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

Team Control Number

48182

Problem Chosen

D

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

2016 Mathematical Contest in Modeling (MCM) Summary Sheet
(Attach a copy of this page to each copy of your solution paper.)

The Evolution and Influence in the Media Network

Abstract

As an important project of network science, information dissemination is influenced by many factors. What's more, with the development of modern science and technology, the relationship between individuals and information has changed a lot. In this paper, our team try to establish models to measure the evolution and influence in society's information network.

Firstly, we build up a media network model: To begin with, we retrospect the development of the society's information network in five important historical periods. To describe the relationship between nodes when disseminating information, we build up an undirected and unweighted graph at first. Next, considering the inequality and non-repetition in intercommunication, we optimize the undirected and unweighted graph to directed and weighted graph, which are named as dynamic media network. Based on the model, we predict the network in 2050: The size of the network will continually become larger and more groups will occur in the network.

Then, we use the degree of the node to measure the probability of propagating the information, and give a definition and measurement of network's capacity. After the simulation, we draw a conclusion that the difference of the size of the information dissemination in different networks is related to the difference of the degree distribution. In addition, after validating the model by using information dissemination algorithm, we simulate the propagation of the information and filter the news from information by our justification.

Secondly, we build up a label propagation network model: In the first step, we define people's interests and opinions as labels and try to observe how label propagates in the media network. We use the label propagation algorithm to simulate the propagation and analyze the impact of it. To explain how community affect the labels' propagation, we make a compare between the community structure network and a null model. From the result, we come up with a conclusion that the community structure is beneficial to the aggregation of opinions.

Thirdly, we analyze the parameters in our paper. To verify the universality and robustness of our algorithms, we adjust numeric value of each parameters and compare it with the initial situation. After analyzing, we find that the parameters in our algorithms are suitable.

Contents

1 Introduction **1**
1.1 Problem analysis . . . 1
1.2 Literature review . . . 2

2 Media Network Model **2**
2.1 Topology structure of media network . . . 3
2.1.1 Defination of basic concept and model . . . 3
2.1.2 Sequential Evolution of network . . . 3
2.2 Dynamic model of intercommunication . . . 6
2.2.1 Inequality in intercommunication . . . 6
2.2.2 Non-repetition in communication . . . 7
2.2.3 Degree-based probabilistic model . . . 7
2.2.4 and capacity of the media network . . . 8
2.2.5 Prediction of communication networks' relationships . . . 8
2.3 Experiment . . . 8
2.3.1 Validation of the model . . . 8
2.3.2 The application to filter the news from information . . . 9
2.4 Evaluation . . . 10

3 Label propagation network model **10**
3.1 Label definition . . . 10
3.2 Lable propagation . . . 11
3.3 How Public Views Affect Information Propagation . . . 12
3.4 How Community Affect the Information Propagation . . . 13
3.4.1 The Introduction of Community . . . 13
3.4.2 Null Model . . . 13
3.4.3 The experiment of the label's propagation . . . 14

4 Sensitive analysis **14**

5 Conclusion **15**
5.1 Strengths . . . 16
5.2 Weaknesses . . . 16

1 Introduction

We live in a media-saturated world and rely on a variety of old and new media for information, entertainment, and connection. The beginnings of mass media and mass communication go back 560 years to the 'print revolution' that occurred in Europe in the fifteenth century. With the development of the technology and human world, mass media connects human and society closer. In this period, the information dissemination has changed gradually:

- **The Print Age**: Around 1450, the development of the printing press was the key technological advance that moved us from the manuscript era to the print era. In this era, people mainly acquired information by newspaper, magazines and books.
- **The Audiovisual Age**: Wireless telegraphy paved the way for radio and television broadcasts. In the 1920s, radios became a more common household item and in the 1970s, televisions came to most households.
- **The Internet and Digital Media Age**: The development of digital code, microprocessors, and fiber-optic cables were key technological advances that made the Internet and digital communication possible. In this new millennium, mobile phone and Internet stepped into our daily life, which played more and more important roles in spreading information.

Reviewing the main means of propagation in every age is beneficial for us to get an overall picture of society. For instance, in 1940, during the U.S. presidential election campaign, the majority of daily newspapers made wrong predictions of the result for the reason of the misunderstanding of the society's information network. Therefore, analyzing the Information circulation way is very important in each era, which is exactly what we want to discuss in this paper.

1.1 Problem analysis

Our team is assigned by ICM to measure the evolution and influence in society's information networks. We seek to understand the evolution of the methodology, purpose, and functionality of society's networks by taking a historical perspective.
Nowadays, information is spread quickly in the tech-connected communications network. On the one hand, it is due to the inherent value of the information itself, and on the other hand, it is due to the information finding its way to influential or central network nodes that accelerate its spread through social media.
Based on this background, to achieve our goal, we need to analyze the relationship between flow of information and inherent value of information based on consideration of 5 periods. The periods contain 1870s, 1920s, 1970s, 1990s and 2010s.
Our tasks:

- Build a society's information network model: Develop a model to explore the flow of the information and find what qualifiers as news. Then validate the model's reliability by using the data from the past, and evaluate the prediction capability by comparing the prediction of today's information communication situation with reality. After that, predict the communication networks' relationships and capacities around 2050.
- Build a label information network model: Develop a model to observe how public interest and opinion can be changed through information networks, and analyzing how the information itself and the network structure be used to spread information and influence public opinion.

1.2 Literature review

'Mass media' is a diversified collection of media technologies that reach a large audience via mass communication. The beginnings of mass media and mass communication go back 560 years to the 'print revolution' that occurred in Europe in the fifteenth century. It is only through technology that mass media can exist. These technologies takes place include a variety of outlets.

- Print mass media: It is the printing press and subsequent technological advances related to paper manufacturing and distribution that led to the establishment of print as the first mass medium. Print media transmit information via physical objects, such as books, comics, magazines, newspapers, or pamphlets. Despite the advent of many other forms of mass media, print is still important as a channel for information and as an industry. For example, in the United States, about 3.1 billion books, 1,400 daily newspapers, and 19,000 magazines are published a year.

- Sound Mass Media: Sound mass media transmit information electronically, via such media as film, radio, recorded music, or television. The origins of sound-based mass media can be traced primarily to the invention and spread of the telegraph. Even more recently, to adapt to changing technologies, radio stations have turned to online streaming and podcasts so their content can still make its way to computers and portable devices such as smartphones. Just as radio caught on quickly, however, so did television and movies.

- Visual Mass Media: It wasn't until about 40,000 years ago that we know our ancestors first began to explore visual media including drawings, paintings, and sculptures. Although we can see that visual media have long been a part of human history, they didn't constitute a mass medium until the late 1800s and early 1900s with the advent of motion pictures and television.

- The Digital Media: Digital media, which began in 1990 and continues today, comprises both Internet and mobile mass communication. Internet media comprise such services as email, social media sites, websites, and Internet-based radio and television. Many other mass media outlets have an additional presence on the web. In short, digital media read, write, and store data (text, images, sound, and video) using numerical code, which revolutionized media more quickly than ever before.

Each form of mass media affected society in important ways. However, we all have to admit that all these media, even print, had to adapt to the advent of the digital age. Copyright violations and pirating have become a problem when old media content was digitized, but makes information more easily reproducible and sharable.

2 Media Network Model

In this section, we firstly put forward a network model for describing the relationship among elements of media, such as newspaper, radio, people etc. Then, we infer the evolution of network from snapshots in five representive periods. Secondly, we build a interaction model to describe the process of information disseminating. After that, we use our model to simulate the information flow in some specific situations and define the standards of filtering news from information. Finally, we compare our model with the reality to evaluate its imitative effect.

2.1 Topology structure of media network

2.1.1 Defination of basic concept and model

- **Node**: The nodes in the network model refer to the information media in the society, which contain individuals and all the other media such as newspaper, radio, webpage and so on. Every node has a different effect on the information's communication and circulation to some extent.

- **Edge**: Edge is a bridge that two points interact with each other. Focused on the society's information network that regards media as points, the most direct connection is the information's emanation and receive. Consequently, the edge should be directed and has different weight based on the nodes' character, and the weight should represent the probability of spreading information.

- **Random Network**

  The ER model[3] of a random network starts with N nodes and connects each pair of nodes with probability p. It creates a graph with approximately $pN(N-1)/2$ randomly placed links.

- **Small-world Network**

  A small-world network[6] is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps. Specifically, a small-world network is defined to be a network where the typical distance L between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes N in the network.

- **Scale-Free Network**

  A scale-free network[2] is a network whose degree distribution follows a power law, at least asymptotically. That is, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as $P(k) \sim k^{-\gamma}(2 < \gamma < 3)$.

  The most notable characteristic in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average, which means most nodes only connect a few nodes in the network, while few nodes connect a huge number of nodes.

- **Gini coefficient**

  The Gini coefficient[4] is a measure of statistical dispersion intended to represent the income distribution of a nation's residents, and is the most commonly used measure of inequality. The specific formula is shown below:

$$G = \frac{\sum\limits_i \sum_j |x_i - x_j|}{2\sum_i \sum_j x_i} \tag{1}$$

  In the paper, we use this index to evaluate the difference level of a model.

2.1.2 Sequential Evolution of network

As time progressed and technology developed, the relationships in the media network have been evolving with time. In this part we aim at retrospecting the development of the mass media at the macroscopic level and introduce the relationship in the media network during different historical periods.

In addition, to observe the information flow in the network more intuitively and concisely, we assume that there will be no change in quantity of nodes and edges in a time unit. The quantitative changes will only happen when time flows between different periods.

To make our time clue more specific, we list several important historical nodes: 1870s, 1920s, 1970s, 1990s and 2010s.

- **The print age**: In the 1870s, when newspapers were delivered by trains and stories were passed by telegraph, the communications among people were not very widely. People only exchanged their information in their living and working areas. We quantify the possibility of existence of the relationship between two people as p. There hardly or no exists information exchange when two individuals were geographic far away from each other, which caused p was fairly small.

  In the meantime, because of the less developed transportation system, newspapers and magazines can only radiate at a limited extent. We quantify the possibility of existence of the relationship between people and newspapers as q_1. It is easy to be accepted that q_1 was also small at that time. However, considering the communication between different newspapers is highly necessary, we believe that the relationship among those mass media is tight and close.

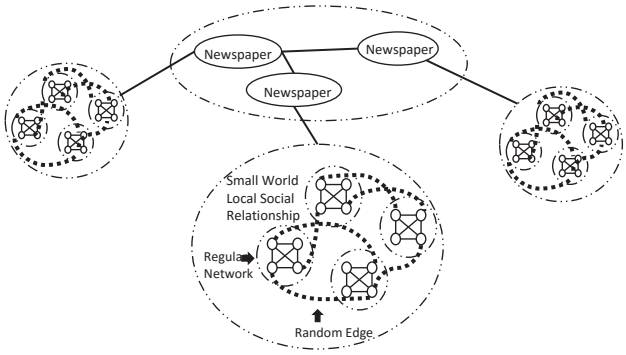

Figure 1: The Network Structure of the Print age

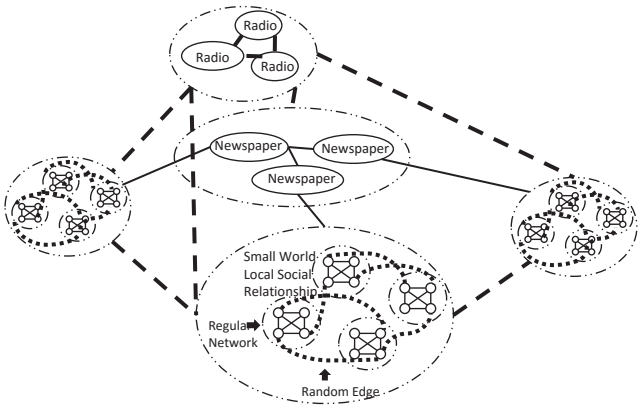

Figure 2: The Network Structure of the Radio Age

- **The radio age**: In the 1920s, radio came to our life. Transportation was more convenient than before, which help people contact with each other more frequently. The probabilities of p became larger. The advantage of the radio is that it can radiate a wider extent

than newspaper or magazine, therefore, people were more likely to get information from magazine or newspaper. We take an assumption that the possibility of existence of the relationship between people and radio is named as q_2, which is larger than q_1.

- **The television age**: In the 1970s, watching TV at home was seemed like a common habit for most of people. The diversification of transportation and other factors helped people gain closer ties to each other. The probabilities of p were larger. Obviously, TV has a more influential effect, so we define it as q_3, which is larger than radios and newspapers

  In the other hand, the rise of new things will cause the demise of the old things. Some statistics indicated that the prosperity of TV lead to the decline of radio and newspaper.

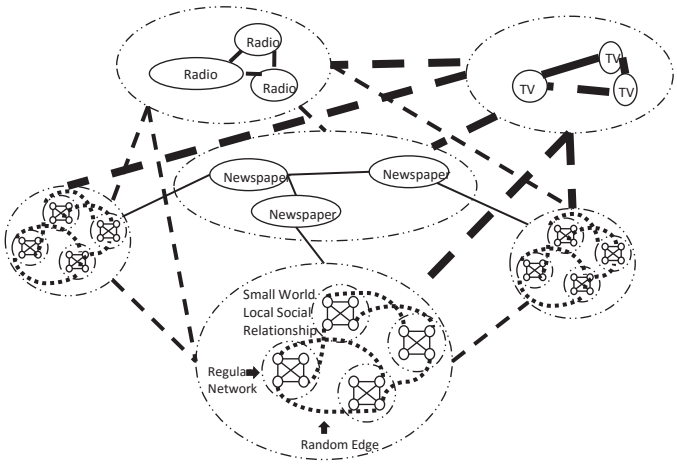

Figure 3: The Structure of the Television Age

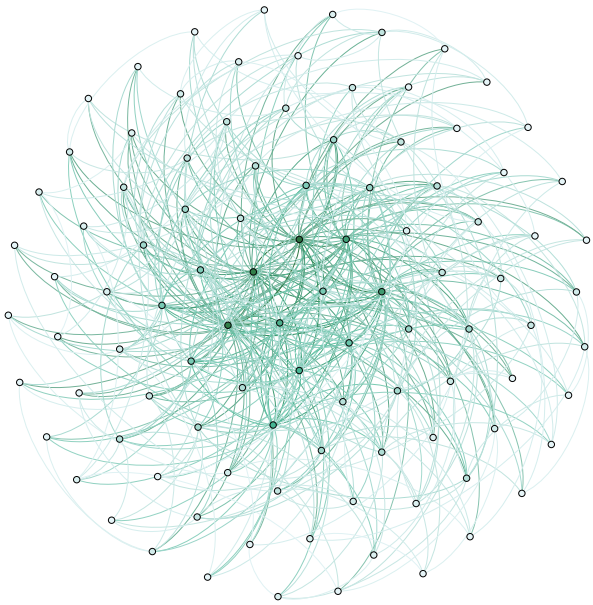

Figure 4: The Network Structure of the Internet Age

- **The internet age**: In the 1990s, a markable event took place. The impact of the Internet is so immense that the distance between people has never been so close. In a way, the internet makes the communication among people much easier, rather than add a new

mean of communication in this relationship network. Instead of paper being the medium for books, radio waves being the medium for sound broadcasting, and cables being the medium for cable television, a person can now read a book, listen to the radio, and access many cable television shows on the Internet. That is to say, p_1, p_2 and p_3 became larger and larger. From a scientific standpoint, the network became a small world.

- **The mobile age**: Since 2010, world has gotten into the digital era. Whereas media used to be defined by their delivery systems, digital media are all similarly constructed with digital, binary code made up of ones and zeros. In short, digital media read, write, and store data (text, images, sound, and video) using numerical code, which revolutionized media more quickly than ever before. Meanwhile, individuals are exposed to vast quantities of information and connected with different media.

  It is necessary to reconsider the structure of the network. Previous research has shown that the recent network is a scale-free model, which we usually call as BA model. The major property of a BA model is when a new node wants to join the network, it will be more likely to connect with the nodes with larger degree.

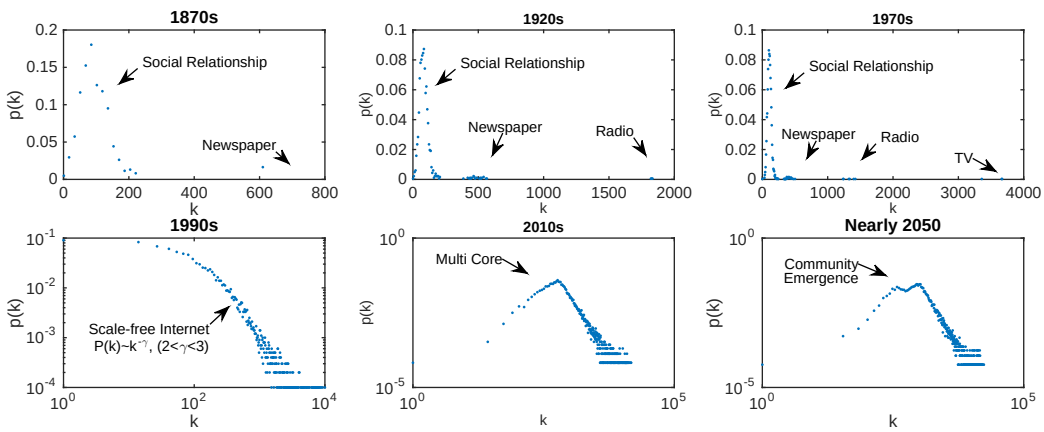

Figure 5: Information dissemination algorithm

2.2 Dynamic model of intercommunication

In this part, we build up an dynamic information dissemination model based on the static media network we talked about in the above. To describe the dynamic process of intercommunication more directly, we decide to expand the undirected and unweighted graph to the directed and weighted graph.

The decision we make is mainly based on the following considerations.

2.2.1 Inequality in intercommunication

In a media network of the real word, a node has different probabilities of receiving and sending a message.

Suppose that there are two nodes: BBC and one of its audience. The relationship between BBC and its audience is a two-way process. Not only BBC can send its information to the audience by TV, but also the audience can share his information with BBC by calling the hotline. Obviously, however, this mutual relationship is not equal. As we can see, BBC is more easily than the audience to disseminate the information to the other.

Because of the information asymmetry, or to be more specific, the inequality in intercommunication, it is more rational to expand the undirected relationship between two nodes to the directed. This will help us to analyze the information spreading respectively.

2.2.2 Non-repetition in communication

Different from virus propagation model that the same virus can be spread many times as long as they have a contact, in a media network, a piece of information will not to be spread by a same access for more than one time.

In other word, there are only two situations in the propagation of information through an edge: send or not. A receiver can choose whether to accept the information from the access with the sender. No matter what the result is, he wont receive the same information from this sender again, however, he still can receive this piece of information from other media.

The reason for that is timeliness of the information. So long as the virus exists, it can transmit in the network, but for the outdated information, little can be propagated by the media. Owing to the mainly purpose of the media is making profits, and considering only updated information can attract lots of audiences, the same information will be eliminated. Similarly, people prefer to discuss something new, rather than something outdated. In a word, this definition is consistent to the reality and necessary.

2.2.3 Degree-based probabilistic model

According to the instance above, we can see that the possibility of sending out a message from BBC and the individual is unequal. To quantify this difference, then we expand the unweighted graph to the weighted graph.

There are four situations when disseminating information: Nodes of large degree send information to the nodes of large degree; Nodes of large degree send information to the nodes of small degree; Nodes of small degree send information to the nodes of large degree; Nodes of small degree send information to the nodes of small degree.

Based on these four situations, we will give a method to calculate the edges weight.On account of that different nodes will have different probabilities of sending the information to other nodes, we add the weight w to the edge in the relationship model, which present the probability of information can spread through the edge. In addition, we regard the degree of the nodes as the most important factors in affecting the probability, so we measure the probability based on the degree of every node. Besides, media is different from each other in the speed of spreading the information, so we also draw into a parameter λ to describe it. In conclusion, the weight between two nodes i and j is shown as follows:

$$w_{ij} = \begin{cases} \frac{1}{k} - \frac{1}{2k(1+deg_i-deg_j)} & \text{, if } deg_i > deg_j \\ \frac{1}{2k(1+deg_j-deg_i)} & \text{, otherwise} \end{cases} \tag{2}$$

where deg_i and deg_j are the degree of node i and node j, and k is a parameter.

To summarize our key points, we still put the same example. BBC, as a famous and influential media, is more likely to send the information to a large number of people, and the people who have access to contact it, may have more probability to receive the information. While an ordinary person in the information network model could only send his information to a few people or public media, which he had a communication at that time. What's more, BBC can send the information to all the people listening to it at the same time, but for a person, it is difficult to send the information to a lot of people in such a short time. In a word, we display the divergence of spreading information by weight in the edges.

2.2.4 and capacity of the media network

We regard the propagation range γ as the percentage of the nodes which has received the information in the network.

$$\gamma = \frac{\sigma_r}{\sigma} \tag{3}$$

where σ_r represents the number of the nodes that receiv the information, and the σ is the total number of the nodes.

The factors which can influence the capacity of a network concludes the number of nodes and edges, are the amount of the information which can be propagated at the same time and the uniformity of the degree distribution. When there are more nodes and edges in the network, the capacity will expand through multiple add. However, keeping the number of nodes and edges unchanged, capacity will still be different, which is resulted from the structure of the network. It is easy to prove that a star network and a regular network with the same number of nodes and edges have different capacity.

Model	#nodes	#edges	Gini Index	Capacity
Star Network	300	300	0.5	1
Random Network	300	1500	0.23	3.67
Regular Network	300	1500	0	5.44

Table 1: Capacity of Different Network Model

2.2.5 Prediction of communication networks' relationships

After evaluating the accuracy of our model, we can make a prediction of what the network would be like around the 2050.

As such development tendency, firstly, the size of the network will be huger and the relationships between nodes will be more complicated. Secondly, a large amount of groups of nodes will emerge. In 2050, the problem many people may face is that there are so many sources of information, though they would want to keep track of, the will have become buried. Because of that, people will be more incline to contact to the nodes they are interested in and that is why there would be many groups in the network. On the other hand, people will have more means of spreading information. That is to say, the influence of an individual will become larger.

2.3 Experiment

2.3.1 Validation of the model

In this part, we want to validate the correction of the model. Firstly, we generate some information randomly, and let the information spread through the model in different periods. Then, we compare the experimental data with real data, and draw a conclusion about the feasibility of our model.

Before we start the experiment, we firstly suppose that the information we will generate later has equivalent influence, for the reason that the value of a piece of information has less effect on the spreading.

In the experimental process, we utilize the following algorithm and methods:

Step 1: Generate the information and let different media propagate it randomly. Define a node set and add all the nodes that are the source of the information's propagation to it.

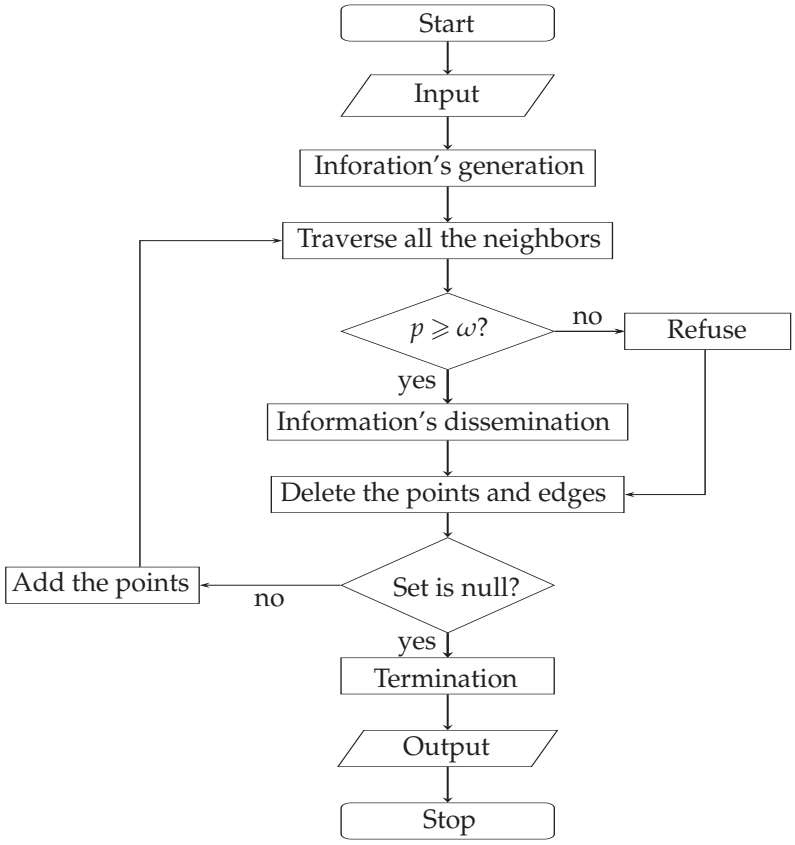

Figure 6: Information dissemination algorithm

Step 2: Every piece of information spread from one node, and firstly traverse the neighbor of the node to judge whether the probability is larger than the weight of the edge. If larger, propagate the information to next nodes. After the traversal is finished, delete the source of the information and the edges that has spread the information.

Step 3: Judge whether the set is null. If so, end the algorithm. Otherwise, add all the nodes received the information to the set, and return to 2 to traverse the neighbors of the nodes in the set.

2.3.2 The application to filter the news from information

News is a subset of information, which has a more specific definition than information. Before we filter the news from information, it still remains some points to be demonstrated:

1. According to the definition of news in the literature sense: noteworthy information about recent events, we hold a view that a piece of information can be called as news when its range of propagation is wide enough. Briefly, news has a wide range of propagation.

2. A piece of information may generate from just one node. We suppose that the repetition wont occur in the network, and different information generate from different nodes in the network. This assumption will help us justify the difference between information. That is to say, what makes them different are not only depended on the audience but also on their sources.

3. When filter the information, we don't consider the internal value of the information itself in the model. People have different hobbies, so they intend to be interested in different information. Everyone has its own justification on the information, therefore individual

variation and imperfection should be allowed. In this case, it is meaningless to discuss the internal value of the information and regard it as a standard of news.

We will process the information as following steps: First, collect the data about the spread range of each piece of information. The spread range of the information can be counted as the percentage of the receivers in the network. Second, list the spread range of each piece of information in descending order. The top 10 percent of this list will be the news under our justification.

Detailed processes are as follows:

Network	#nodes	#edges	Gini Index of Degree
1870s	609	2742	0.43
1920s	1829	21059	0.29
1970s	3668	41784	0.20
1990s	9920	199790	0.54
2010s	14841	342627	0.43
2050s	17247	398489	0.37

Table 2: Statistical Description of Media Network in Different Period

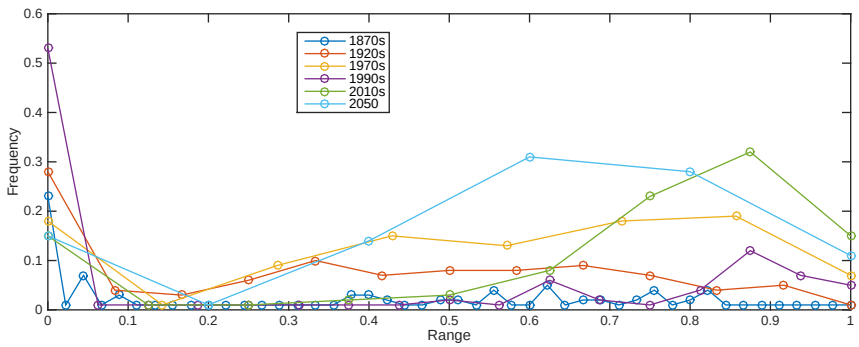

Figure 7: Range Frequency of Different Networks

2.4 Evaluation

In order to evaluate our model, we build up an artificial dataset based on scale-free network, and compare its range with the real the dataset from facebook[7].

Network	#nodes	#edges	Gini Index of Degree
BA Model	63000	800000	0.61
Facebook	63731	817035	0.64

Table 3: Statistical Description of Media Network in Different Period[1]

3 Label propagation network model

In this section, we focus on the phenomenon that public interest and opinion can be changed through information networks. Our aim is optimizing the media network model and giving

a rational explanation to the phenomenon. So firstly we put forward the definition of labels attached to every node, which are on behalf of the opinions of the nodes. Secondly, we further research the labels' propagation in the network model and make an experiment. After that, we discover a phenomenon of aggregation, and attempt to find how the community can affect the label. Finally, we draw the conclusions by comparing with null model.

3.1 Label definition

All the nodes in the communication model, no matter individual or mass media such as newspaper, are the information transmission terminators. Besides of spreading the information completely, they also add their personal opinions into the information. What's more, they will also interpret the information according to their own comprehensions and positions. For instance, newspaper will translate some complex information for audiences in some situations. Unconsciously, individuals' opinions, interests and even lifestyle would be influenced or changed to some extent by such special information.
For example, in a president campaign, John thought that the republican presidential contender performed better than the democratic at first, but lots of his friends as well as some influential media had the opposite opinions of his. Affected by their reasons and views of those two candidates, he maybe change his opinion at last.
For the sake of depicting people's opinions or interests mathematically, we attach a label to every node in the network. To be more specific, when facing a situation, different people hold different views on it. We code these different views as 1, 2, 3 and so on. In this way, handling contents becomes more easily. Label has diverse meanings. That is to say, each node's label represents its own opinion.

3.2 Lable propagation

In a media network, individual or social media often send messages or information with their subjective perceptions. This kind of information with specific label can be propagated through any accesses, and the receivers would be influenced by this special information.

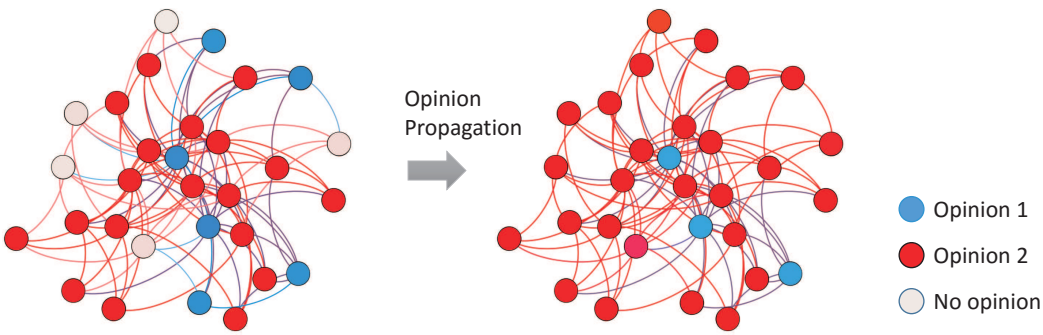

Figure 8: Opinion propagation

In this part, we try to use the theories and concepts of information influence on networks to model how the label can be propagated and make an influence on other nodes in the network. We represent out media network using a $N \times N$ affinity matrix W, where the value of w_{ij} means the strength of link from j to i. People' initial view label matrix in this network is denoted as T, whose (i, k)-th element is 1 if node i belong to category $k\,(k = 1, 2)$. All zeros rows means we do not know the nodes' opinions. Furthermore, we define the label propagation score matrix F, where $0 \leq f_{ik} \leq 1$ represents the score of student i in label category k after label propagation

procedure. We can formulate our label propagation procedure as [8]:

$$f_i^{t+1} = \eta \sum_{j \in \mathcal{N}(x_i)} w_{ij} f_j^t + (1-\eta) t_i. \tag{4}$$

Where $0 < \eta < 1$ is the parameter of label information that node i receives from its neighbors. After we concatenate each vector to matrix, we can update our iteration equation as

$$F^{t+1} = \eta W F^t + (1-\eta) T \tag{5}$$

Then, we are supposed to demonstrate that F will converge to $F^* = (1-\eta)(1-\eta W)^{-1} T$. Based on the initial condition that $F^0 = T$, we derive

$$F^t = \eta W^{t-1} T + (1-\eta) \sum_{i=0}^{t-1} (\eta W)^i T \tag{6}$$

Because of $w_{ij} \geqslant 0$ and $\sum_{j=1}^{N} w_{ij} = 1$, from the theorem of Perron-Frobenius [5], we know that the spectral radius of W, which is $\rho(W) \leqslant 1$. What's more, because of $0 < \eta < 1$

$$\lim_{t \to \infty} (\eta W)^{t-1} = 0 \tag{7}$$

$$\lim_{t \to \infty} \sum_{i=0}^{t-1} (\eta W)^i = (D - \eta W)^{-1} \tag{8}$$

where D is the degree matrix of this network, i.e., $D_{ii} = \sum_{j=1}^{N} w_{ij}$. Then, the final prediction matrix can be calculated as

$$F = (1-\eta)(D - \eta W)^{-1} T \tag{9}$$

Thus, the value f_{ik} means the final prediction score of node i in label category k. The label of node i can be determined as $y_i = \arg\max_k f_{ik}$. **Step 1**: Attach a label representing the opinions

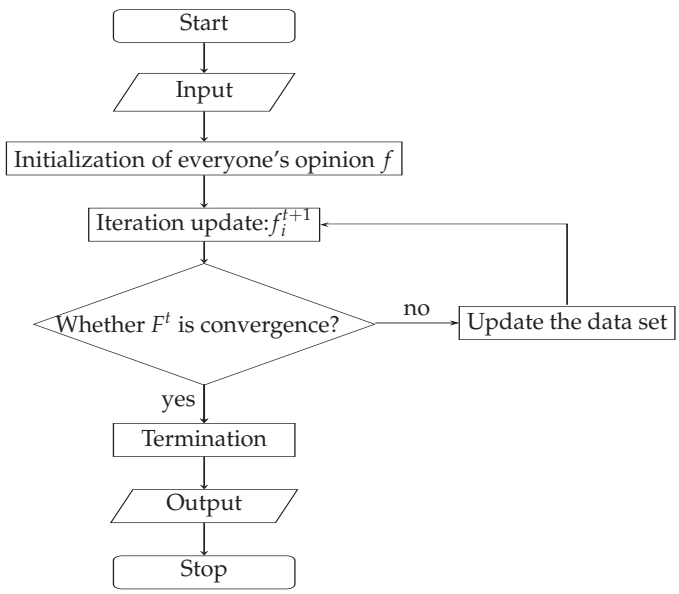

Figure 9: The Algorithm Flow of Label Propagation

to every node. All the labels form a matrix F_t. For instance, someone agree with an event, while other people disagree with it, and others even don't know about this thing, so we denote

the three conditions respectively as: $[1\,0]$, $[0\,1]$ and $[0\,0]$. All the matrixes $f_{1\times 2}$ combine to the matrix $F_{n\times 2}$

Step 2: Use the iterative formula of F_t to update the matrix. According to the formula, every node's label will be affected by the others and make changes to some extent. So, this process can simulate the spread of labels.

Step 3: Repeat the iteration update until F_t is convergence and then draw the conclusion.

3.3 How Public Views Affect Information Propagation

Based on the label propagation algorithm we described above, in this part, we are supposed to discovery the influence of public views to ones' opinion. Thus, we compare the rate of positive views before and after label propagation. The result are shown in Table. 4.

Initial Rate	1870	1920	1970	1990	2010
0.1	0.024	0.022	0.008	0.012	0.015
0.2	0.023	0.033	0.068	0.056	0.047
0.3	0.239	0.261	0.132	0.115	0.198
0.4	0.309	0.520	0.183	0.132	0.234
0.5	0.600	0.561	0.600	0.498	0.532
0.6	0.636	0.722	0.880	0.789	0.796
0.7	0.758	0.911	0.981	0.975	0.945
0.8	0.946	0.998	0.989	0.992	0.983
0.9	0.985	0.999	0.999	0.997	0.996

Table 4: Comparison the Influence of Different Community Model

Therefore, it is obviously that majority's interests have impact on others.

3.4 How Community Affect the Information Propagation

In this part, we want to reveal the cause of how communities affect labels propagation and propose our views based on the simulate results.

3.4.1 The Introduction of Community

To determine how peoples opinions and interests, strength of the information network in metropolitan neighborhood, state, and occasionally a region could be used to influence public opinions and spread information, we first introduce the community structure into our model.

Retrospecting the development of the social media network, we find a phenomenon that the nodes will congregate with their neighbors in flocks. Considering the enormous quantity of the information nowadays, people can obtain any information they are interested. Meanwhile, recommendation is more precise and popular, so people tend to aggregate with others who have the same opinions and interests. We name those different flocks as communities.

In the study of complex networks, a network is said to have community structure if the nodes of the network can be easily grouped into sets of nodes such that each set of nodes is densely connected internally. This implies that the network divides naturally into groups of nodes with dense connections internally and sparser connections between groups. In the further research, we find out that many real networks have obvious community structure.

We try to find how a community can affect its members opinions, interests and other attributes, which are called as labels. If so, we want to do a further research of how a community affect its members.

3.4.2 Null Model

In mathematics, in the study of statistical properties of graphs, the null model is a graph which matches one specific graph in some of its structural features, but which is otherwise taken to be an instance of a random graph. The null model is used as a term of comparison, to verify whether the graph in question displays some features, such as community structure, or not.

In our experiment, we make use of the characters of the null model in which all links were randomly rewired. We maintain the number of the nodes and the edges, and research the character of the label propagation network model through random rewiring algorithm.

3.4.3 The experiment of the label's propagation

Many real-world networks have an important property of community structure, in which there are stronger connections than those between the clusters. In this experiment, we want to find out whether the community structure has an impact on the labels propagation.

We use the null model to compare with a community structure model. This null model has the same amount of nodes and edges with the community structure, but the difference is that the edges in the null model are connected randomly. In a word, the null model is a network with no community structure. To measure the impact, we put forward an evaluation metric first.

Considering a pair of nodes in the network, there are two situations: Two nodes have the same label or have the different label. We count the front situation as 1 and the later situation as 0. After traversing all the pairs of nodes in the network, we can get a total amount. We regard this amount as a metric to evaluate the labels propagation in the network.

By comparing the metric of the null model and the community structure model, we can clearly see whether the community structure has an impact on the propagation of the label. Furthermore, we construct the null model in the following three cases:

- **Region**: We construct the null model of region by reconnecting all edges in a community of our network.
- **Country**: The null model of country is constructed by reconnecting all edges in some regions of our network
- **World**: We construct the null model of world by reconnecting all edges of our media network.

The result of these three cases is shown in Tab 5

Model	Metric
Real	0.902
Random Region	0.857
Random Country	0.821
Random World	0.770

Table 5: Comparison the Influence of Different Community Model

As the testing results show: The community structure does have an effect on the propagation of labels.

With the network becoming more and more enormous, a node tends to contact with only a part of the nodes which have some similarities with it and is more easily to be influenced by these nodes.

4 Sensitive analysis

Our model defines some random numbers and parameters determined by experience. This may lead to uncertainty effects and some errors, so we fluctuate the index within limits to see whether the result will be changed a lot.

Uncertain effects produced by stochastic simulation

To better describe the information dissemination, we define the index k to amend the formula and let k equals to be 2, 3, 4, 5, 6. From the graph below, we notice that with the variation of k, the range of the informations dissemination doesnt change obviously. That means the

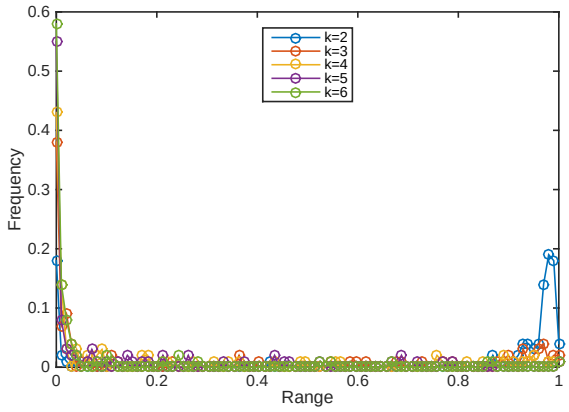

Figure 10: Sensitive Analysis of Parameter k

value of k is stable, and it is proper to set this value as parameter.

In label propagation algorithm, we put forward a parameter η to describe the impact of the label propagation. We suppose that an individual agrees or disagrees with an opinion, which we quantify them as 1 and -1. By changing the value of , we draw a curve to observe the change of the ratio of nodes which hold the view 1 and -1 is small or not. According to that the real network has homogeneity, we choose the lowest point as the ideal point.

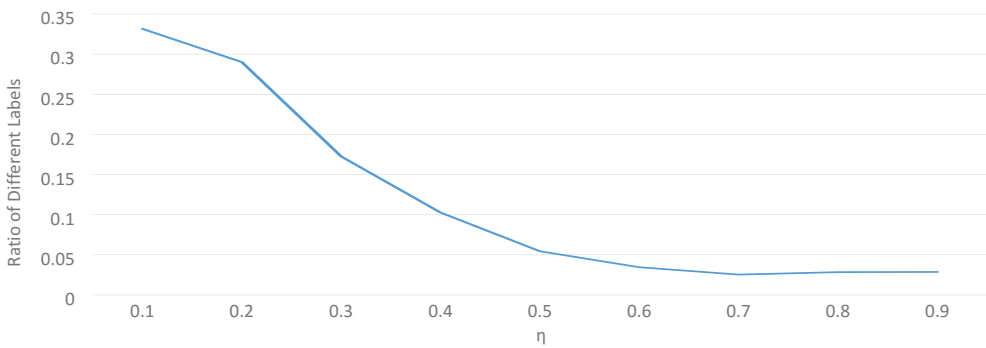

Figure 11: Sensitive Analysis of Parameter η

5 Conclusion

In todays tech-connected communications network, information spreads quickly and extensively. We live in a media-saturated world and rely on this huge network for information, entertainment, and connection. We all have a need for the information to satisfy curiosity, reduce uncertainty, and better understand how we fit into the world.

Our team is assigned to analyze the relationship between speed or flow of information and inherent value of information. After completing the tasks, we draw three conclusions to summary our main works in this paper.

1. The difference of the size of the information dissemination in different networks is related to the difference of the degree distribution. We regard the propagation range as the percentage of the nodes which has received the information in the network. Generating the same amount of the information in several networks with different degree distribution, we find that the degree of deviation of the information disseminates larger than that in the network with homogeneous degree distribution. This conclusion can help us to explain why sometimes a piece of information is harder to be propagated to a node than before.

2. The factors which can influence the capacity of a network concludes the number of nodes and edges, are the amount of the information which can be propagated at the same time and the uniformity of the degree distribution. When there are more nodes and edges in the network, the capacity will expand through multiple add. However, keeping the number of nodes and edges unchanged, capacity will still be different, which is resulted from the structure of the network. It is easy to prove that a star network and a regular network with the same number of nodes and edges have different capacity.

   To explain our conclusion more specifically, we display the capacity by the number of information spreading in the network, and define the measurable indicator as iterations. Finally, we find out that the greater the difference, the less the capacity, and this result is accord to different network models.

3. The community structure is beneficial to the aggregation of opinions. Whats more, the community structure can influence its members opinions, interests or other attributes. As we have discussed before, a node is more likely to receive the message which is sent in the same community, rather than from different communities. Nodes in a community will influence and be influenced by their neighbors, which may have a potential impact on their characteristics.

5.1 Strengths

1. We build a media model to research the information propagation in the network. Other than the epidemic model, our model consider the information's timeliness. For one thing, we can study the spread of the information. For another, we fully consider the internal characteristic of the information. Therefore, our model is close with the real situation.

2. We establish a label propagation model to take every node's opinions into consideration. By this way, we can easily observe the process of the propagation of people's opinions and interests through the media network and describe their changes immediately. In addition, we further observe the community's effect in the label propagation network and find some valuable attributes to explain the phenomenon.

3. The details in algorithms, such as the algorithms of the label's propagation, are in accord with real conditions, and the parameters used in the paper is reasonable. The precision of the established model is well, the fitting effect was perfect and extrapolation predictive value was reliable. What's more, the results of the model are accord with the subjective cognition.

5.2 Weaknesses

1. Because of the time limits, we only build the models for local spreading rather than the whole world.

2. Because of the irritation of the algorithm, the time complexity is high and it is time consuming.

References

[1] Facebook friendships network dataset – KONECT, May 2015.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[3] P ERDdS and A R&WI. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.

[4] CW Gini. Variability and mutability, contribution to the study of statistical distribution and relaitons. *Studi Economico-Giuricici della R*, 1912.

[5] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[6] Mark EJ Newman and Duncan J Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.

[7] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. On the evolution of user interaction in Facebook. In *Proc. Workshop on Online Social Networks*, pages 37–42, 2009.

[8] Fei Wang and Changshui Zhang. Label propagation through linear neighborhoods. *Knowledge and Data Engineering, IEEE Transactions on*, 20(1):55–67, 2008.