

Measuring the Evolution in Society's Information Networks

Anran Zheng

School of North China Electric Power University Baoding, Baoding 071000, China;

zhenganran1995@126.com

Keywords: Network simulation, Topology structure, Machine learning

Abstract. This paper sets up the communication network model to explore social information network's evolution. The communication network model consists of topology structure, communication capacity analysis and topology prediction. Nodes are composed of five kinds of media. Links indicate that there is information dissemination relationship between different nodes. The network communication capabilities are determined by the degree distribution, average degree and average path length of topology structure. It adopts four indexes including GDP, gross enrolment ratio, patent applications and urban population to represent the characteristics of an era, inputs the index data and media market share of media to machine learning and set up topology structure. It uses simulation to study and analyze the evolution of network communications capacities in five periods. After predicting the proportion of each medium in 2050, we obtain the network topology structure. The communication capacity of network will be improved greatly in 2050 by 1.12 times the average degree.

Introduction

We live in a media-saturated world and rely on a variety of old and new media for information, entertainment, and connection^[1]. However, before the invention of newspaper and other media was invented, news got around by word of mouth on crossroads, at campfires or at markets. With the development of technology, various media such as newspaper, radio, televisions and Internet had been produced which had caused that the flow of information differed in different times.

To explore the evolution and influence in society's information networks, we set up the communication network model. We study network topology and then change parameters of the network which are evolved over times to get the networks in five periods. Finally, we predict the topology structure and communication capacity of network in 2050.

Communication Network Model

Concepts. *Node:* We define different media source as node, which can receive and transmit information. We use newspaper offices, broadcasting stations, television stations, websites and human all these five class as media sources.

Link: If there is a relationship about information spread between two nodes, it means that they have link. The existing links have no differences, so the link is unweighted. Also, the link is undirected, because people will share information if they have friendly relationship and media will focus on each other to maintain its competitiveness.

Network Topology Structure. *Node:* we use simulation to build a small network including 1040 nodes. The composition rules of nodes are as follows:

(1). The proportion of five class nodes change with time. At different periods, the structure of the communication network is different.

(2). Hub is the mainstream media form in the studying period. It is characterized by very few in numbers, but very large in degree. Different periods have different degree of dependence on different media, and the influences of media are different.

(3). Human node consists of three kinds of people: young, middle-aged and elderly people. As long as the network has a variety of media co-existence, the relationship between different types of

people and different media is different. The composition of three kinds of people in people node will change with time.

Link: the connection rules are as follows:

(1). The link between people: we establish the relationship among people based on small-world model. Small-world network is sparse network with high clustering coefficient, relatively short average path length and entropy which can be extended^[2]. Here, we generate a small-world network about relationship among people by following method:

Step 1: Constructing a regular graph with proximity coupling, which means that each node is connected to N nodes in their right and left sides.

Step 2: Based on proximity coupling, we use random probability P to reconnect. The values of N and P are based on the level of development of the time.

(2). The link between people and other media: combining the ways that people use media in real world, we develop following rules about the links among people and other media: a). People tend to use the media that they are used to. Therefore, the links between people and media are bound with times. b). Young people are the main force to use media hub node. Based on the two rules above, we use the following method to make links:

Step 1: According to the characteristics of the times, we set the proportion of young people, middle-aged and elderly person as a_a, a_m, a_o and the probability of their connection with the four media as p_{ij} .

Step 2: Use the Monte Carlo to confirm whether there is link between the two. We generate a random number between 0 and 1, and determine whether the random number is within the range of the connection probability.

(3). The link between media: each media have links to other media nodes.

In the current era, for example, the network topology is as follows:

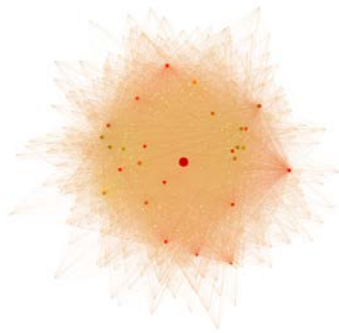


Figure 1 network topology structure of 2012

Dynamic Network Parameters. The market share of four types of media (%) depends on the degree of the development of the times. We use four indexes, namely, GDP per capita (current US\$), gross enrolment ratio, patent applications and urban population to reflect the development of the times which reflects economic, education, technology and the ability to accept new media respectively^[3]. Having considered the fact that the data is limited, we decide to use neural network learning to describe their relationship. To reduce complexity, we choose the most basic BP neural network. The BP model, however, has a major setback of slow convergence^[4]. Hence we introduce Particle Swarm Optimization (PSO) to help the neural network define best weights of linking edges. In the BP neural network, there are:

(1). Input layer: the neurons on the input layer are GDP per capita, gross enrolment ratio, patent applications and urban population.

(2). Output layer: the neurons on the output layer are the market share of newspapers, radios, televisions and Internet.

After using data to train the BP neural network, we can use it to get the market share of the various media corresponding to the input.

Network Communication Capacity. We measure the network communication capacities by following three indexes, namely, degree distribution $p(k)$, average degree $\langle k \rangle$, the average path

length L .

Evolution of Society's Information Networks. We simulate the communications network in five periods based on the market share of various media and their ability of spread ^[5]:

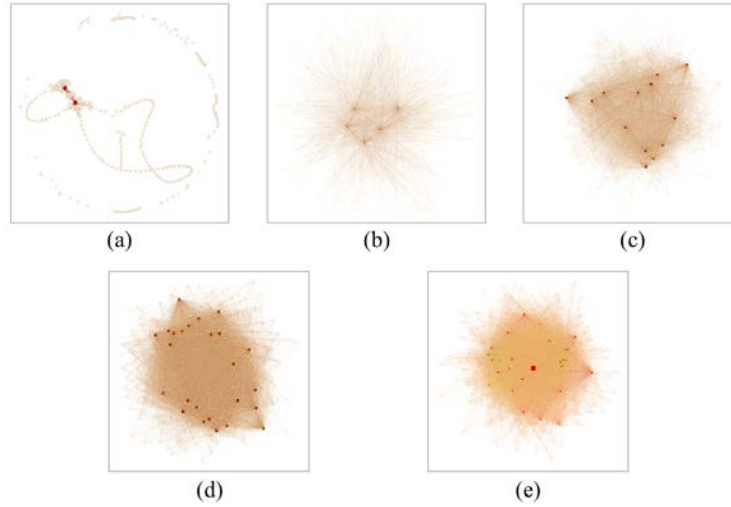


Figure 2 Network topology structures of five periods
(a)1870s.(b)1920s.(c)1970s.(d)1990s.(e).2010s

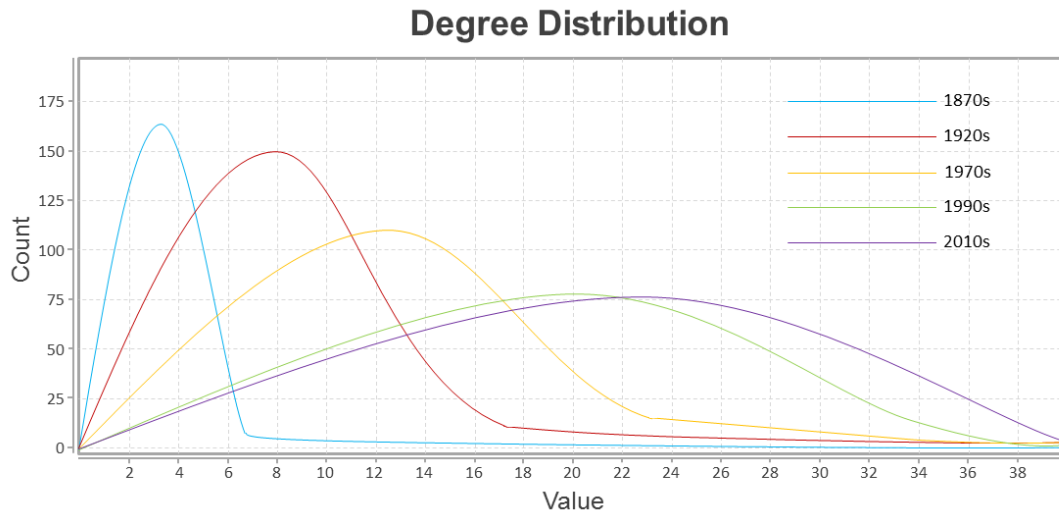


Figure 3 Degree distribution curves of five periods

Table 1 Average degree and average path length of five periods

	1870s	1920s	1970s	1990s	2010s
Average Degree $\langle k \rangle$	3.329	10.62	21.983	40.344	41.938
Average Path Length L	66.717	39.192	25.42	13.305	9.232

From the figure and table above, we can get the characteristics and evolution process of communication network:

(1). The average path length has decreased gradually over time. Due to the increase of the mutual links, the path length between nodes shortens obviously and the range of spread becomes wide.

(2). The degree of human nodes has increased greatly over time. From the degree distribution curves, we can see that the peak position constantly moves to the right.

(3). The degree of hub has changed in the order of magnitude. Hubs play a more and more important role in the networks which enhance the connections between the human nodes and make entire network closer.

Prediction Capability. We use regression analysis to fit function for four input indexes respectively. Then we get the values from functions: 88449(GDP per capita), 98.8(gross enrolment ratio), 644622(patent applications), 88.315(urban population). After comparing them with prediction values from Goldman Sachs ^[6], we find the difference is small. Input the values of four indexes to BP model and we get the results:

Table 2 predicted values and number of media node of 2050

	Market Share(%)	Percentage(%)	Nodes Number
Televisions	37.33	26.9	11
Papers	7.85	5.7	2
Radios	12.14	8.7	3
Internet	81.61	58.7	24

The communication capacity is: average degree $\langle k \rangle = 47.246$, average path length $L = 5.282$ and the degree distribution which is shown in figure 4

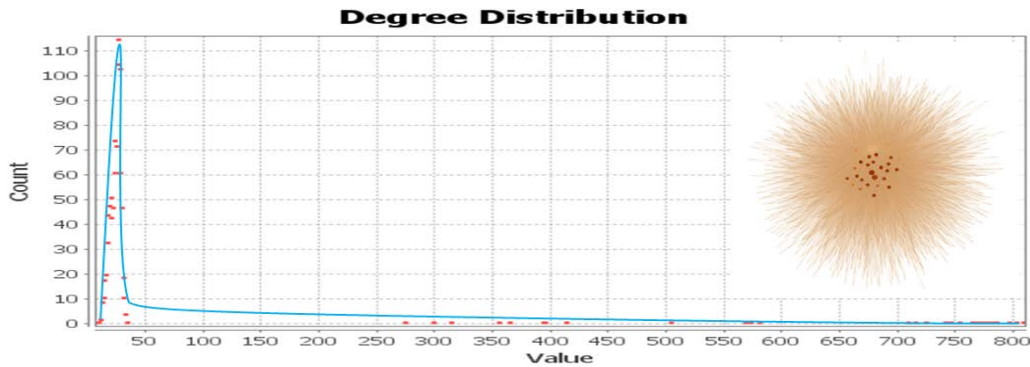


Figure 4 Degree distribution and network topology of 2050

Summary

From the results of Communication Network Model, we can get the characteristics and evolution process of communication network: the dominant media become more influential. The human nodes have better communication abilities. The communication capacities of networks have experienced evolution from weak to strong.

References

- [1] Bowditch J L, Buono A F, Stewart M M. A primer on organizational behavior[J]. 1997.
- [2] Lewis T G. Network science: Theory and applications[M]. John Wiley & Sons, 2011
- [3] World Bank, <http://data.worldbank.org.cn/>
- [4] Feng Shi, Xiaochuan Wang, Lei Yu, Yang Li, YanLiang Zhang, 2013, MATLAB neural networks: 30 case studies, pp21-35, Beijing: BUAA Press, 7-5124-0034-4
- [5] Watching, Reading and Listening to the News <http://www.people-press.org/2012/09/27/Section-1-watching-reading-and-listening-to-the-news-3/>
- [6] Wilson D, Trivedi K, Carlson S, et al. The BRICS 10 years on: halfway through the great transformation[J]. Goldman Sachs Global Economics Paper, 2011, 208.