Semantic Based Network Growth in Instant Messaging Environment

Wei Ren
College of Computer and Information Science, Southwest University
2 Tiansheng Rd, Beibei, Chongqing, China, 400715
Email: oicq@swu.edu.cn

Yuhui Qiu
College of Computer and Information Science, Southwest University
Email:yhqiu@swu.edu.cn

***Abstract*—The instant information exchanging network of the Internet is interpreted as the consequences of invisible connection between humans. In the graph based studies the nodes are human beings and the edges represent various social relationships. The interactions among users can be interpreted via the formation and evolution of semantics. The interactive as well as intertwined behaviors are the foundation of network itself; at the same time, they shape the way how and where the network will evolve. This paper proposes a network growth model based on the semantic similarity as well as popularity of nodes. In our model, the nodes represent *Sina Weibo* blogs and are with semantics, the links are subscribing hyperlinks between nodes. The probability of link establishment between two nodes then calculated from the similarity between nodes. The data and experiments are based on *Sina Weibo* blogs, which are the continuous results of interactions by users. We collect data using WebCrawler from *Sina* API, obtaining a portion of the whole network. Results show that the statistic properties of *Sina Weibo* are in close analogy with that of social network and also the characteristic complex network. The studied network contains a number of very high-degree nodes; these nodes are the cores which small groups strongly clustered, and low-degree nodes at the fringes of the network. However, some nodes with too much semantics (especially under one category) are in decreased chances of having links from newly added nodes. The reason may lies in that the over-abundant semantics remains confusion for knowledge acquiring.**

***Index Terms*—Semantic, Network Growth, Instant Messaging Network**

I. Introduction

The social network in the Internet is interpreted as the system of invisible connection and interactions between humans. Fads and ideas spread through the network, the nodes are human beings and the edges represent various social relationships. Users' behaviors form the forums, build up the communities and further on establish small world networks that finally come as a whole as the Internet network system. The interactions and the topology between network users are the key issue in the study of network growth model of online social network. From these interactions one can obtain the inner structure of the network users, that is, the unseen relationship of users with one another, the possibility the two users establish a connection and the trends and ways the network might evolve.

Scientists have notices the online social networks, especially the instant messaging networks are the true and in-time reflection of real human world. The establishing of connection, exchanging information as well as forming small communities are the same as in that in the human society. So the online interactions networks provide researchers a very direct and effective way to study human society.

The graph based web discovery methods have the major limitations such as no demonstration of semantic relations between two connecting communities and the meaning of the found communities are invalid until semantics are assigned. Semantic Link Network (SLN) [1, 27] is proposed for better resource discovery on the basis of the WWW, which aims to achieve higher efficiency of community discovery and reproduce relationships between communities in a semantic way. However, how different semantic communities in an semantic based network establish connections with each other still remains as an unsolved issue.

Semantics originally is the study of meaning. It focuses on the relation between *signifiers*, like words, phrases, signs, and symbols, and what they stand for, their denotation. Zhuge's define semantics as: a semantic node can be a notion, a URL or a semantic community. And the semantic link is a kind of relational knowledge represented as a pointer with a tag describing semantic relations such as *causeEffect, implication, subtype, similar* or *equal* and so on [1]. Growing semantic communities and the links between them enlarge the scale as well as enrich the semantics of the network. How a semantic based network may evolve is an issue less discussed but crucial for study of our human interaction and society.

In this paper, we studied the characteristics of network growth model of semantic based network in a dynamic network scenario. By using the WebCrawler we collected

doi:10.4304/jcp.9.10.2387-2394

empirical blogs data of 42 days, analyses the mechanism of semantic similarity based evolution of *Sina Weibo* network; then we studied the nodes' indegree and outdegree of network and experiments show that the indegree and outdegree are not much the same as we assumed.

The rest of the paper is organized as follows. In Section Ⅱ, we present an overview of previous work. In Section Ⅲ, we give the detailed design of proposed routing algorithm. In Section Ⅳ, we present our simulation model and its results. Finally, we offer conclusion in Section Ⅴ.

II. Related Work

According to the social network perspective, individual behavior is the consequences of others. Thus, to understand individual behavior, we need to "describe patterns of relationships between actors, analyze the structure of these patterns, and seek to uncover their effect on individual behavior" [14]. In the paper they used social network analysis as a method to evaluate the society level structures and processes in an instant messaging network environment. On the other point, they give a glimpse of an emerging field of knowledge building and acquisition processes. Social network analysis is a commonly used method to study social interactions of online groups at an individual level as well as group level [6, 7, 23].

The online instant messaging network is different from the traditional social network observed in the real life. The interactions are quick and far more frequent. The network such as *Youtube*, *Yahoo!* and *Twitter* gain tremendous amount of online users globally, which therefore provide powerful means of sharing, organizing, and finding content and contacts. The popularity of these sites provides an opportunity to study the characteristics of online social network graphs at large scale.

In this paper we focus on *Sina Weibo*. *Sina Weibo* is now the most popular and prevailing online social platform in China. *Sina Weibo* based on user relations; the information shared fast and instantly, and the cost can be ignored. These advantages enhanced good user experience and become the *No.1* online social platform of instant information sharing network.

A. Social Network Growth Models

There have been considerably many studies on the network growth model [2]. Price first described in 1965 the archetypal model of a scale-free network, by studied the network of citations between scientific papers. The paper takes the idea of Simon [3, 4] and applies them to the growth of a network. Findings show that the in- and out-degree of a vertex is follows the power-law distribution. That the rate at which a paper gets new citations should be proportional to the number that it already has. This phenomenon is referred as Matthew Effect.

Barabási and Albert [12] proposed in the journal of Science a network growth mechanism similar to Price's but with one major difference, which they call preferential attachment. In the model of Barabási and Albert, the edges are undirected, so there is no more distinction between in- and out-degree. Compared with that of Price's, the model misses a crucial feature of the networks observed in real world but solved Price's problem of how a paper gets its first citation or a Web first gets its first link. This preferential attachment received a considerable attention and become the mostly popular mechanism under which the network growth models are studied.

Krapivsky and Redner [5] analysed the model of Barabási and Albert in details in which the probability of attachment to a vertex of degree k. The study shows the older vertices in a network have higher expected mean degree and the overall power-law degree distribution of the whole network is a result primarily of the influence of these earliest vertices.

Adamic and Huberman [6] analyses the model of WWW by applying preferential attachment and found that, using actual Web data, there is no fixed correlation. This may caused as the dynamics of the Web are much more complicated for the simple model of Barabási and Albert's, at the same time the degree of vertices is also a function of their intrinsic worth. Some Web sites are useful to more people than others and so gain links at a higher rate.

The network growth model (also known as preferential attachment) by Barabási and Albert is widely accepted and well studied by researchers. The SLN is much in similarity of social network as the semantic links are formed as a result of human interactions. This paper studies the network growth model of SLN using the model of Barabási and Albert.

As Capocci et al. [7] have pointed out that preferential attachment usually associated to networks with the features that triggered by local events, the interactions between uses are most crucial issue that as soon as the interactions begin, the semantics are generated, then the semantic links are formed, finally the semantic link network itself is expanded [10]. The network growth model is then used in the generation of semantics links in an SLN [8]. The most important feature of this network growth model is that the generations of links in the network are based on the reasoning rules rather the graph-based theory measures.

B. Semantic Based Network

The concept of Semantic Link Network (SLN) is proposed by Zhuge in [1] to support discovery and learning in a semantic context. A semantic node in SLN can be a notion, a URL or a semantic community. And the semantic link is a kind of relational knowledge represented as a pointer with a tag describing semantic relations such as *causeeffect, implication, subtype, similar* or *equal* and so on [1]. Here we see the users are the nodes of network and their interactions are the links between nodes. The semantics come from the users' interests or the topics they focus, and also the direction of links can be semantic, too. The semantics of the nodes and links help the researchers to understand the intrinsic relations of network, at the same time it forms the semantic reasoning basis on which the preferential attachment is applied.

Various explicit and implicit semantic relations in the world constitute networks, which can be formalized into a loosely coupled semantic data model for managing various web resources. It consists of semantic nodes, semantic links between nodes, and a set of relational reasoning rules like $\alpha \cdot \beta \Rightarrow \gamma$ (i.e., the connection of semantic relation α and semantic relation β implies semantic relation γ).

The *Sina Weibo* is a social network in abstract, which is in a instant messaging based platform but just the network forming ways are different. This inspires the idea that we apply measures in social network study on the network growth model of *Sina Weibo*.

III. Semantic Based Network Growth

With all the attributes that are similar to social network, we use the most prevailing network growth model in social network, the Barabási and Albert model, to model the way in which SLN may evolve.

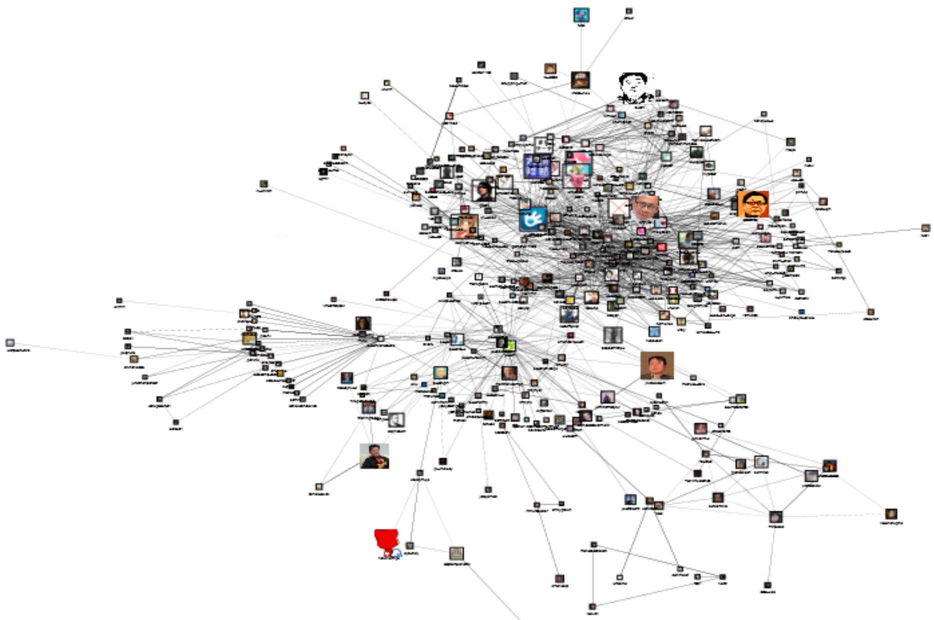


Figure 1: A part of *Sina Weibo* network

A. Barabási and Albert's Network Growth Model

Barabási and Albert's network growth model is based on the idea that two nodes are more likely to form a link if they share more same features or neighbours [22, 24]. The basic premise is that the probability a new edge involves a node is proportional to the shared features or neighbours.

The model of Barabási and Albert first demonstrated the power-law distribution. And the algorithm of the model is as the following:

(1) Growth: Starting with a small number (m_0) of nodes, at every time step, adding a new node with m ($\leq m_0$) edged that link the new node to m different nodes already present in the system.

(2) Preferential attachment: to choose the node or nodes to which a new node should connect with, the rate is the proportional to the number that it already has.

Let p_k be the fraction of nodes in the network with k degree with $\sum_k p_k = 1$, m is the mean degree of nodes with degree k, then the original probability of a new link attaches to any of the nodes with degree k is

$$\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}$$

As in the model the edges are undirected, so there is no in- and out-degree, and then the probability is

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

The mean degree of the network this time is $2m$, since there are m edges for each node added and each edge is now undirected, which then contributes two ends to the degrees of network of nodes. At this moment when a new node with m links is adding into the network, the mean number is $1/2\, kp_k$, which is independent of m.

Note that the number of nodes with degree k increases because of influx from nodes previously of degree $k-1$ that have also acquired a new link, except those nodes with degree m, which in this case we use $m = 1$ for simplicity.

Let $p_{k,n}$ represents value of p_k, where the network contains n nodes, then the network changes as:

$$(n+1)p_{k,n+1} - np_{k,n} = \begin{cases} 1/2(k-1)p_{k-1,n} - 1/2kp_{k,n} & \text{for } k>m \\ 1 - 1/2mp_{m,n} & \text{for } k=m \end{cases}$$

There is no node with $k<m$.

For stationary solutions let $p_{k,n+1} = p_{k,n} = p_k$, we have

$$p_k = \begin{cases} 1/2(k-1)p_{k-1,n} - 1/2kp_{k,n} & for \quad k > m \\ 1 - 1/2mp_{m,n} & for \quad k = m \end{cases}$$

B. Semantic Based Sina Weibo Growth Model

Instant messaging network shares great similarities with that of social network. So the network growth mechanisms are like each other at some points [13, 21, 25, 26]. The network expansion is based on the interactions of people.

People are nodes in the network, their actions and reactions are the links between them. All their actions, in our case, the actions are mainly message posting, replying, forwarding and *etc.* Then we can easily abstract the semantics within and results of semantic deductions can happen. Figure 2 is the simplified reasoning mechanism.

In semantic reasoning, we have three instances *A*, *B*, and *C*; there are semantic links between *A* and *B*, *B* and *C*. We want to find out if there is any inner relationship between *A* and *C*.

Then we use $R= \{\alpha, \beta, \gamma\}$ to represent the result of reasoning. α, β are the semantic link of instance *A* and B, α, β can often be considered as a relation between each pair. This kind of relation is directed, no reverse can be accepted as the semantic between the two will change *and* the result γ will alter. The γ, on the other hand, do not necessarily be an instance or utility γ, but instead γ is rather a semantic link or relation.

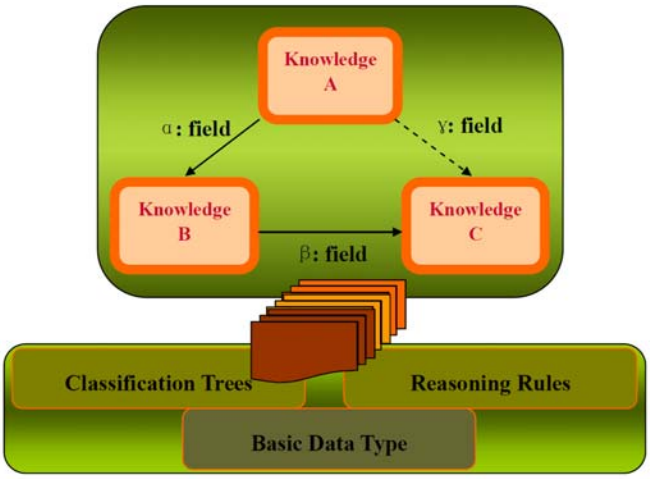


Figure 2: Inner reasoning mechanism of SLN

In semantic network specifically, the nodes are with semantic meanings, the semantics can be “a concept, an instance of concept, a schema of data set, a URL, any form of resources” [1]. And the links have semantics too as the nodes (users) have different relationships between each other.

Here we define the semantic of a node as the category or categories it belong and the tags they have (note that in our study one blog can focus on two different topics and thus is classified into two categories). We then define the richness of a node’s semantics by the number of descriptive words that the node may have. For convenience we call the words *Se-word.* The semantic of a node v_i is then S_i; S_{ij} is the j^{th} descriptive word of node v_i. Then the semantic similarity between two nodes γ are represented as $\gamma= (S_1 \cap S_2)/ (S_1 \cup S_2)$; S_1 and S_2 are the semantics of the related nodes.

On another aspect, the links a node has is also a critical issue as we observed in real life. Portal websites have enormous links connected to them and often get far more new links than those of nodes which own less links. Here we use w represents the popularity of a node. The popularity is the proportion of links which are pointed to the node. And we have $w = q / c$, q is the number of links and c is the mean number of links in the network. w is $[0,1]$. w is calculated as follows:

$$c = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

$$w = q / \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Often, the c is difficult to evaluate precisely as the scale of a network can be very large, so here we calculate c approximately. In this study c is based on the sample numbers we collected from *Sina Weibo*.

Now the semantics of a node, the similarity between the nodes and the likeness of a new adding node can be interpreted with straightforwardness. Thus the statistical properties of a semantic based network can be observed.

In our experiment we collect data from *Sina Weibo* blogs at a period of 42 days. By observing the data we noticed that the number of *Se-words* of one blog can be more than a few, so we give an upper limits u of node semantics. And u can be varied as to the characteristic of a network. If a node has more than u *Se-words*, we give credit to the first u ones. In our case, after the empirical experience, we have $u= 5$.

In a semantic based network such as *Sina Weibo*, a new adding node’s semantics is known forward. This is because the node will have semantics as soon as the user start the blog and focus on the topic he/she interested in. This process is usually done as users click the “follow” option, meaning that this user is collected to another blog. By this way the user submits his/her own interests to the network. And the node now owns semantics. This crucial characteristic allows us the convenience of ignore the process of calculating the original connecting probability that needed in Barabási and Albert’s model. The indegree of one node now does not necessarily have something to do with the link establishing probability.

So we have:

$$p = w \cdot \gamma$$

w is the popularity of a existed node in the network, γ is the semantic similarity between a new adding node and the current one. p is the probability the link between the two nodes.

In our study, because the semantic of the new adding nodes are known forward, so each new adding node has different prior probabilities.

As shown in Figure 3, in semantic based network such as social science, for instance, the node with semantics "Psychology, Education, Sociology" has more links than the node with semantics "Development psychology". Yet, a new node with semantics "adult psychological problem" will more likely link to the node with semantics "Development psychology", than the also relevant node "Psychology, Education, Sociology".

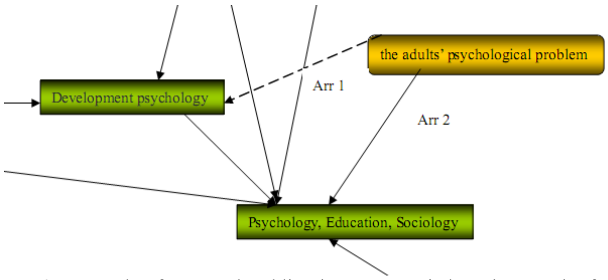


Figure 3: Example of new node adding in an semantic based network of social science

The link formation in a semantic based network, according to Zhuge [1] , given the situation of two nodes in the network containing a same semantic, the more semantic a node has, on the contrary, the less likelihood this node will have establishing a link with the new adding node [16, 17, 18].

In this paper we propose the instant messaging network growth model by calculating the semantic similarity between nodes. During the network growing process, the semantic similarity determines to which node a newly generated node will connect. Barabási and Albert model studies the probability of old nodes in a network to which a new node with certain amount (*m* as described above) of links will connect. By analyzing semantic similarity as well as the relations between nodes we may found the way how an instant messaging semantic based network might grow. Hope the finding of this work can be useful for better semantics resources and communities discovery in the future.

IV. Experiment And Analysis

We study *Sina Weibo* in our experiments. *Sina Weibo* is an instant messaging information network for sharing and discovering events and personal opinions. It enables users to post, send and reply messages. These messages are displayed on the author's profile page and pushed to the author's subscribers.

Each user is a semantic node in the network of *Sina Weibo*. There are two kinds of links a blog have: *"follow"* and *"fan"*. One user's *"follow"* are the other links lead to his or her profile page. And the *"fan"* links are other blogs this user chooses to follow. The links can be tracked and messages can be accessed by other users. By exchanging messages or pictures under the same interests, a community is formed. Any other user who is interested in the topic can subscribe and join in. Once the action is succeeded, a link is formed. With different blogs the user chooses to subscribe, different semantics are formed.

We use $S=\langle C,L\rangle$ to represent the network, C is the nodes in the network and L are the links between nodes. Each $l\in L$ is a semantic link between semantic nodes. Semantic nodes are the blogs' names and tags, semantic links are the hyperlinks that each blog has. At discrete times the system will add into numbers of new blogs with different semantics. We collected the multiple interactions with different time stamps $t(i)$. For t_0 to t_1, $S=[t_0,t_1]$ is the classification trees within time stamp t_0 to t_1, and $S=[t_1,t_2]$ is the classification trees within time stamp t_1 to t_2 with new semantic nodes (blogs) added.

In the experiment we collect 1514 bloggers' profile data in entertainment industry and 3017 links in total. Table 1 shows the details of the network.

TABLE 1:
SIZE OF COLLECTED *SINA WEIBO* NETWORK

	nodes	links
media	235	628
music	256	134
director	335	778
producer	167	245
manager	164	867
broker	357	365
total	1514	301 7

The collected network just only a small portion of the whole. Here we focused on the people in the entertainment industry as they appeared more active more than people in other careers. We record each blog the semantics, updating frequency, number of subscribers and the forwarding and replying between them.

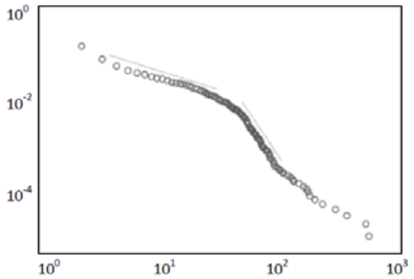
The network is growing every time stamp. First we collect 30 days of data then after that we collect the data for analyze especially on new added subscribers, of their focus of interests (semantics), their own subscriber, and most importantly, of which user or users they subscribe to.

We monitor the network above and collected data discreetly from Jan 10th, 2010 to Feb 21st, 2010. And the observed network has gained 1191 new links and 217 new subscribers. Table 2 is the gained number of blogs and subscribers to each topic.

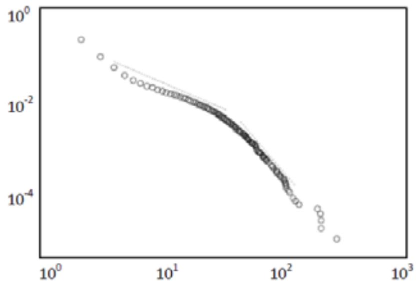
TABLE 2:
NUMBER OF BLOGS OF EACH TOPIC AND THE SUBSCRIBERS OF EACH TOPIC IN TOTAL

People	Blogs	Subscriber
media	36	176
music	27	256
director	56	149
producer	26	275
manager	25	226
broker	46	109
total	217	1191

Then we study the "*fans*" as well as "*followers*" of the *Sina Weibo* nodes.



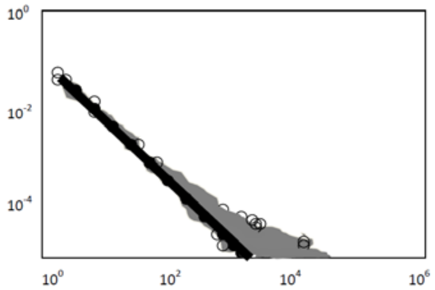
(a) indegree



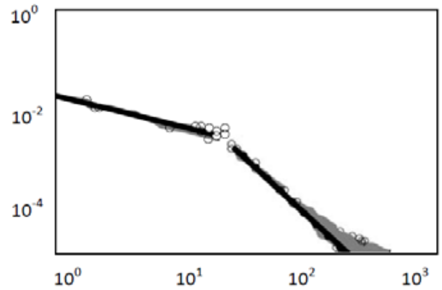
(b) outdegree

Fig.4. Indegree and outdegree of *Sina Weibo*

Fig.4 shows the indegree and outdegree of one node that are not much alike, and the power law distribution of both indegree and outdegree are not straight distribution. This shows the *Sina Weibo* network we sampled is self-similar demonstrates different distribution at different time.



(a) indegree



(b) outdegree

Fig.5. Distribution of in & out degree of network in a 42 days period

From Fig.5, the result shows the overall indegree and outdegree in the network with a period of 42 days. The distribution follows the shape of power law distribution. The indegree of nodes shows saturation of small variables phenomenon. By the method of least squares, the indegree distribution is -1.421. As for the outdegree of nodes, it shows a two-stage distribution, which may be resulted as some emergencies happen in the network.

Result shows the network growth model of *Sina Weibo* network follows the *Matthew Effect*. The nodes already own large amount of links tend to get more nodes and the links are more tempted points to them. The nodes in *Director* category which have more than 176 subscribers gained 83.54% new subscribers while the other nodes gained more than 16.7%. The more links the node has, the more accession it will tend to get.

This kind of nodes we call *popular* node, have a very large amount number of *"follow"* than the average. And as the results show, the chance of a *popular* node getting a new link from other nodes is much higher than that of an average node. And the new nodes are apt to establish a link first to the *popular* nodes in the network than to an ordinary one. We also found that of all the links (*"follow"*) of a certain node (blog), the node which will get more new links is the one that owns a relatively more links itself.

Findings show the network growth model of an instant messaging network such as *Sina Weibo* is much like that of the social network. The nodes with more links of its own are apt to getting more links. The growth of network follows the *Matthew effect*, "the richer get richer", which is *preferential attachment* in Barabási and Albert model [12].

The results also demonstrate some diversity other than similarity with that of social network:

- Only the nodes that with proper number of semantics have the highest probability of getting new links. The nodes with too much semantics have less chance establishing a link; this phenomenon happens everywhere in every life, the growing of a semantic will not last forever. At some point at some time, the increment reaches a certain level, then growth will stop and decrement is ready taking place. When a node happens to carry too much semantic all in itself, the probability the link formation between this node and others will be less contrary to traditional hypothesis. In this case only the blogs with proper number of semantics will get the largest amount of subscribers.

- The outdegree of nodes in *Sina Weibo* network also show segmental characteristic, which may resulted in the existence of "star node" and recommend mechanism.
- Overall the Network follows the Matthew Effect. The indegree and outdegree follow the power law distribution. This suggests that the network may have the *"scale free"* characteristic of complex network.
- Random events effect the growth of network. A number of new communities might emerge and become very active in a very short time yet they also die out very quickly, even more quickly than that they emerge.

As other cyber networks or real life networks that exist, the networks are the results of interactions of human users. The time series of same behaviors' can have diverse even opposite results as the semantics are changed. The real life events affect the growth of network and cause different topologies. For example the scandal of famous investor XUE Manzi cause the astonishing fierce objection against him and the blogs related to him gained greatest number of postings in a very short time.

However, the results also show that the *Matthew effect* is not the only reason contributes to the network evolution. Big events in real life and random occurrences affect the expansion as well. For example, nodes on Entertain and Travel increase a lot as Vancouver Winter Olympic was held on Feb 13th. Random events in real life can cause great access number and links in a very short time and this will affect the network growth model, too.

V. Conclusions and Future Work

Online instant messaging social networks such as *Twitter, Sina Weibo* are now the most prevailing tools in people's social lives. These tools affect us both in cyber space as well as in real interaction when we with others. Yet the network growth model of semantic based network is less discussed. This paper studies the network growth in a dynamic network scenario, analyses the mechanism in which the network evolves, at the same time provides a deeper understanding of the online human interaction. We analyzed the network evolution based on *Sina Weibo* blogs. We abstract the blogs into semantic nodes and the links are also with semantics. We set time stamps to observe the changes. Results show the network growth model follows the *Matthew effect* but also shaped by real life events and sometimes random events.

Human interactions are always been affected by the surroundings and not always show a constant stability. The samples we studies are just part of reflections in real human society. There must be some other issues also making sense but yet haven't been studied. How to approach the network growth model with more precision is still an open question.

Acknowledgment

I want to thank XIAO Yacheng for his kindly advice. This work was supported in part by supported by "the Fundamental Research Funds for the Central Universities" grant XDJK2014B013.

References

[1] H. Zhuge, "Communities and emerging semantics in semantic link network: discovery and learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 6, pp. 785-799, 2009.

[2] M. E. J. Newman, "The structure and function of complex networks", *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[3] H.Zhuge and X.Li, "Peer-to-Peer in metric space and semantic space", *IEEE Transactions on Knowledge and Data Engineering*, vol.19, no. 6, pp. 759-771, 2007.

[4] H.Zhuge, "Interactive Semantics", *Artificial Intelligence*, vol.174, pp. 190-204, 2010.

[5] H. A. Simon, "On a class of skew distribution functions", *Biometrika*, pp. 425-440, 1955.

[6] S. Bornholdtand and H. Ebel, "World Wide Web scaling exponent from Simon's 1955 model", *Physics Review E*, vol. 64, no. 3, 2001.

[7] P. L. Krapivsky and S. Redner, "Organization of growing random networks", *Physics Review E*, vol. 63, no. 6, 2001.

[8] H.Zhuge, Y.Xing and P.Shi, "Resource Space Model, OWL and Database: Mapping and Integration", *ACM Transactions on Internet Technology*, 8/4, 2008.

[9] H.Zhuge and B.Xu, Basic Operations, Completeness and Dynamicity of Cyber Physical Socio Semantic Link Network CPSocio-SLN, *Concurrency and Computation: Practice and Experience*, DOI:10.1002/cpe.1623.

[10] L. A. Adamic and B. A. Huberman, "Power-law distribution of the world wide web", *Science*, vol. 287, no. 5461, pp. 2115, 2000.

[11] A. Capocci, V. D. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, "Preferential attachment in the growth of social networks: the internet encyclopedia Wikipedia", *Physics Review E*, vol. 74, no. 3, 2006.

[12] P. Mika, "Social networks and the semantic web", WI '04: *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 285-291, 2004.

[13] H.Zhuge and J.Zhang, "Topological Centrality and Its e-Science Applications", *Journal of the American Society for Information Science and Technology*, DOI: 10.1002/asi.21353.

[14] L. A. Adamic, "The small world web", ECDL '99: *Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*, pp. 443-452, 1999.

[15] M. E. J. Newman, "Clustering and preferential attachment in growing networks", *Physics Review E*, vol. 64, no. 2, 2001.

[16] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality", *Physics Review E*, vol. 64, no. 1, 2001.

[17] A. L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science,* vol. 286. no. 5439, pp. 509 - 512, 1999.

[18] A. L. Barabási, "Linked: the new science of networks", *Perseus*, 2002.

[19] Zhixing Huang and Yuhui Qiu, "A multiple-perspective approach to constructing and aggregating citation semantic

link network", *Future Generation Computer Systems*, vol. 26, no. 3, pp. 400-307, 2010.
[20] Zhixing Huang, Yan Yan, Yuhui Qiu and Shuqiong Qiao, "Exploring emergent semantic communities from DBLP bibliography database", ASONAM '09: *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, pp. 219-224, 2009.
[21] X.Shi and H.Zhuge, "Cyber Physical Socio Ecology", Concurrency and Computation: *Practice and Experience*, accepted.
[22] L. Getoor and C. P. Diehl, "Link mining: A survey", *ACM SIGKDD Explorations Newsletter*, vol. 7, no. 2, pp. 3-12, 2005.
[23] J. Scott, "Social network analysis: A handbook", *Sage Publications Ltd, 2nd edition*, 2000.
[24] A. Mislove, M. Marcon, etc, "Measurement and analysis of online social networks", IMC '07: *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement,* pp. 29-42, 2007.
[25] H.Zhuge, X.Sun, et al, "A Scalable P2P Platform for the Knowledge Grid", *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 12, pp. 1721-1736, 2005.
[26] H.Zhuge, "Socio-Natural Thought Semantic Link Network: A Method of Semantic Networking in the Cyber Physical Society", *24th IEEE International Conference on Advanced Information Networking and Applications*, 2010.
[27] H.Zhuge and J.Zhang, Automatically Discovering Semantic Link Network on Documents, Concurrency and Computation: *Practice and Experience*, DOI: 10.1002/cpe.1624.

Wei Ren is a associate professor at College of Computer & Information Science, Southwest University, Chongqing. She is a member of Chongqing Computer Association, and also the member of Chongqing Youth Federation for Science & Technology. Her major research interests include complex network and online social network study.

Yuhui Qiu, Professor at College of Computer & Information Science, Southwest University, Chongqing. His contributions are mainly on SLN and social networks.