Team Control Number

47876

Problem Chosen

D

Abridge the Distance between Human Minds
—Research on Social Information Circulation

Summary

Information circulation network is as complicated as the human brain, while brain with wisdom seeks to explore the mysteries of the information circulation network.

First, we establish a model of information circulation network (ICP). Then we establish five network topology graphs and qualitatively analyze evolution of five periods. Based on classical epidemic model, we introduce the super-spreader, which can effectively accelerate the circulation of information as the central node. The differential equations and their results give us reasonable circulation laws: the density of the spreader and the super-spreader rapidly increase at the beginning, then they reach a peak and decline rapidly; the ignorant density shows a rapid downward trend while the immune density generally increases.

We build the fuzzy comprehensive evaluation model (NF) to filter what qualifies as news. We propose audience awareness index (AAI) to indicate the inherent value of information. Then we use four given examples to test the model. Undoubtedly, the assassination of Abraham Lincoln was qualified as news; and so does the important person's assassination today. As for information of Taylor's style transition, it becomes news today, but it might not be news in 1860.

Then we use the neural network prediction model (ICNP) to predict the networks' relationships and capacities. We get the node number and the node degree via previous data from three states in USA. And the results of ICNP model are: in the fifth period, node number of New York is 22535000, and its relative error is 15.32%; in the fifth period, node degree is 13533000, and its relative error is 20.58%. So it validates the reliability of ICP model. In around 2050, node number of New York is 34885000 and node degree is 23766000.

Finally, we build our model (PIINI) considering the interest attenuation mechanism and social strengthening mechanism. Taking the interaction of information circulation network and public interests into account, we establish differential equations similarly. And it comes to interaction laws: the density of the spreader and nodes in connection state rapidly increase at the beginning, they reach a peak and then decline rapidly. The ignorant density shows a rapid downward trend while the immune density generally increases.

As a part of ICM's Information Analytics Division, we have accomplished the given assignments.

Keyword: Information Circulation Network   Epidemic Model   Differential Equation, BP Neural Networks

Abridge the Distance between Human Minds
—Research on Social Information Circulation

Content

1. Restatement of the Problem

Information spread quickly in today's tech-connected communications network; sometimes it is due to the inherent value of the information itself, and other times it is due to the information finding its way to influential or central network nodes that accelerate its spread through social media.

❖ Explore the flow of information and filter or find what qualifies as news.

❖ Validate your model's reliability.

❖ Predict the communication networks' relationships and capacities around the year 2050.

❖ Explore how public interest and opinion can be changed through information networks in today's connected world.

❖ Determine how information value, people's initial opinion and bias, form of the message or its source, and the topology or strength of the information network in a region, country, or worldwide could be used to spread information and influence public opinion.

Interpretation of these problems

● We should build a model to indicate how information circulate in social network.

● We should figure out an improved model to filter what qualifies as news based on first model.

● We should predict present network's relationship and capacity to validate first model by using our model's prediction ability, and predict future network's relationship and capacity in around 2050.

● We should build another model to explore the interaction of public interest and information network.

2. Introduction

Social information circulation network is a very active research area. Building the social information circulation network is the hot area of complex network research. That research has very important value for understanding the dynamic behavior of the network.

Previous work

In previous analysis, most of models are based on the classical epidemic model SI, SIS and SIR, etc. [1, 2] The IC model [3, 4] proposed by Jacob Goldenberg and the LT model [5] proposed by Mark Granovetter is most widely studied currently. And Jaewon

Yang and other researchers proposed a linear influence model [6] by the analysis of Twitter's users' behaviors.

From the aspect of social psychology, the mass information circulation process is placed in a environment affected by a variety of social forces and psychological factors. Massa described the analysis and prediction of perceived user behavior based on trust framework [7]. Jamali and other researchers [8] use TrustWalker model to predict behaviors' of users. Josang A [9] build a user trust prediction system to describe the development of trust between different users.

3. Assumptions

Table 1 assumptions

	Assumptions
Overall	1) We do not take time delay during information circulation between two nodes into account.
ICN model	2) We think the public interest is consistent.
	3) We do not take social strengthening into account.
	4) We do not consider the difference between different information.
	5) We think the influence between two nodes will not affect our model.
NF model	6) We think the audience awareness is the only indicator of public interest and opinion.
ICNP model	7) We do not consider secondary factors.
	8) We think the network is universal around the world.
PIINI model	9) We do not consider secondary factors.

4. Justification of Our Approach

ICN model

Dynamics of epidemic disease can effectively portray on the circulation characteristics of complex network, which is an important basis for the current network circulation dynamics. Super-spreader plays an important role in the information circulation network. On the basis of classical epidemic model we introduce of the super-spreader and establish differential equations which can effectively portray information circulation network's characteristics.

NF model

News is objective report of what audiences are concerned about. The concept of audience awareness is vague, but we can qualitatively analyze audience awareness index has a positive influence on inherent value of information. The fuzzy comprehensive evaluation model can characterize the transition state between news and information. Therefore, it is appropriate to use NF model to filter what qualifies as news.

ICNP model

Information circulation network is a complex and random system, influenced by many factors. Using the statistical forecasting model to predict might have defects

especially when the sample data are not enough and when historical data are volatile. And we use BP neural network to build ICNP model, which can overcome this shortcoming.

PIINI model

Strengthening social mechanism is an important distinction between the information circulation and the spread of disease. Social interest attenuation mechanism is also significant in information circulation network. Proposed PIINI model takes social interest attenuation and social strengthening into account, so it can be effectively and reasonably reveal laws in the information circulation network.

5. Symbol Descriptions

There are some major symbols appear in the model, as shown below:

Table 2 symbol descriptions

Index	Definition	Formula
Node number(N)	The number of nodes include connected and disconnected nodes.	--
Node degree(k_i)	The degree of nodes is the number of connected nodes. And give adjacent matrix $A = \left(a_{ij} \right)_{N \times N}$.	$k_i = \sum_{j=1}^{N} a_{ij} = \sum_{j=1}^{N} a_{ji}$
Node degree average (<k>)	The average of all node degrees of the network.	$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \sum_{i=1}^{N} a_{ji}$
Network density(D)	Network density is a symptom of network's completeness.	$D = \frac{2L}{N(N-1)}$
Clustering coefficient(C_i)	Clustering coefficient represent the probability of any two nodes being neighbors. E_i is the edges between node i and neighbors k_i.	$C_i = \frac{E_i}{\left(k_i (k_i - 1) \right)/2} = \frac{2E_i}{\left(k_i (k_i - 1) \right)}$

Notes: we will explain other symbols when we use them.

6. The Model

6.1 Information circulation network model (ICN)

Social network is a complex network, which is composed of social individuals as well as the social relationships between individual members.[10] In the related social network research, researchers used $G = (V, E, W)$ graph to build model. V is the set of network nodes; $E \subseteq V \times V$ is the set of relationships between social individuals; W is the weight of social individuals.

6.1.1 Topological Properties

Topological properties of network are irrelevant with size, location, shape and function of nodes. However, they are only relevant with number of nodes and the connection features between connected nodes.

In the research of social networks, directed graph has a better expressive ability and applicability, so we use a directed graph to characterize social networks. And based on this, we present the following indices.

6.1.2 The model construction

There are lots of classical epidemic models, such as SI, SIS, SIR, IC and LT. We take into account the central node (such as the Internet) has higher spreading function than ordinary spreader. We then introduce super-spreaders to our network to improve the SIR model. We divide the user into four parts: **ignorant**, **spreader**, **super-spreader** and **immune**. And the definition of these are as follows[11]:

Table 3 Four type of users

Types of users	Definition
Ignorant	The ignorant has not received any information from neighbors.
Spreader	The spreader receives information, the spreader will probably spread the information to neighbors.
Immune	The immune will not spread the information after receiving the information.
Super-spreader	The super-spreader is able to quickly deliver information to more people.

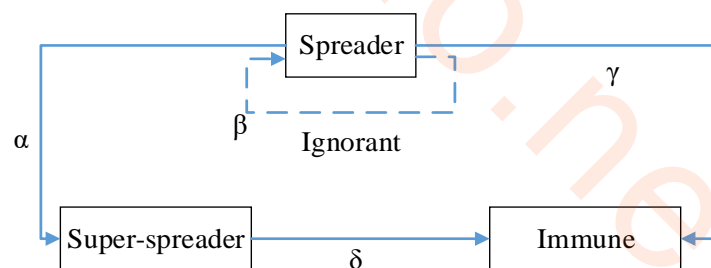**The rules of information circulation**

Fig 1 the propagation rules of the model

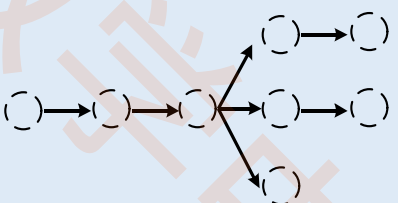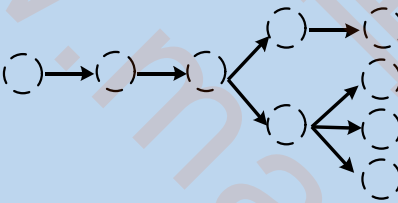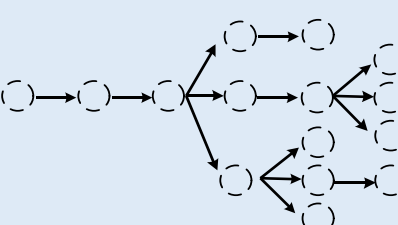The rules of information circulation are shown above. We can conclude the rules as follows:

- If ignorant connects to a spreader, they will turn into spreader with a probability β.
- For publicity and other reasons, some spreaders will turn into super-spreader with a probability α.
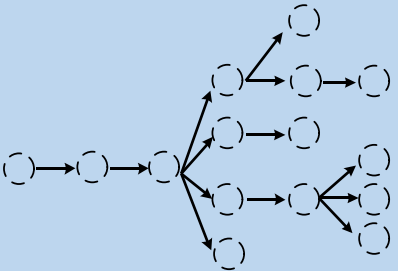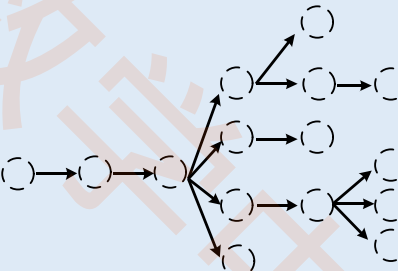- After contacting with other spreaders, the super-spreader will turn into the immune with probability δ.

➕ After contacting with other spreaders, immunes or super-spreaders, the spreader will turn into the immune with a probability γ.

The information circulation network in five periods.

On the basis of the establishment of information circulation network model, considering the characteristics of modes information circulation in five periods, we list the network circulation information model in five periods as follows:

Table 4 the information circulation graph and characteristics

Periods	Modes	The information circulation graph	Characteristics
1870s	Newspaper & Telegraph		(1)The number of node, the degree of node, internet density and clustering coefficient are small (2)"Newspaper office" node is regarded as super-spreaders and plays an important role in the information network. (3)The closeness of information circulation is poor. Regional links between users are strong. (4)Spreading flow is small and spreading speed is slow.
1920s	Radios		(1)Same to the first period, the number of node, the degree of node, internet density and clustering coefficient are small. (2)"Radio" node is regarded as super-spreaders and plays an important role in the information network. (3)The closeness of information circulation is poor. Regional links between users decrease. (4)Compared with first period, the spreading flow and spreading speed increase signally.
1970s	TV		(1)The number of node, and internet density are large, while the degree of node and the clustering coefficient are small. (2)"Television station" node is regarded as super-spreaders and plays an important role in the information network. (3)The closeness of information circulation is better. Regional links between users is poor. (4)Compared with second period, the spreading flow and spreading speed increase signally.

| 1990s | Internet | | (1)The number of node, the degree of node, internet density and clustering coefficient are large. (2)The condition of super-spreader is inconspicuous. (3)The closeness of information circulation is well. Regional links between users is poor. (4)Compared with third period, the spreading flow and spreading speed increase signally. |
| 2010s | Phone | | (1)Compared with the fourth period, the number of node, the degree of node, internet density and clustering coefficient are larger. (2)Some nodes in internet are regarded as super-spreaders and play an important role in the information network. (3)The closeness of information circulation is well. Regional links between users is poor. (4)Compared with fourth period, the spreading flow and spreading speed increase signally. |

Notes:

✧　Which nodes do not enter the network are not shown in the graphs.

✧　Solid circle represents the super-spreader, and dotted circle represents the spreader.

The differential equation of information circulation network.

Based on three circulation principles, we build differential equations as follows:

$$\begin{cases} \dfrac{dI(t)}{dt} = -\beta I(t) \\[2mm] \dfrac{dS(t)}{dt} = -\alpha S(t) + \beta I(t) - \gamma S(t) \\[2mm] \dfrac{dJ(t)}{dt} = \alpha S(t) - \delta J(t) \\[2mm] \dfrac{dR(t)}{dt} = \gamma S(t) + \delta J(t) \end{cases} \tag{1}$$

In these equations, $I(t)$, $S(t)$, $J(t)$ and $R(t)$ respectively represent, at time t, the density of the unknown, the spreader, the super-spreader and the immune. The density of four types of users meets the condition of normalization: $I(t) + S(t) + J(t) + R(t) = 1$. When $\alpha = 0$, $\delta = 0$, this model will turn into classical SIR model. With the evolution of the system, when the number of spreaders is zero, the process stops and spread of network ends. And the network will only contain those who have not yet received any information and the immune who are transformed from the spreader and super-spreader.

6.1.3 To solve the model

We use Matlab 2014b software to solve differential equations. By means of numerical simulations, the number of configured nodes is 1000 in the network, the initial number of spreaders is 1.

In the initial stage, $S(t=0)=1/1000$, $I(t=0)=999/1000$, $J(t=0)=R(t=0)=0$, and we assume $\alpha=0.1, \beta=0.7, \delta=0.5, \gamma=0.3$ [17]. Fig 1 shows general trends of four types of nodes' density.
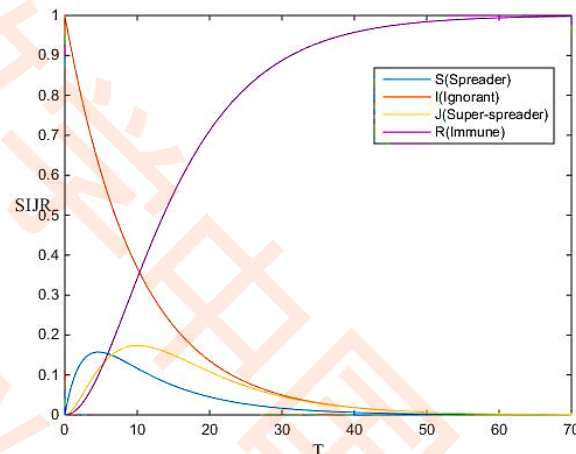

Fig 2 densities of different groups in the model

According to the analysis of fig 2, we can conclude:

- **Law one**: Spreader' density rapid increase at the initial stage. After it reaches a peak, it declines rapidly, until the spreader enters into the immune state.
- **Law two**: The evolutionary trend of super-spreaders like that of ordinary spreaders, and it also experiences an increase, and then decrease until it disappears. However, the curve changes gently. The peak density is also much smaller than ordinary spreaders. This shows that in the online social networks, super-spreaders are a small number of "star nodes."
- **Law three**: Ignorant density shows a rapid downward trend. And it eventually settles to a non-zero value. And that reflects the information circulation's breadth and depth are limited and it is impossible to cover all users.
- **Law four**: The trend of immune density is opposite to that of the ignorant. It experiences a rapid increase stage and then gradually stabilizes to a value close to 1. Rapid increase of the immune density is mainly because spreaders and super-spreaders continue turning into the immune state. And the increasing trend slows because there is no new spreader.
- **Law five**: It is worth noting that, when information circulation stops, the density of the spreader and the super-spreader are equal to zero, but the density of the ignorant and the immune are not zero. It indicates that there are users who does not receive the information in the network. The size of information circulation has a certain limit, and it is impossible to "infect" all users.

6.2 News filter model (NF)

News is defined as a report for what audiences focus [12]. The dogs bite human beings is normal, while that human beings bite dogs will become news. The reason is we focus more on human beings biting dogs. So in the light of audience's awareness, we define **audience's awareness index** (AAI). By determining the value of AAI, we can filter what qualifies as news.

6.2.1 Influential factors of audience's awareness index

The number of audiences, the political status of audiences and the influence of information have an influence on AAI [13]. And the analysis of these factors are as follows:

 + The number of audiences

The number of audiences is the number of persons who are concerned about this information. The more the number of the audiences is, the higher the AAI is.

 + The political status of audiences

Higher political status of audiences, the more it can arouse the audience's attention, so AAI will be higher.

 + The influence of news

The higher the influence of news, the higher the value of AAI.

According to foregoing analysis, there are positive correlations between the AAI and these factors.

6.2.2 The model construction

In order to calculate, we build the fuzzy comprehensive evaluation model. First, we should determine index weight matrix. Then we improve the max min algorithm and the principal of maximum membership degree among evaluation results. We build the set of factors, the set of evaluation, membership function and the set of weight [13]. Last, we comprehensively evaluate AAI.

Step 1: Determine the set of evaluation object.

$$P = \{\text{Information to be evaluated}\}$$

Step 2: Build the set of evaluation index.

$$U = \left\{u_1, u_2, u_3\right\}$$

u_1, u_2, u_3 is the number of audiences, the political status of audiences and the influence of news respectively.

Step 3: Determine the universe of evaluation degree.

Build a set of evaluation v, each level can correspond to a fuzzy subset。

$$V = \left\{v_1, v_2, v_3\right\} = \{A, B, C\}$$

Step 4: Calculate index weight

Use paired comparison method and 1-9 comparative scale to calculate index weight, we have $A=[0.4,0.3,0.3]$

Step 5: Calculate fuzzy relationship matrix R

Evaluate U, and then figure out the membership degree. Finally get the fuzzy relationship matrix R.

Step 6: Calculate the vector of fuzzy evaluation results B

Use the weighted average of fuzzy arithmetic operators $M(\bullet,\oplus)$ to determine index weight matrix A and fuzzy relationship matrix R, and the formula is:

$$b_i = \sum_{i=1}^{p}(a_i \cdot r_{ij}) = \min\left(1, \sum_{i=1}^{p} a_i \cdot r_{ij}\right), j = 1,2,\cdots m \tag{2}$$

Synthesizing the A and R will get the vector of fuzzy evaluation results B. According to the definition, if $Bs>Bt$, the value of information s is higher than that of information t. The value of vector B is AAI.

Step 7: Calculate the value of AAI

We use $CAAI=(B_i^{'} \cdot [3,2,1])/3$ to indicate the value of AAI.

6.2.3 The process to test the model

We use the given four examples to test our model:
- ✓ Information a: country-turned-pop singer Taylor Swift's possible engagement today.
- ✓ Information b: country-turned-pop singer Taylor Swift's possible engagement in 1860.
- ✓ Information c: an important person was assassinated today.
- ✓ Information d: assassination of US President Abraham Lincoln in 1865.

Due to space limitations, the solution process omits steps 1-4. And according to step 5, we get the evaluation matrixes:

$$R_a = \begin{pmatrix} 0.1 & 0.6 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.5 & 0.4 & 0.1 \end{pmatrix}, \quad R_b = \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.5 & 0.4 \end{pmatrix},$$

$$\tag{3}$$

$$R_c = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.8 & 0.2 & 0 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}, \quad R_d = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 1 & 0 & 0 \\ 0.9 & 0.1 & 0 \end{pmatrix},$$

We calculate the result of evaluation vector, and we have:

$$B_a = [0.30, 0.54], \quad B_b = [0.10, 0.46, 0.44],$$
$$B_c = [0.69, 0.24, 0.07], \quad B_d = [0.89, 0.07, 0.04]$$

Then according to step 7, we have:

$$CAAI_a = 0.71, \quad CAAI_b = 0.55, \quad CAAI_c = 0.87, \quad CAAI_d = 0.95$$

According to previous analysis, the greater value of CAAI, the greater number of audiences will be concerned about given information. So there are more reasons to explain why this information is qualified as news. In this four examples, that the US president was assassinated undoubtedly became news; and so does the important person occurred today. As for information of Taylor transition style, it is still is a news today, but if it happened 1870, perhaps it will not be concerned about.

6.3 Information circulation network prediction model (ICNP)

Information circulation network is a complex and random system, influenced by many factors. Using the statistical forecasting model to predict might have defects especially when the sample data are not enough and when historical data are volatile. Therefore, we improve the ICN model by adding BP neural network prediction model to previous model to predict information circulation network's node number and node degree.

6.3.1 Influential factors

Comprehensive analysis of factors that influence the spread of information networks is a prerequisite to improve ICN model by using its ability of prediction.

Based on the reality of social life and the complexity of finding data, we will select the following main factors:

➢ Level of regional **economic** development

The level of regional economic development is measured by regional GDP. The higher the level, the more beneficial to information circulation.

➢ Level of regional **communication** development

The level of regional communication development is measured by information communication technology development index (IDI) [15]. The larger the IDI, the more beneficial to information circulation.

➢ Level of regional **education** development

The level of regional education development is measured by education index (EI). The larger the EI, the more beneficial to information circulation.

6.3.2 The model construction

According to the foregoing ICN model, we will build ICNP model to predict the number of node and the degree of node. The number of node determine the size of

network, and the degree of node determine the capacity of the network.

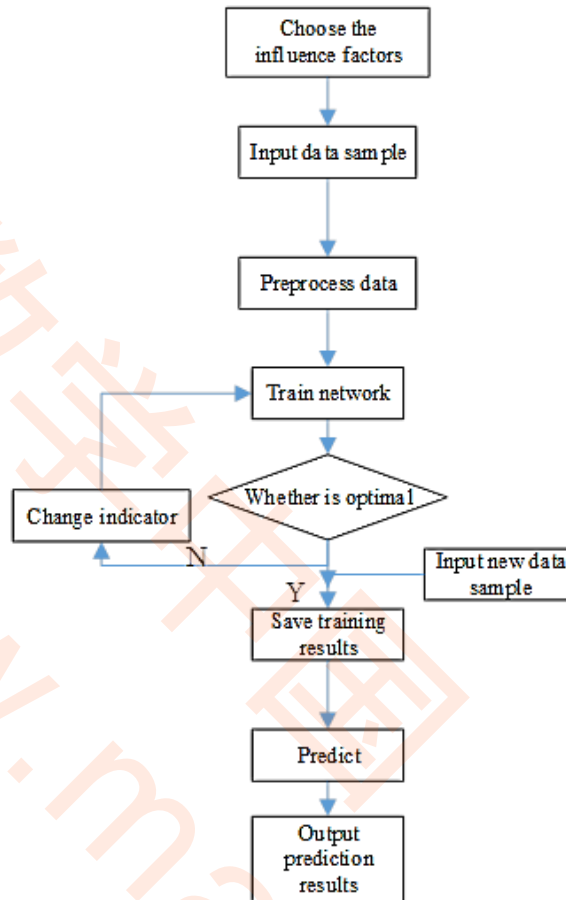Similar to the general neural network model, our ICNP model's prediction process is as follows:

Fig 3 ICNP model's prediction process

Data preprocess

In this paper, data are normalized by using maximum and minimum normalization method. The data normalized formula is:

$$X_i = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4}$$

Determine the input number of neurons

According to Hurvich's study, when sample data is less, if using ACI criteria, there will be great estimation bias. So, based on ACI criteria, Hurvich proposed AICC criteria specifically applicable to small samples. And the formula is :

$$AICC(m) = N \ln E + \frac{N+m}{N-m-2} \tag{5}$$

Among the formula:

$$m = \begin{cases} \dfrac{N}{2}, & N = 20 \sim 40 \\[2mm] \dfrac{N}{3} \sim \dfrac{N}{2}, & N = 40 \sim 100 \\[2mm] \dfrac{N}{\ln N}, & N = 100 \sim 200 \end{cases} \tag{6}$$

6.3.3 To solve the model

● Collect data

We collect from internet, and collect data in three states: California (CA), Texas (TX) and New York (NY) in USA. And these three states have top three largest population among fifty states in USA. The reason we choose these three states are because these data are easier to find and these data samples are larger.

The data and data source are shown in the appendix.

● Validate ICN model

We use data excluding the data from New York at fifth period. And then we use the ICNP model to predict the data in New York at fifth period and compare these data with present data to validate our ICN model.

Table 5 during the sample period prediction

	From ICN model	From predict model	Relative error (%)
Number of node(person)	19541453	22535000	15.32
Degree of node(person)	11223538	13533000	20.58

● Predict future network's relationships and capacities in 2050.

Table 6 during the sample period prediction

	ICNP model
Number of node(person)	34885000
Degree of node(person)	23766000

6.4 Public interest and information network interaction model (PIINI)

In the process of social information circulation networks, the circulation of information affects individuals' preferences. We also think that individuals' different preferences affects information circulation networks. That is same as what question said that public interest can be changed through information networks. In the meantime, information value, people's initial opinion and bias, form of the message or its source, and the topology or strength of the information network in a region, country, or worldwide could influence public opinion.

6.4.1 The model construction

Information circulation is intrinsically different from the characteristics of the spread of disease [17]. And its characteristics should consider a variety of factors. We consider only consider two factors: public interest decay and social strengthening effect Based on these two factors, we build PIINI model to reveal the interaction of public interest and information network.

Public interest attenuation mechanism

The public interest decay mechanism refers to the public interest decreases when the times of receiving same information increase [18]. If users repeatedly received the same information from neighboring nodes, users will acquiesce in information being widely disseminated. Then there exists attenuation coefficient [19] which indicates users will gradually decrease their behavior of circulating information. Therefore, we introduce a saturation function to represent public interest attenuation coefficient:

$$\zeta(m) = a - be^{-\delta m} \tag{7}$$

And a is the maximum attenuation rate, b is the adjustable coefficient, the δ is public interest attenuation coefficient and m is the number of connected users. The saturation is a non-linear increasing function. When the maximum attenuation rate $a = 1$, $b = 1$ adjustable coefficient, as m increases and nonlinear growth, indicating that the loss of interest and forwarded into a connected state more and more users. When $a = 1$, $b = 1$, when m and δ increases, the number of users losing interest increases. As shown in fig 4.
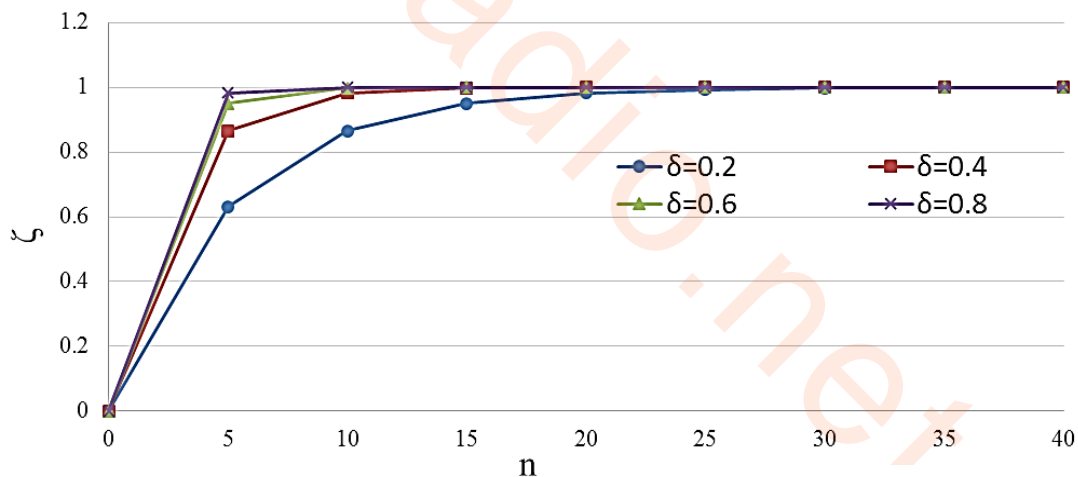

Fig 4 the mechanism of interest decay

Social strengthening mechanism

Social strengthening have nonlinear cumulative effect which have a significant effect on public interest [20]. Because of this mechanism, an individual, users will be affected by neighbors' cumulative effects before they adopt received opinions [21].

In our paper, we define the social strengthening mechanism as: the probability of a user's behavior (including the acceptance and dissemination of information) will

decrease while strengthening factor generated from neighbors' discouragement. The larger the strengthening factor, the lower the probability of a user's behavior (including the acceptance and dissemination of information). We use the following nonlinear function to explain the social strengthening mechanism.

$$\xi(n) = \frac{\beta n}{1 + \eta n} \tag{8}$$

β indicates the probability of a user's behavior (including the acceptance and dissemination of information), n indicates the number of connected node, η is the strengthening factor, $0 < \eta < 1$.
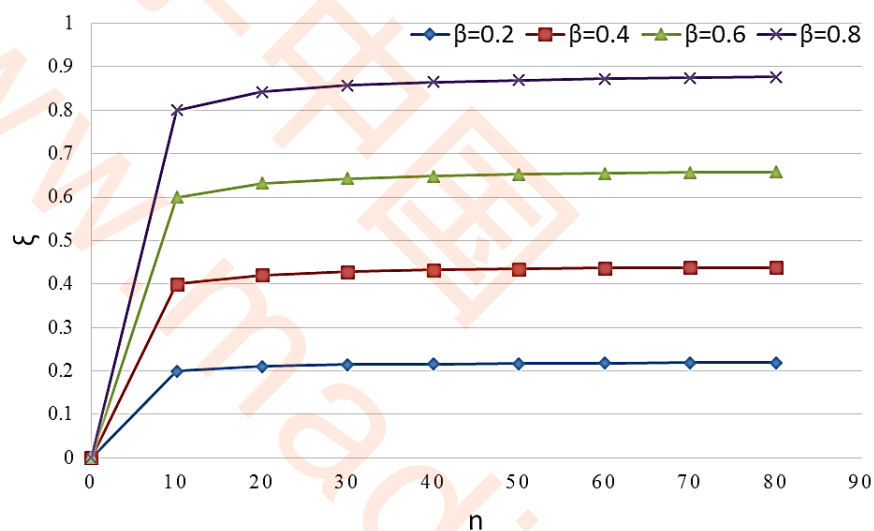

Fig 5 the social strengthening effect changed with β

As fig5 shows, the larger value of β, the larger value of n. In the meantime, the larger value of η, the smaller value of ξ.
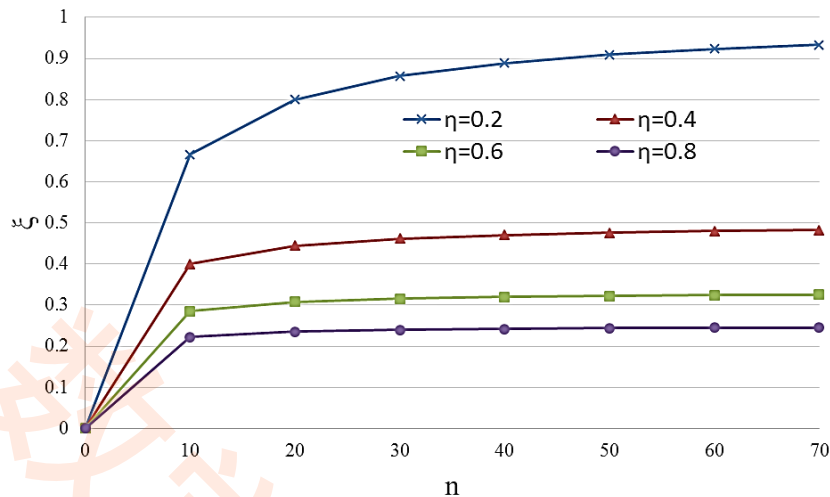
Fig 6 the social strengthening effect changed with η

As fig 6 shows, when the probability β is constant, the value of ξ will decrease when the value η increases. It shows that by increasing the effect of social strengthening, more users will stay in the unconnected state. And therefore the number of spreaders will decrease.

Improve the information circulation model

The state of connection defined as a state that users have already received information but they are not sure whether to spread the information. And the information circulation rules are as follows:

- If ignorant is concerned about received information, the ignorant will turn into spreader with a probability β; if ignorant is not concerned about received information, the will turn into the immune with a probability γ.
- With times of circulation increasing, the spreader will gradually lose interest in transferring information to others and turn into the immune with a probability of ξ.
- When the spreader meets with other spreader or immune, the spreader might not transfer the information with a probability of θ.
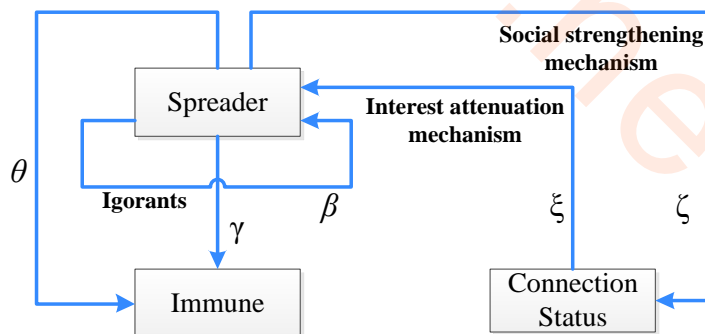

Fig 7 the circulation rules of PIINI model

The differential equations in PIINI model

$I(t)$、$S(t)$、$C(t)$ and $R(t)$ indicate the density of ignorant nodes, spreader nodes,

nodes in connection state and immune nodes when node degree is *k* at time *t*. The equations are shown below:

$$
\begin{cases}
\dfrac{dI(t)}{dt} = -\beta I(t) \\[2mm]
\dfrac{dS(t)}{dt} = \beta I(t) - (\theta + \gamma)S(t) - \zeta S(t) + \xi C(t) \\[2mm]
\dfrac{dC(t)}{dt} = \zeta S(t) - \xi C(t) \\[2mm]
\dfrac{dR(t)}{dt} = (\gamma + \theta)S(t)
\end{cases}
\tag{9}
$$

When $\gamma = 0, \zeta = 0, \xi = 0$, this model will degenerate to classical model SIR.

6.4.2 To solve the model

We use Matlab 2014b software to solve differential equations. By means of numerical simulations, the number of configured nodes is 1000 in the network, the initial number of spreaders is 1.

And $S(t=0) = 1/1000$, $I(t=0) = 999/1000$, $J(t=0) = R(t=0) = 0$ and we assume $\alpha = 0.1, \beta = 0.7, \theta = 0.5, \gamma = 0.3, \xi = 0.3, \zeta = 0.15$ [17]. Fig 8 shows general trends of four types of nodes' density.
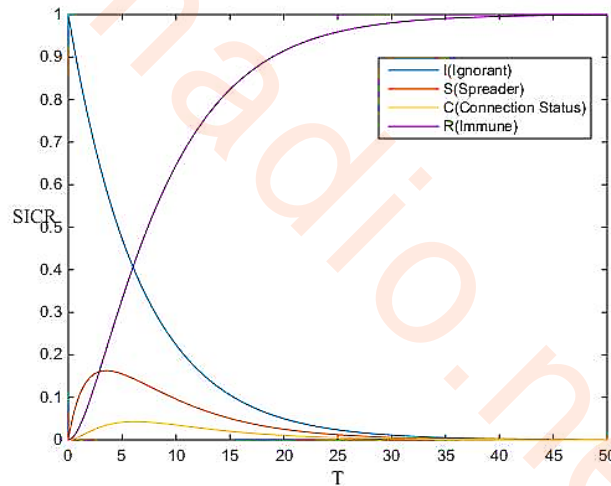

Fig 8 densities of different groups in the model

According to the analysis of fig 8, we can conclude:

+ **Law one**: Spreader' density rapid increase at the initial stage. After it reaches a peak, it declines rapidly, until the spreader enters into the immune state.
+ **Law two**: The evolutionary trend nodes in connection state like that of ordinary spreaders, and it also experiences an increase, and then decrease until it disappears. However, the curve changes gently. The peak density is also much smaller than ordinary spreaders. This indicates that in today's online social network nodes in the connection state are a small part of all nodes. And this state of node will change quickly into the spreading state.

- **Law three**: Ignorant density shows a rapid downward trend. And it eventually settles to a non-zero value. And that reflects the information circulation's breadth and depth are limited and it is impossible to cover all users.
- **Law four**: The trend of immune density is opposite to that of the ignorant. It experiences a rapid increase stage and then gradually stabilizes to a value close to 1. Rapid increase of the immune density is mainly because spreaders and super-spreaders continue turning into the immune state. And the increasing trend slows because there is no new spreader.
- **Law five**: It is worth noting that, when information circulation stops, the density of the spreader and the super-spreader are equal to zero, but the density of the ignorant and the immune are not zero. It indicates that there are users who does not receive the information in the network. The size of information circulation has a certain limit, and it is impossible to "infect" all users.

This laws are similar with previous ICN model when the super-spreader is introduced.

According to previous research, the size of information circulation [17]:

$$R = \sum_k P(k)\left(1 - e^{-k\varphi(t=\infty)}\right) \tag{10}$$

The formula indicates that the ultimate size of information circulation depends on the degree distribution $P(k)$. And the node degree plays an important role. It also verify our ICNP model which uses node degree as a viable scientific indicator to measure information circulation network. We introduce the auxiliary function $\varphi(t)$. So we need to calculate $\varphi(t=\infty)$ to indicate the size of information circulation.

7. Sensitivity Analysis

The ICN model

In ICN model, we set the parameters of differential equations, and obtain a stable solution. And it successfully describes the process of information circulation. In order to test the stability of the model, we will change the value of $\alpha,\ \beta,\ \gamma,\ \delta$, and change is set to $\pm 5\%$ to observe system's stability status.

We change only one parameter every time, and other parameters remain unchanged. Due to space limitations, we only list sensitivity analysis to parameter β with biggest change.

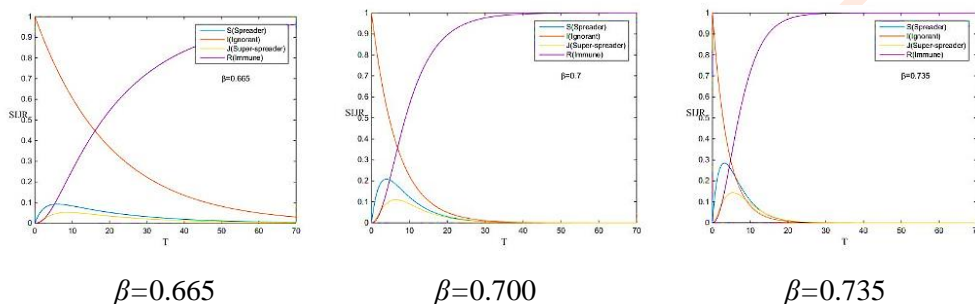β=0.665 β=0.700 β=0.735

Fig 9 the sensitivity analysis of β

According to fig 9, when adjust the value of parameters, the overall trend is consistent and the system can quickly reach a steady state, only the time of reaching the steady state is different. The larger the value of β, the faster the convergence.

For other parameters, by adjusting their values, change is not large. Through the above sensitivity analysis we validates our model is credible within a certain range of variation, which can be ignored.

The PIINI model

Similar to the previous analysis, we change only one parameter every time, and other parameters remain unchanged, to examine whether the status of the entire system is stable. During experiments similar conclusions can be obtained that the system is not sensitive to change of parameters. So our model is credible.

The ICNP model

We use the fit test to accomplish the error analysis. And coefficient of determination (R) is used in fit test to indicate the accuracy of our model. The bigger the R, the better the model results are. According to the fig 10, R is equal to 0.99, so the model is reliable.
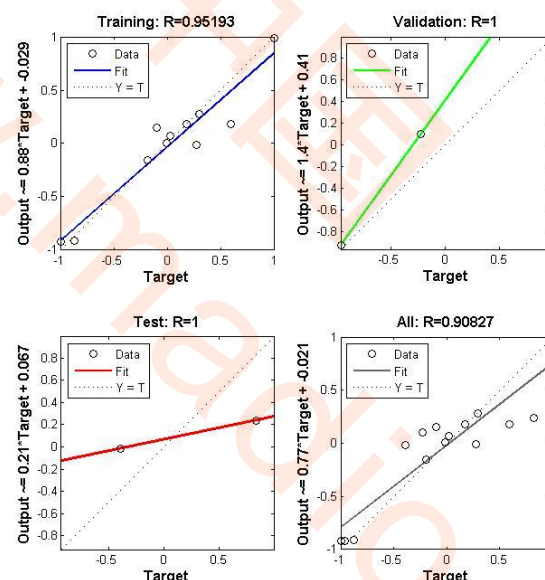

Fig 10 the error analysis of ICNP model

Due to space limitations, we only discuss number of hidden layer nodes, and coefficient of determination and epoch. The sensitivity analysis result is shown in table 7.

Table 7 the sensitivity analysis of ICNP model

Node number	1	2	3	4	5	6	7	8	9	10	20	50
Relative error	0.62	0.49	0.34	0.14	0.28	0.31	0.24	0.47	0.52	0.22	0.24	0.77
R	0.74	0.89	0.85	0.91	0.87	0.85	0.82	0.69	0.86	0.62	0.59	0.71
Epoch	7	9	17	14	10	8	11	15	7	8	6	5

8. Strengths and Weaknesses

Table 8 strengths and weaknesses

	Strengths	Weaknesses
ICN model	(1). Considering previous model, we take super-spreader into account. (2). The results of model conform to the reality well according to scond model results.	(1). Ignore influence between each nodes. (2). The result of differential equation is not stable.
NF model	(1). This model is simple, but universal and effective to solve given problem.	(1). The determination of index weight is subjective
ICNP model	(1). Compared with the statistical forecasting model, the ICNP has higher accuracy. (2). The ICNP model can successfully validate the first model.	(1). Collected sample data are not enough.
PIINI model	(1). The PIINI model can successfully indicate the interaction of public interest and information network.	(1). We only consider two major factors, but we ignore some other factors.

9. Conclusions

As a part of ICM's Information Analytics Division, we have accomplished the given assignments. Here is the overall results of our paper.

Based on improved epidemic model, we established five network topology graphs and come to several information circulation laws: the density of the spreader and the super-spreader rapidly increase at the beginning, then they reach a peak and decline rapidly; the ignorant density shows a rapid downward trend while the immune density generally increases.

In NF model, we conclude that the assassination of Abraham Lincoln undoubtedly becomes news; and so does the important person's assassination today. As for information of Taylor's style transition, it becomes news today, but it might not be news in 1860.

The ICNP model validate the reliability of ICP model and predict many parameters in the information circulation network. And the results of ICNP model are: in the fifth period, node number of New York is 22535000, and its relative error is 15.32%; in the fifth period, node degree is 13533000, and its relative error is 20.58%; in around 2050, node number of New York is 34885000 and node degree is 23766000.

And we establish differential equations and solve them and we come to interaction laws: the density of the spreader and nodes in connection state rapidly increase at the beginning, they reach a peak and then decline rapidly. The ignorant density shows a rapid downward trend while the immune density generally increases.

References

[1] Diekmann O, Heesterbeek J A P. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation [M]. Wiley, 2000.

[2] Girvan M, Callaway D S, Newman M E. Simple model of epidemics with pathogen mutation [J]. Physical Review E. 2002, 65 (3): 031915.

[3] Goldenberg J, Libai B, Muller E. Talk of the network: A complex systems look at the underlying process of word-of-mouth [J]. Marketing letters. 2001, 12 (3):211–223.

[4] Goldenberg J, Libai B, Muller E. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata [J]. Academy of Marketing Science Review.2001, 9 (3): 1–18.

[5] Granovetter M. Threshold models of collective behavior [J]. American journal of sociology. 1978: 1420–1443.

[6] Yang J, Leskovec J. Modeling information diffusion in implicit networks [C]. In Proceedings of the 10th International Conference on Data Mining. 2010: 599–608.

[7] Massa P, Avesani P. Trust-aware Recommender Systems. Proceedings of the ACM conference on recommender systems, 2007, 17-24.

[8] Jamali M, Ester M. TrustWalker: a random walk model for combining trust-based and item-based recommendation. Proceedings of the 15th ACMSIGKDD international conference on knowledge discovery and data mining, 2009, 397-406.

[9] Josang A, Ismail R, Boyd C. A survey of trust and reputation systems for online service provision. Decision support systems,43(2), 2007, 618-644.

[10] http://en.wikipedia.org/wiki/Social_network.

[11] YangR, WangB-H, Ren J. Epidemic spreading on heterogeneous networks with identical infectivity[J]. Physics Letters A,2007,364(3-4):189-193.Saito.

[12] https://en.wikipedia.org/wiki/News.

[13] https://en.wikipedia.org/wiki/Public_opinion.

[14] Yonghong Hu, Enhui He, Comprehensive evaluation metho[M].Beijing: Science Press, 2000: 167-188.

[15] http://www.itu.int/zh/Pages/default.aspx.

[16] http://www.undp.org/.

[17] Chaosheng Tang, Online social network information propagation modeling and prediction of forwarding[D]Yanshan: Yanshan University,2014.

[18] Liu Y, Xiao L, Liu X. Location Awareness in Unstructured Peer-to-peer Systems [J]. IEEE Transactions on Parallel and Distributed Systems, 2005,16(2):163-174.

[19]Dashu Tang, Mingke He, Qingsong Meng, Based on Power Law Distribution and Small World Character unstructured P2P network searching method [J]. Computer Research and Development,

2007, 44(9): 1566-1571.

[20] LüL, Chen D-B, Zhou T. The Small World Yields the Most Effective Information Spreading [J]. New Journal of Physics, 2011, 13(12): 123005.

[21] Karrer B, Newman M E J. Message Passing approach For General Epidemic Models [J]. Physical Review E, 2010, 82(1): 016101.

[22] Nekovee M, Moreno Y, Bianconi G. Theory of Rumor Spreading in Complex Social Networks [J]. Physica A: Statistical Mechanics and its Applications, 2007, 374(1): 457-470.

Appendix (data and data source)

Table 8 Collection data

	1870s			1920s			1970s		
Index	CA	TX	NY	CA	TX	NY	CA	TX	NY
IDI	1.2	0.9	0.8	1.3	1.2	1.1	2.1	1.9	1.7
EI	0.736	0.723	0.722	0.834	0.823	0.832	0.856	0.845	0.845
GDP(million US$)	1006	615	594	11492	7027	6788	46640.4	28522.2	27553.8
Number of node(person)	4574962	3067451	2418760	24465442	16403752	12934761	29943110	20076455	15830777
Degree of node(person)	2407875	1614448	1273032	4286735	2874198	2266376	5926493	3973634	3133308

Table 9 Collection data

Index	CA	TX	1990s	2010s	TX	NY
IDI	3.4	3.2	2.9	4.39	3.87	2.34
EI	0.876	0.877	0.867	0.968	0.976	0.987
GDP(million US$)	77734	47537	45923	193640	120743	115650
Number of node (person)	33978000	22781795	17964004	36961664	24782302	19541453
Degree of node (person)	10183745	6828065	5384096	21228751	14233594	11223538

Data source:

http://news.bbc.co.uk/2/hi/technology/8552410.stm

https://www.census.gov

http://www.bea.gov/national/index.htm#gdp

http://www.people-press.org/2012/09/27/section-1-watching-reading-and-listening-to-the-news-3/

http://data.worldbank.org.cn/

http://www.undp.org/

http://www.itu.int/zh/Pages/default.aspx