founded indicators.

### Weaknesses

- (1) We lack the research on the characteristics of the network under different media.
- (2) In the *information flow and filter network*, the feedback function of the filter has few descriptions. The noise impact in the process of information dissemination has not been taken into consideration. In addition, the age structure of the node isn't studied in-depthly.

## References

- [1] D. E. KNUTH The TeXbook the American Mathematical Society and Addison-Wesley Publishing Company , 1984-1986.
- [2] Lamport , Leslie , L<sup>A</sup>T<sub>E</sub>X: " A Document Preparation System " , Addison-Wesley Publishing Company , 1986.
- [3] Wikipedia, Information flow (information theory). [https://en.wikipedia.org/wiki/Information\\_flow\\_\(information\\_theory\)](https://en.wikipedia.org/wiki/Information_flow_(information_theory)). 2016
- [4] Jure Leskovec, Andrej Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>. jun,2014.
- [5] Wikipedia, Information filtering system. [https://en.wikipedia.org/wiki/Information\\_filtering\\_system](https://en.wikipedia.org/wiki/Information_filtering_system). 2016
- [6] Schwarz J, Morris M. Augmenting web pages and search results to support credibility assessment[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011: 1245-1254.
- [7] Stanovich K E. What Intelligence Tests Miss:The Psychology of Rational Thought[M]. Yale University Press, 2009.
- [8] Wikipedia, Barack Obama on social media. [https://en.wikipedia.org/wiki/Barack\\_Obama\\_on\\_social\\_media](https://en.wikipedia.org/wiki/Barack_Obama_on_social_media). 2016
- [9] Pewresearch, State of the News Media 2012. <http://www.pewresearch.org/2012/03/19/state-of-the-news-media-2012/>. 2012
- [10] International Telecommunication Union, ICT Data and Statistics (IDS). <http://www.itu.int/ITU-D/ict/statistics/explorer/index.html>. 2016
- [11] Wikipedia, Quantum information science. [https://en.wikipedia.org/wiki/Quantum\\_information\\_science](https://en.wikipedia.org/wiki/Quantum_information_science). 2016