Information Diffusion in Complex Network

May 27, 2018

Contents

1 Introduction

1.1 Background

Nowadays, we are encompassed by information network in the Information Age. Information networks acting as links interacts in seemingly complex ways, which facilitates the spread of information. From newspapers to smart phones, from the wired networks to the wireless networks, the flow of information has been easier and more widespread. It is generally accepted that there exists a strong relationship between the flow of information and its value throughout the history. In this situation, it is very meaningful to learn the information diffusion in complex network.

1.2 Discussion

The issue of information flow has been studied extensively. For example, Granovetter proposed linear threshold model to analyze the effects of network structure[1]. Daron Acemoglu et al extended the model to better capture the notion of path dependence[2]. In "Viral Spread Models for Multimedia Content Propagation in Social Networks", Dinuka Aminda Soysa put forward ICM (Independent Cascade Model) to predict the future access pattern of multi-media content in social network[3].

The most classic model for this issue must be the **SIR Model** (Susceptible Infected Recovered Model), which was built by Kermack and Mckendrick in 1927[4]. In this classic model, people are divided into three groups: **Susceptible**, **Infective** and **Removal**. In our model, we would divided people into similar groups: **S** (uninformed people), **I** (informed people who would spread the information), **R** (informed people who would not spread the information). Then we can derive transform rates between groups with some statistics.

The importance of SIR Model is to uncover the interpersonal diffusion rate, which describes how rapidly the information propagates in complex network. Therefore, we choose the SIR Model to analyze the flow of information. The infectious recovery SIR Model of online social networks is proposed by John Cannarella, who draw analogy to the dynamics that govern the spread of infectious disease[5]. Stauffer and Sahimi studied the mechanism of the spread of extreme opinions using the epidemic model with an exposed period and simulated the diffusion process in the scale-free network.

Lastly, the Bass Diffusion Model (BDM), first introduced in 1969, views diffusion as occurring through a combination of innovation and limitation, describing the external influence and internal influence to a population[6]. It can be used to understand diffusion of products and innovations. Hence, we choose the Bass model to predict the communication networks' relationships and capacities in the future.

1.3 My Work

In this project, I am supposed to build some models to explore the diffusion process of information and predict certain relationships in complex network. Our model should not only consider the trend of information flow in complex networks, but also how public interest and opinion can be changed through information networks and other factors.

Firstly, we modify the original SIR Model and apply it into the flow of information, as well as the interaction between diffusion speed and information value. We collect some data to test the model and results prove to be good. In this way, we build a simple and practicable model to simulate the information diffusion process in complex network.

Next, we utilize the Bass Diffusion Model to predict the information diffusion situation in daily life, and results turn out to match perfectly with the reality. And then we predict the number of users of different media in near future. This would give us a direct view about how complex social network changes.

Finally, some factors are listed since I think they may influence information diffusion, including information value, people's initial opinion, form and source of the message and the structure of the complex network. And we make sensitivity analysis for 3 parameters: the time constant, the innovation and imitation coefficient in Bass Diffusion Model.

2 Assumptions

Our model makes the following assumptions:

1. Each individual in the complex network is regarded equally, which means that people have the same probability to receive a certain piece of information and be "infected" by it.
2. The information value will not change during the diffusion process.
3. People who have received certain information can be divided into two groups: those would spread the information and those do not spread the information.
4. The transmission rate that uninformed people turn into informed group will not change during the diffusion process.

3 Nomenclature

In this paper, symbols and definitions used in SIR Model and Bass Diffusion Model are listed in Table 1 and Table 2 respectively. The detailed explanation of symbols can be found in the following sections.

4 Information Diffusion with SIR Model

In order to construct a better model to demonstrate information diffusion process, we first consider how people spread information in reality. Some people learn information through many chan-

Table 1: Symbol Description for SIR Model

Symbol	Definition
$S(t)$	Percentage of uninformed people
$I(t)$	Percentage of informed people who would spread the information
$R(t)$	Percentage of informed people who would not spread the information
$F(t)$	Percentage of all informed people
λ	Transmission rate from S to I
γ	Transmission rate from S to R
μ	Transmission rate from I to R
ω	Rate of information spreading from other media
σ_{ij}	Information flow rate between district i and j

Table 2: Symbol Description for the Bass Diffusion Model

Symbol	Definition
$F(t)$	Percentage of adopters
$S(t)$	Cumulative number of adopters
m	Eventual number of adopters
$G_i(t)$	The user percentage of the ith generation technology
δ_{ij}	Transition factor from the ith generation to the jth
τ_i	Introduction time of the ith generation

nels and we divide them into two groups: those who would like to spread information and those who won't. The former will tell the uninformed at certain rate, which will increase the number of informed people. Due to some factors, those who would like to spread information may be unwilling to spread information at a certain rate.

4.1 Basic Model

In basic model, we make assumption that contemporary social media has not appeared. Information are transmitted by word of mouth and writing material.

As I have said before, people are divided into three groups:

- **S**: Uninformed people, who don't know the information at time t;
- **I**: Informed people, who would like to spread the information;
- **R**: Informed people, who don't like to spread the information.

I have already defined λ, γ, μ as transmission rates, which represent population flow from S to I, from S to R and from I to R respectively.Since different people have various attitudes toward received information, we can divide people into n categories and give them different parameters. Since people would live in different areas, we set λ_i, γ_i and μ_i denote corresponding parameters in the district i. And σ_{ij} represents the rate of information flow between district i and district j.

The flow chat of our basic SIR Model is shown in Figure 1.

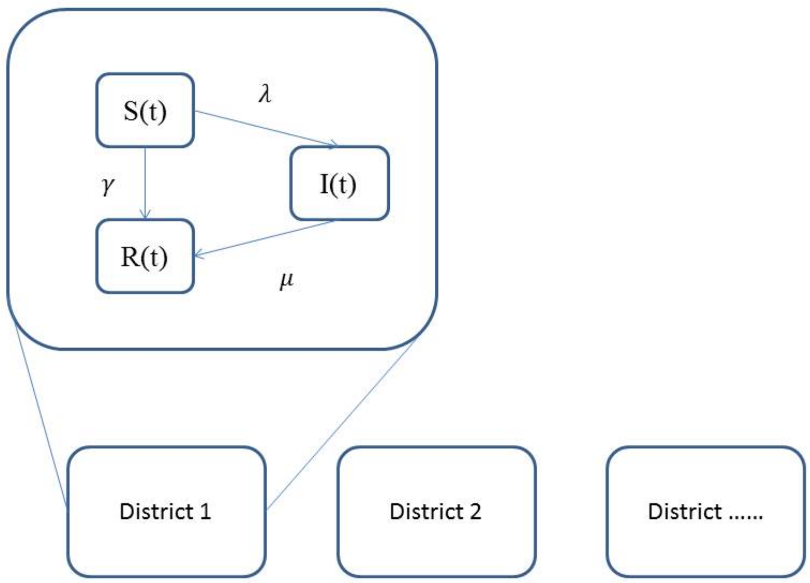

Figure 1: Flow chart of SIR Model

Based on the SIR Network Model we get the following set of ordinary differential equations:

$$\frac{\mathrm{d}S_i(t)}{\mathrm{d}t} = -\lambda_i S_i(t) I_i(t) - \gamma_i S_i(t) I_i(t) - (\lambda_i + \gamma_i) \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) \tag{1}$$

$$\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} = \lambda_i S_i(t)I_i(t) - \mu_i I_i(t) + \lambda_i \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) \tag{2}$$

$$\frac{\mathrm{d}R_i(t)}{\mathrm{d}t} = \mu_i I_i(t) + \gamma_i S_i(t)I_i(t) + \gamma_i \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) \tag{3}$$

$$F_i(t) = I_i(t) + R_i(t) \tag{4}$$

4.2 Traditional Information Diffusion Model

The appearance of traditional media such as newspaper, television and telephone greatly facilitate information diffusion process, making the diffusion network much more complicated.

We should also focus on the fact that the timeliness is quite important for information value. People are prone to show more interest in updated news. Also, for the same information, the spread speed in each diffusion channels vary a lot. Some channels, such as online live and social application, are the first to report and spread the information, while traditional channel such as newspaper and television will be relatively slow.

Some symbols are defined for our refined model:

- ω_n: the rate of information diffusion through newspaper and other paper media;
- ω_t: the rate of information diffusion t; hrough television, telephone and so on;
- ω_i: the rate of information diffusion through Internet;
- $f(t)$: the timeliness of certain information at time t.

For function $f(t)$, we use the exponential decay function to represent the timeliness of certain information.

$$f(t) = e^{-k(t-T)} \tag{5}$$

In this function, k is a decay coefficient. Larger the k is, faster the information fades its value. T represents the time the information diffusion happen through certain media.

Then we would refine our basic model by taking traditional media, such as newspaper, television, telephone into consideration. With the help of these media, location limitation is broken. People in different districts can learn about certain information miles away through these channels. We build the refined SIR Model in Figure 2.

In Figure 2, we set traditional media as a higher "district", which can connect to other locations directly in a faster pace. We assume in this district, the rates of percentage of people who are willing to spread information and people who don't like to are also represented by λ and γ. We use $f_n(t)$ and $f_t(t)$ to represent the timeliness of newspaper and the timeliness of telephone an television.

The revised SIR ordinary differential equations are listed as follows:

$$\frac{\mathrm{d}S_i(t)}{\mathrm{d}t} = -\lambda_i S_i(t)I_i(t) - \gamma_i S_i(t)I_i(t) - (\lambda_i + \gamma_i) \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) - (\lambda_i + \gamma_i)\omega_n f_n(t) - (\lambda_i + \gamma_i)\omega_t f_t(t) \tag{6}$$

$$\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} = \lambda_i S_i(t)I_i(t) - \mu_i I_i(t) + \lambda_i \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) + \lambda_i(\omega_n f_n(t) + \omega_t f_t(t)) \tag{7}$$

$$\frac{\mathrm{d}R_i(t)}{\mathrm{d}t} = \mu_i I_i(t) + \gamma_i S_i(t)I_i(t) + \gamma_i \sum_{j \neq i} \sigma_{ij} S_i(t) I_j(t) + \gamma_i(\omega_n f_n(t) + \omega_t f_t(t)) \tag{8}$$

$$F_i(t) = I_i(t) + R_i(t) \tag{9}$$

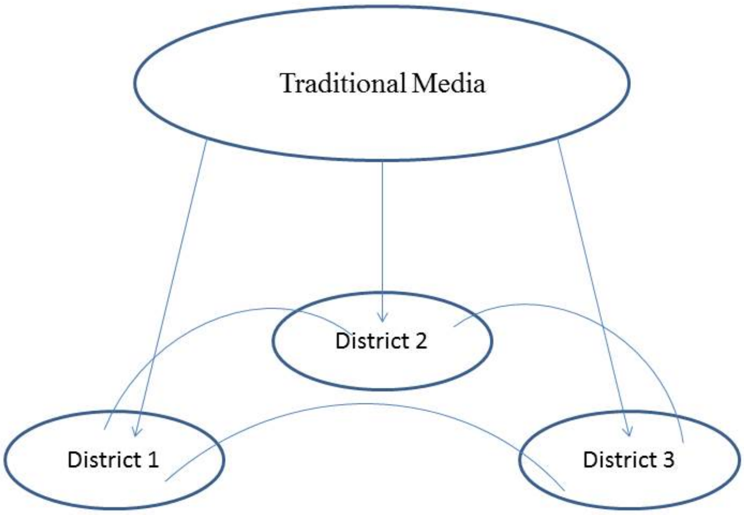

Figure 2: SIR Model with Traditional Media

$$F(t) = \sum_{i=1}^{n} F_i(t) \tag{10}$$

4.3 Complex Network Information Diffusion Model

Finally, we would consider the effect of Internet on information diffusion. Nowadays, Internet does shorten the distance between people in different districts, making the world a global village. What is more, people themselves can be the source of information on the Internet and other people can receive the information. Given all properties of Internet, we would put it on the highest level and refine our model. Figure 3 shows the final SIR Model.

The final SIR ordinary differential equations are listed as follows:

$$\begin{aligned}\frac{\mathrm{d}S_i(t)}{\mathrm{d}t} &= -\lambda_i S_i(t)I_i(t) - \gamma_i S_i(t)I_i(t) - (\lambda_i + \gamma_i)\sum_{j\neq i}\sigma_{ij}S_i(t)I_j(t) \\ &\quad -(\lambda_i+\gamma_i)(\omega_n f_n(t) + \omega_t f_t(t) + \omega_i f_i(t))\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} &= \lambda_i S_i(t)I_i(t) - \mu_i I_i(t) + \lambda_i \sum_{j\neq i}\sigma_{ij}(1+\omega_i)S_i(t)I_j(t) \\ &\quad +\lambda_i(\omega_n f_n(t) + \omega_t f_t(t) + \omega_i f_i(t))\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\mathrm{d}R_i(t)}{\mathrm{d}t} &= \mu_i I_i(t) + \gamma_i S_i(t)I_i(t) + \gamma_i \sum_{j\neq i}\sigma_{ij}(1+\omega_i)S_i(t)I_j(t) \\ &\quad +\gamma_i(\omega_n f_n(t) + \omega_t f_t(t) + \omega_i f_i(t))\end{aligned} \tag{13}$$

$$F_i(t) = I_i(t) + R_i(t) \tag{14}$$

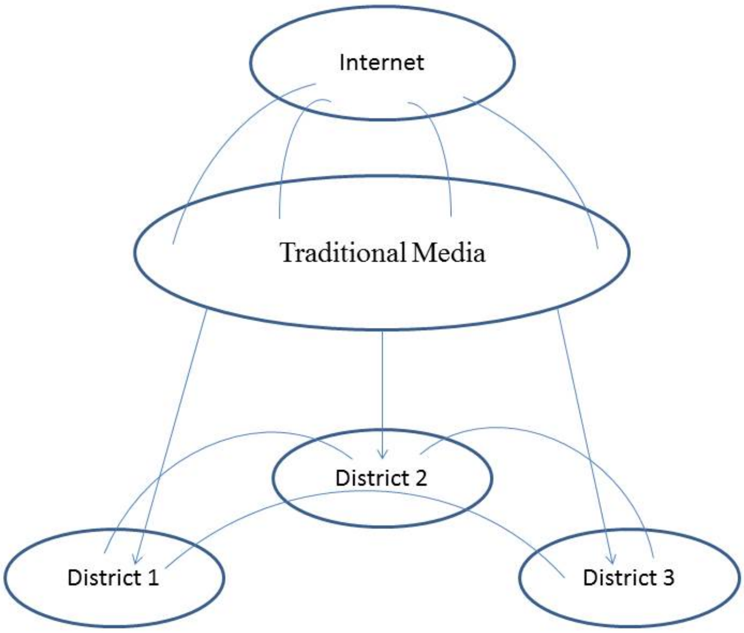

Figure 3: SIR Model with Internet

$$F(t) = \sum_{i=1}^{n} F_i(t) \tag{15}$$

$$F_i(t) = I_i(t) + R_i(t) \tag{16}$$

$$F(t) = \sum_{i=1}^{n} F_i(t) \tag{17}$$

4.4 Model Validation and Analysis

To test the validity of our model, we would use some data to train this model. Since it is difficult to obtain data about the diffusion speed of traditional channels, we assume that the diffusion situation through tradition channels is similar to that in Internet. Hence, we would use data from Tumblr and make some qualitative analysis.

The number of registers in Tumblr is approximately 20 million. We assume that some nodes in the complex network are high active and other nodes are common. We set active nodes have $\lambda_1 = 10000$ and common active nodes have $\lambda_2 = 30$. This is actually a rational assumption. In our social complex network, stars and celebrities can make a huge difference on information diffusion. Common people, however, have very limited diffusion ability.

We also set the rate $\mu = 0.99$ and the ratio of active nodes is 10^{-6}. The result is shown as below.

From Figure 4, we can see that the information diffusion process is gentle in first 5 days. Then the process is at its highest speed from day 6 to day 15. The diffusion speed fades from day 15 to day 25. Finally, after day 25, nearly all people in Tumblr receive this information.

However, it is unrealistic to let a single piece of information spread for that long time period. Most heated topics can only last for about 7 days. And due to information explosion, this time period would be even shorter. Reducing the immune probability μ, increasing the infected rate λ

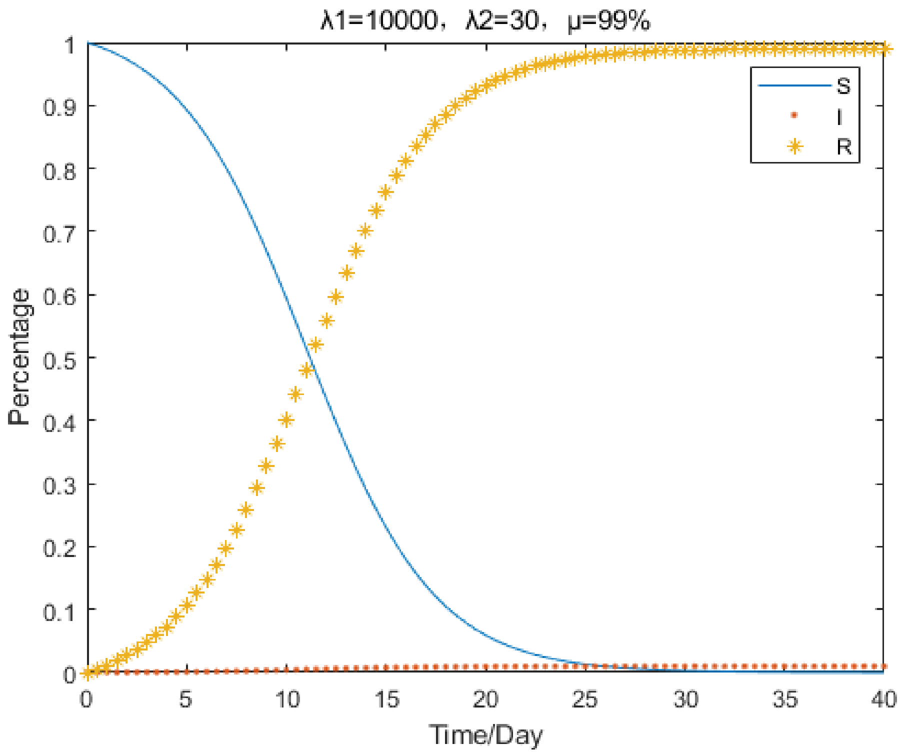

Figure 4: SIR Model Simulation

and increasing the ratio of active nodes can accelerate the speed of information diffusion. We would discuss these details in **section 6**.

5 Information Diffusion with Bass Model

To gat a deep understanding of information diffusion in complex network, we not only need to model the information diffusion through each media, but also need to predict the future development of information diffusion in each media. In this section, we would apply the Bass Diffusion model to predict the population growth of netizen and validate our model with data from United States Data Source. Then we would derive an equation to represent the future development of traditional media, and compare their trends together.

5.1 The Development Trend of Internet

The Bass Diffusion Model is proposed by Frank. M. Bass in 2004 [7], which describes the process of how new products get adopted in a population. The model presents a rationale of how current adopters and potential adopters of a new product interact. It is useful in determining the patterns of innovations in a population and predicting the diffusion of future innovations with information about the spread of older innovations. In Bass Diffusion Model, the term "innovation" can refer to any novel idea and product. And it is a trend to apply Bass Diffusion Model to social network.

In Bass Diffusion Model, adopters are divided into two groups: innovators and imitators. Innovators are the initial group of individuals adopted the innovation. Imitators are the group of individuals influenced by the previous adopters. The basic model formulation can be described as below:

$$\frac{F'(t)}{1-F(t)} = p + qF(t) \tag{18}$$

where $F(t)$ denotes the adoption fraction,which means the fraction of adopters at time t in the

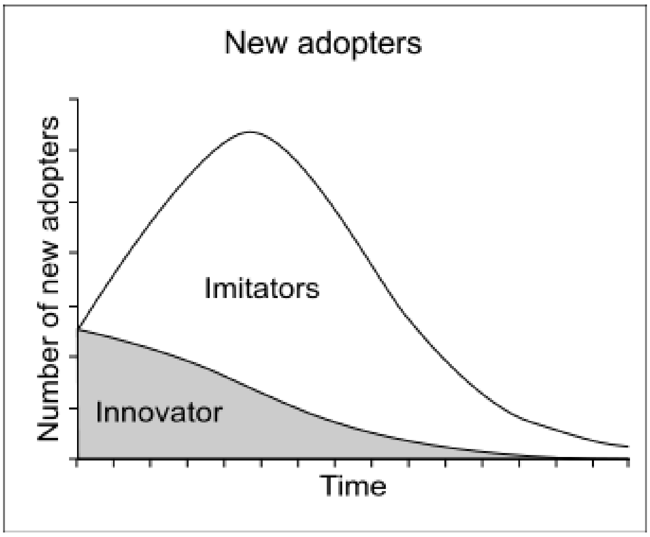

Figure 5: Bass Diffusion Model[8]

eventual adopters. p is the innovation coefficient, reflecting the influence by factors outside of the population. q is the imitation coefficient, representing the extent to which the adopters emulate other members.

We can calculate the detailed expression of $F(t)$ from Equation 18. The expression is shown below.

$$F(t) = \frac{p - pe^{-(t-C)(p+q)}}{p + qe^{-(t-C)(p+q)})} \tag{19}$$

where C is a constant.

To test the validity of the Bass model, we compare the result of the model with the actual data of Internet users in US[9]. The detailed data can be checked in **Appendix A**. We would use the curve fitting app in MATLAB to determine the parameters in the equation. The fitting curve and the factual data points would be displayed in Figure 5. The fitting parameters are listed in Table 3.

From Figure 6 we can see that the model fits the data of Internet users in US very well. We compute the R2 and get R2 = 0.9842. The calculated function gives an estimate of the number of Internet users in 2016, which is 284,154,941, and it is also very close to the actual data (286,942,362). However, because the data points are concentrated in the steep part of the curve, the accuracy of fitting may be affected.

5.2 Media Technology Influence

The Bass Diffusion Model of adoption and substitution, proposed by Frank Bass, study the process of diffusion and substitution of innovations like high technologies[10]. It successfully models the substitution of some hi-tech products. Nevertheless, it is not suitable for the substitution of media technologies, because users of different media technologies overlap.

Based on his substitution model, we make a new model introducing a transition factor δ in order

Table 3: Parameter Values for the Bass Diffusion Model

Parameter	Value
p	0.1816
q	-0.1198
C	1996

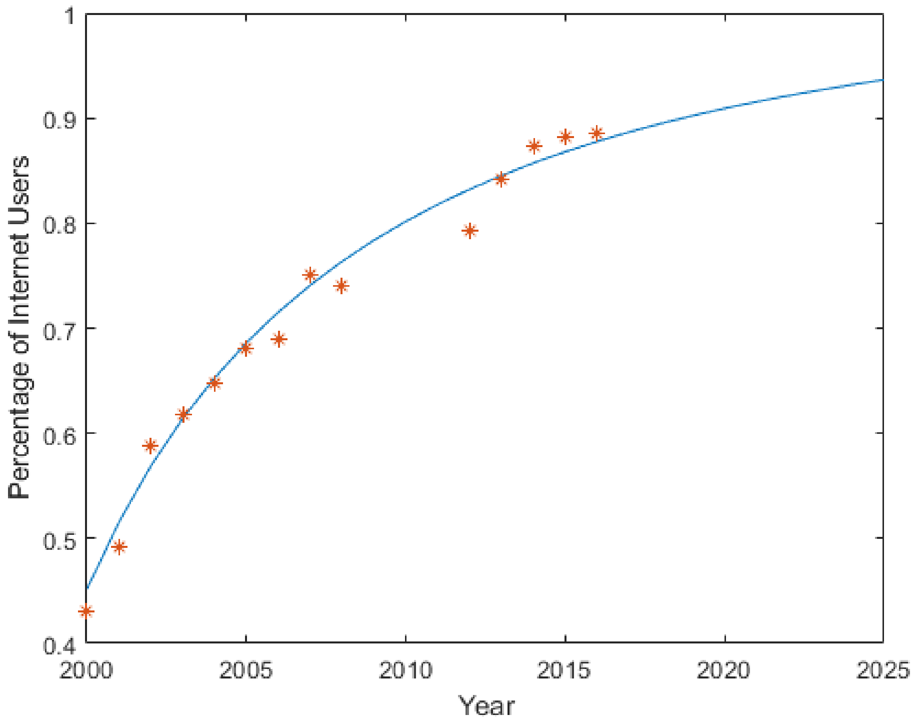

Figure 6: Actual and Estimated Internet User Percentage in US

to describe the trend of media technologies. Equations are listed as below.

$$G_i(t) = F_i(t - \tau_i) \prod_{j=i+1}^{m} (1 - \delta_{ij} F(t - \tau_j)) \quad (i = 1, 2, \ldots, m-1) \tag{20}$$

$$G_m(t) = F_m(t - \tau_m) \tag{21}$$

$$F_i(t) = \frac{p_i - p_i e^{-t(p_i+q_i)}}{p_i + q_i e^{-t(p_i+q_i)}} \tag{22}$$

Here, m denotes the total number of generations. $G_i(t)$ denotes the actual fraction of individuals using the ith generation of technology. $F_i(t)$ denotes the fraction of individuals using the ith generation of technology without the effect of other technologies. δ_{ij} is the transition factor, reflecting the extent to which users of the ith generation of technology will deviate his choice due to the effect of the jth generation. τ_i denotes the introduction time of the ith generation, and we assume $t = 0$ is the time when the first generation is introduced.

We obtain data concerning how sources of news change over year in the US[11] and use it to validate our model. We use the curve fitting app in MATLAB to fit the data, and the result are shown in Figure 8. In Figure 8, point '.' represents television; blue '+' represents newspaper and purple '*' represents Internet.

6 Information Diffusion Influence Factors

In this section, we would briefly discuss factors that would influence information diffusion process.

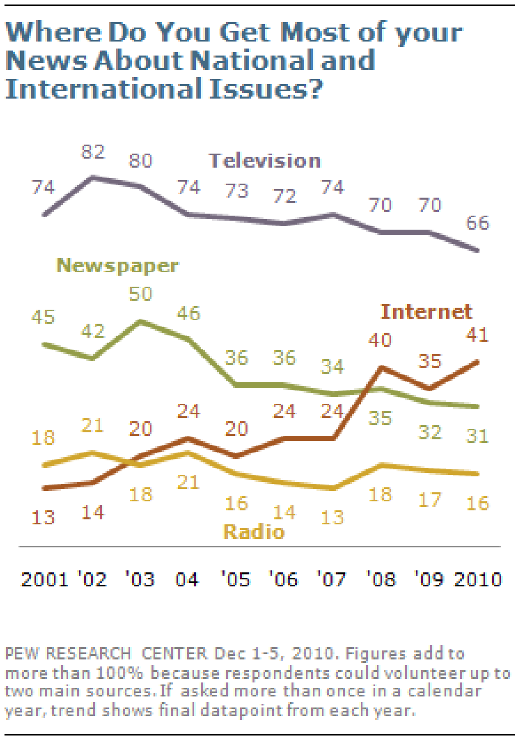

Figure 7: Media as News Source in US[11]

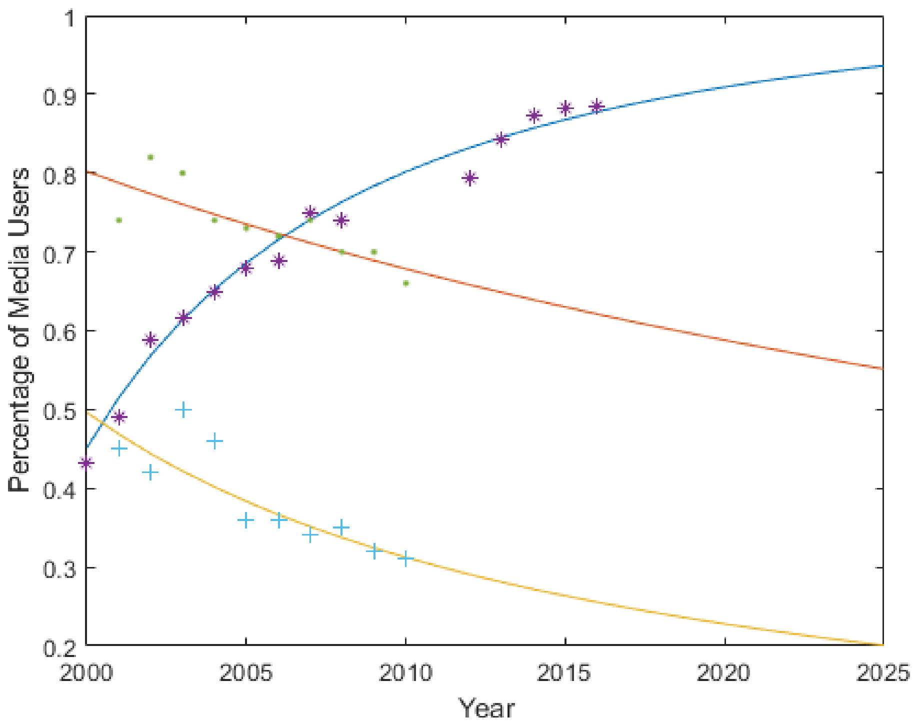

Figure 8: Trend of Media as News Sources in US

6.1 Information Value

As we have discussed above, the information value is proportional to the speed and coverage of dissemination, which means that people are more willing to spread the information. In this way, when the information value increase, both λ and the ratio λ/γ would increase. People are more likely to become **Infective** rather than **Removal**. And also, mass media would like to report certain information and ω increases.

We show the results in Figure 9. The two bottom lines show that the larger the rate between λ and γ is, the more informed people are.

People's initial opinion and bias are often reflected on the parameter λ. If the mass hold the view, the information would be easier to spread. However, if people reject certain opinion, the information diffusion would be hard. We show the result in Figure 9.

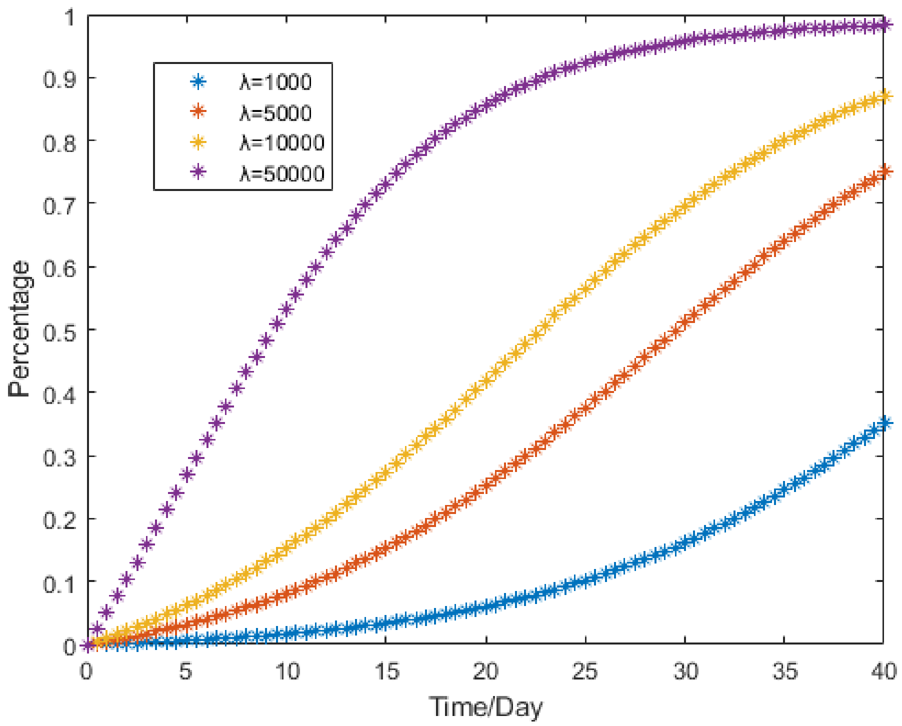

Figure 9: Influence of Information Value

6.2 Ratio of Active Nodes

It is a common practice that advertisers would invite some stars or famous web celebrities to make massive propaganda. This is actually improve the number of active nodes in the network. We first set $\lambda_1 = 10000$ and $\lambda_2 = 10$. Then we set the ratio of active nodes to be 10^{-6} and 10^{-5}. The result is shown in Figure 10.

From Figure 10, we can see that active nodes do play an important role in information diffusion and we cannot neglect their influence. Also, when it comes to the influence capacity of a certain active node, higher λ shows much more influence. In reality, these nodes are "Super V" in our social network.

6.3 Topological Structure of Complex Network

In SIR Model, we assume that uninformed people receive certain information at the rate λ. To take the topological structure into consideration, we make following assumption

$$\lambda = p \times v \tag{23}$$

In Equation 23, p is the probabiliy that an uninformed individual receives the information in one time unit, and v represents the average number of uninformed people that an informed individual

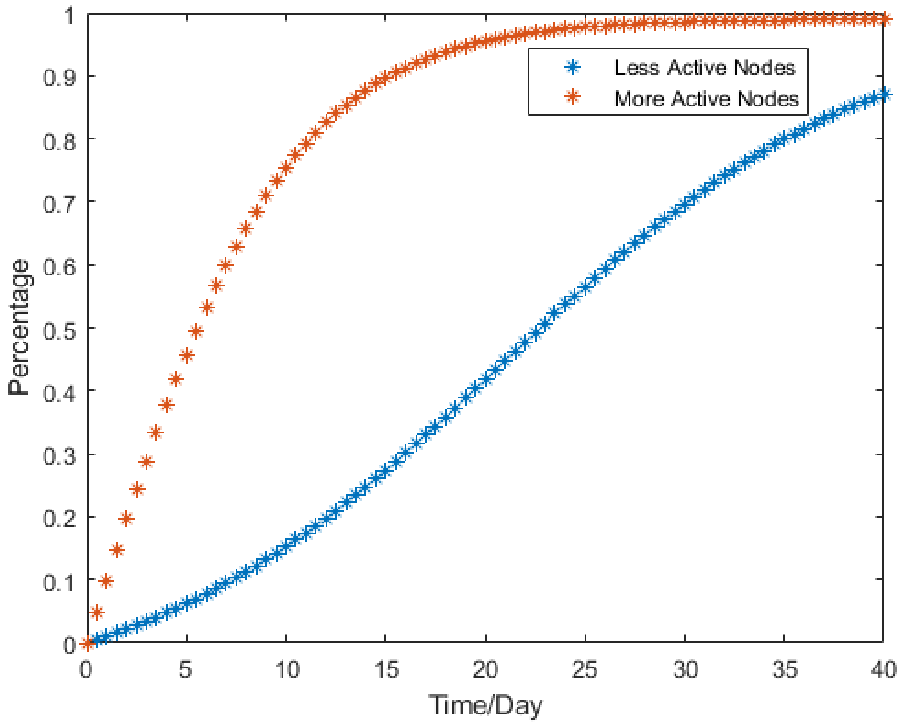

Figure 10: Influence of Active Nodes

can spread the information to in one time unit. To get an approximate value of v, we use following equation.

$$v = \frac{1}{n}\sum_{i=1}^{n}\deg(u_i) \tag{24}$$

In Equation 24, n denotes the total number of individuals and u_i represents the ith individual. $\deg(u_i)$ is the degree of the ith individual, which means the the number of individual ith can contact with.If the network has a dense graph, the speed of information diffusion is quite fast. Nevertheless, if the topogical structure of the network is sparse, the speed of information diffusion is slow.

6.4 p and q in Bass Diffusion Model

In Bass Diffusion Model, p is the innovation coefficient and q is the imitation coefficient. We get the values of p and q by curve fitting function in MATLAB. Now we will change the values of p and q and figure out their influence in final results.

We would take a close look at the estimated value of the percentage of Internet users in 2025 with p and q fluctuating by 0.05. The result is shown in Figure 11.

From Figure 11, we can see that although p and q only change a little, the estimated number has a fluctuation range of around 5.73%. Therefore, the estimated number is very sensitive to the value of p and q. Also, the accuracy of the estimation rely on that of curve fitting.

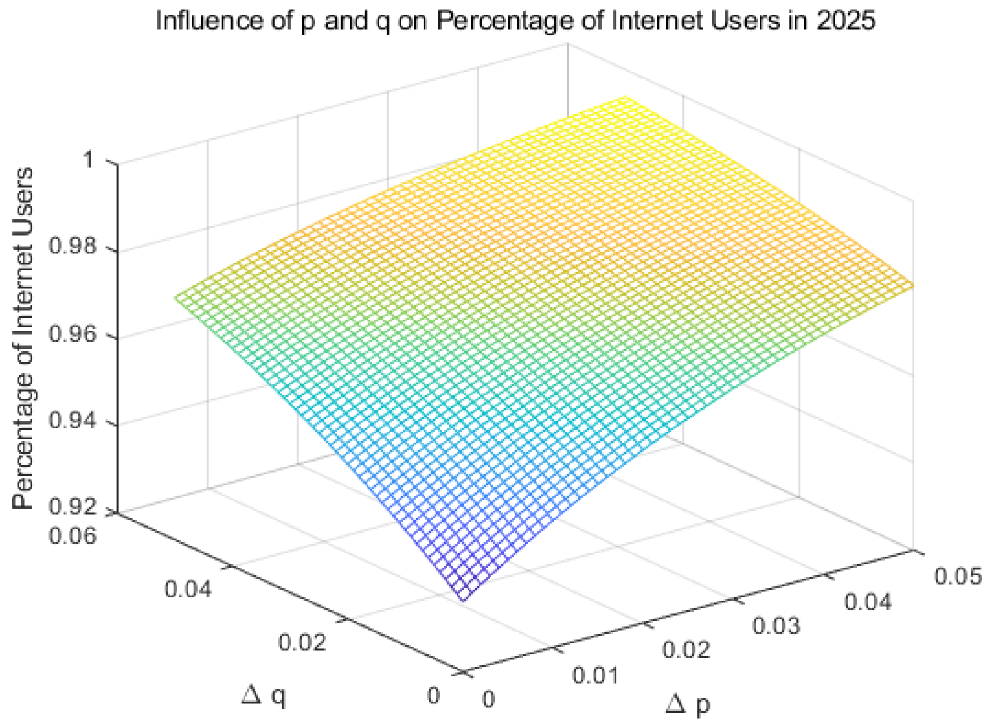

Figure 11: Influence of p and q on Percentage of Internet Users in 2025

7 Strengths and Weaknesses

7.1 Strengths

1. We build not only a SIR Model for information diffusion, but also a Bass Diffusion Model for the trend of media technology. Thus we can get an insight into how information will spread in complex network in future.
2. We revise the SIR Model for the spread of virus and apply it into the information diffusion. This SIR Model can be extended to other fields.
3. We successfully predict the trend of media technology with Bass Diffusion Model. Though it may seems simple, the result shows practicability of this method.

7.2 Weaknesses

1. Due to the lack of data, the estimation of some parameters may not be so accurate. Also, curve fitting may introduce some deviation in our model.
2. We do not take the population growth into account since the Bass Diffusion Model is usually used for diffusion in a short period. It is necessary to improve the model so that it can apply to a long term diffusion.
3. As for how people's opinions can be changed in today's complex social network, we only offer a rough framework and the specific equations and results need to be complemented.

8 Future work

Since there are many weaknesses in our work, plenty of efforts can be spent in the future to improve our work.

First, we may come up with a concrete method to define the information value or take more factors into consideration to measure the information value.

Secondly, we can propose a quantitative method to define the speed of information diffusion and test its relationship with the information value.

Third, we can quantify our illustration for the process of how opinions are changed with more detailed equations and illustration for further analysis.

Furthermore, we can also test the possibility of applying our model to more specific communication channel such as Facebook. By doing so, we will probably improve our mode even better.

9 Conclusion

In this paper, we propose the SIR Model for information diffusion in complex network and the Bass Diffusion Model to predict the trend of media. Actual data is collected to validate our models, which show that our models fit the reality well. We then consider factors which may influence our model and make detailed discussion.

References

[1] M. Granovetter,"Threshold models of collective behavior," *American journal of sociology*, pp. $1420 - 1443$, 1978.

[2] D. Acemoglu, A. Ozdaglar, and E. Yildiz, "Diffusion of innovations in social networks," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pp. 23292334, IEEE, 2011.

[3] D. A. Soysa, "Viral spread models for multimedia content propagation in social networks," 2013.

[4] Kermack WO1, McKendrick AG, "Contributions to the mathematical theory of epidemics–I. 1927," *Bulletin of Mathematical Biology*, $53(12) : 3355$.

[5] J. Cannarella and J. A. Spechler, "Epidemiological modeling of online social network dynamics," arXiv preprint arXiv:$1401.4208, 2014$.

[6] F. M. Bass, "Comments on "a new product growth for model consumer durables the bass model," *Management science*, $vol.50, no.12, pp.18331840$, 2004.

[7] Bass, Frank M, "Comments on "A New Product Growth for Model Consumer Durables": The Bass Model". *Management Science*. $50(12) : 18331840$, 2004.

[8] "Bass diffusion model" https://en.wikipedia.org/wiki/Bass_diffusion_model.

[9] "United states internet users." http://www.internetlivestats.com/internet-users/united-states/.

[10] J. A. Norton and F. M. Bass, "A diffusion theory model of adoption and substitution for successive generations of high-technology products," *Management science*, $vol.33, no.9, pp.10691086$, 1987.

[11] A. Kohut, C. Doherty, M. Dimock, and S. Keeter, "Internet gains on television as public's main news source," *Washington, DC: Pew Research Center*, 2011.

Appendices

Appendix A Internet users in US

Year	Internet Users	Penetration (% of Pop)	Total Population	Non-Users (Internetless)	1Y User Change	1Y User Change	Population Change
2016	**286,942,362**	88.50%	324,118,787	37,176,425	1.10%	3,229,955	0.73%
2015	**283,712,407**	88.20%	321,773,631	38,061,224	1.70%	4,642,081	0.73%
2014	**279,070,327**	87.40%	319,448,634	40,378,307	4.50%	12,041,883	0.73%
2013	**267,028,444**	84.20%	317,135,919	50,107,475	7%	17,392,468	0.74%
2012	**249,635,976**	79.30%	314,799,465	65,163,489	14.60%	31,807,857	0.77%
2011	**217,828,119**	69.70%	312,390,368	94,562,249	-1.90%	-4,322,107	0.81%
2010	**222,150,226**	71.70%	309,876,170	87,725,944	1.80%	4,015,534	0.86%
2009	**218,134,692**	71%	307,231,961	89,097,269	-3.20%	-7,175,434	0.91%
2008	**225,310,126**	74%	304,473,143	79,163,017	-0.40%	-931,839	0.93%
2007	**226,241,965**	75%	301,655,953	75,413,988	9.80%	20,233,843	0.94%
2006	**206,008,122**	68.90%	298,860,519	92,852,397	2.30%	4,727,778	0.92%
2005	**201,280,344**	68%	296,139,635	94,859,291	5.90%	11,194,860	0.89%
2004	**190,085,484**	64.80%	293,530,886	103,445,402	5.90%	10,543,491	0.87%
2003	**179,541,993**	61.70%	291,005,482	111,463,489	5.90%	9,963,241	0.88%
2002	**169,578,752**	58.80%	288,470,847	118,892,095	20.90%	29,307,602	0.94%
2001	**140,271,150**	49.10%	285,796,198	145,525,048	15.10%	18,402,034	1.03%
2000	**121,869,116**	43.10%	282,895,741	161,026,625	21.50%	21,589,192	1.13%