Human Flesh Search Model
Incorporating Network Expansion and GOSSIP with Feedback

Bing Wang, Bonan Hou, Yiping Yao, Laibin Yan
School of Computer Science
National University of Defense Technology
Changsha, P.R. CHINA
{wangbing, bnhou, ypyao, lbyan}@nudt.edu.cn

***Abstract*—With the development of on-line forum technology and the pervasive participation of the public, the Human Flesh Search is becoming an arising phenomenon which makes a great impact on our daily life. There arose big research interests in social, legal issues resulted from HFS, however, very little work has been conducted to understand how it comes into being and how it dynamically evolves. This paper proposes a modeling and simulation approach incorporating network expansion and GOSSIP propagation with feedback for a better understanding of the human flesh search phenomenon. Based on the acquisition and analysis of the netizens' surfing behavior data, the evolution of the HFS is modeled as a network growth process with proper dynamic input, which is characterized by heavy-tail and burst-oriented distribution, modeling as a Weibulloid process. Then, an improved GOSSIP model with feedback is proposed to represent the information propagation, processing and aggregation during the HFS. New insights for HFS are gained through a set of simulation experiments.**

Keywords-human flesh search(HFS); social network; comeplex network; distributed problem solving; crowdsourcing; GOSSIP with feedback

I. INTRODUCTION

The uncovering of fake pictures of "South China Tiger" and series of events happened during the past two years show the great power of crowds of netizens online. These eye-catching phenomenons, rising recently over the Internet, are called Human flesh searching (HFS), which is literally translated from the Chinese phrase 'RenRouSouSuo' [3]. Human flesh search is a mass campaign, which comes into vogue through the medium of internet, targeting at searching for the identity of a certain person or the truth about a certain event, whose data collection depends partially on the human force to filtering the information gained from the search engine , and partially on the anonymous or real-name information announcement, "Human flesh" illustrates that the role of human force played in the search and the distinction between itself and the traditional machine-based search engine [4]. With the advent of Web 2.0 technology and the online community, everyone can publish, share, review, and comment at his/her wish. An interesting or controversial topic would always attract a great number of online participants gathering together to dig the "truth" out. This process usually involves hundreds and thousands of netizens and occurs in a relatively short time, e.g., in several hours or a couple of days. Surprisingly, in most cases, the answer they pursue is accurately found out, no matter how hazy the problem seems to be.

Although HFS has attracted numerous controversies, and even triggered legal dilemmas between privacy violation and public opinion, it has already become a phenomenon of profound impact. It is necessary to acquire better understanding of its structure and evolving process. However, the HFS is very complex due to the intrinsic nature of the phenomena itself. Despite a great interest has been paid on the HFS phenomena, little knowledge on how it arise, how people involved are organized and how to describe the evolution process have already been gained.

Several researches have been conducted on the Web 2.0 related phenomenon, such as blogosphere [9], crowdsourcing [20][22], Guanxi [15], to name a few. Complex Network theory has been adopted as a powerful approach to study these systems that are best described as networks with complex topology. Different variants of complex network have been proposed. Traditionally, random network, where the vertices are randomly connected, was widely used, but the predictions were rarely justified in the real world. Small-world models were introduced by Watts and Strogatz [11] to describe the six degree of separation. The scale-free (SF) networks [2], proposed by Barabási and Albert, are regarded to be the best suitable models for large amount of systems from the Internet to social networks. SF networks are characterized by the vertex connectivity following a power-law distribution.

These approaches could be utilized as an infrastructure of the HFS phenomenon, but great challenges are still in the way towards a better understanding of the phenomena. First, it's still hard to define and classify the HFS. As a novel and arising phenomena, the HFS overlaps multi-principles including Crowdsourcing, Social Search [18], Distributed Problem Solving, and collective intelligence [17]; second, some characteristics of HFS, such as the vast number of participants and the dramatic evolving process, bring difficulties to a well explanation through traditional social research methods. For these reasons, quantitative analyse through modeling and simulation of the phenomena and reproducing the dynamic process, is a promising approach to acquiring new insights.

In this paper, we propose a network expansion model to describe the HFS, focusing on the growth and evolution of the network topology formed by the posted messages.

Further, we improve the GOSSIP model with feedback by taking the propagation, the processing and the aggregation of the information into consideration.

The remainder of this paper is structured as follows. In section II, we explain the background and related works, especially the HFS phenomena and the participation pattern of the netizens in HFS. Section III proposes a network expansion model of HFS process with proper dynamics input. Section IV describes an improved GOSSIP model with information feedback. Section V discusses the results of the HFS simulation. Finally, conclusions and future works are described in section VI.

II. Background and Related Work

With the rapid prevalence of Internet, the number of netizens is rapidly increasing. According to the "24th Statistical Survey Report on the Internet Development in China", by the end of June of 2009, the amount of Internet users in China had reached 338 million, with the increase rate of 32.1% during the past half year [19]. Additionally, online Forums, BBS, Blogs and such social network sites (SNS) facilitate people to express their thoughts, voice their opinions, and share their experiences and ideas. These media are easily and widely accessed by the public, even more, every netizen could be both information provider and reader. The content created by one netizen could be seen by others immediately, which may intrigue further reactions. The interactions among huge amounts of people have emerged "collective wisdom" or crowdsourcing which acts as the storehouse of overwhelming amounts of knowledge about the topic. All these conditions fertilize the rising-up of HFS. Figure 1 shows the popularity of the HFS in China in recent years. The data reflects the search amount of HFS worldwide through Google, the keyword is typed in Chinese and the result is gained from the Google trend analysis (http://www.google.cn/trends).

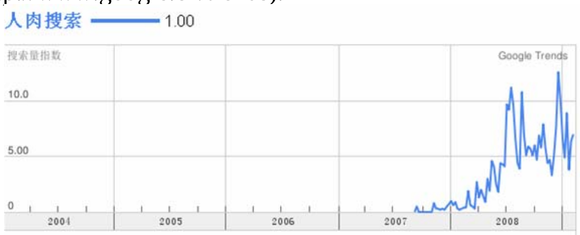


Figure 1. The google trends for "RenRouSouSuo"(HFS).

It is still hard to give the definition of HFS. As a Web 2.0 related phenomena, however, HFS has similar characteristics with crowdsourcing, social search, collective intelligence, and distributed problem solving. HFS is a particular form of crowdsourcing, outsourcing a task to an undefined, generally large group of people or community in the form of an open call [22]. Also it can be treated as a distributed problem solving paradigm which takes advantage of numerous voluntary contributions to acquire the right answer. HFS can also be classified into social search in which users utilize social interactions with others in the search[18].

The HFS dynamics is jointly determined by the netizens' Internet information seeking habit and the popularity of social network sites. The typical process of human flesh search is described below:

First, someone publishes an open question, or part of a question, with some information or clue, such as pictures or videos on the Blog or online forums, the content of the question may be of interest or controversy, clear or opaque.

Then, the online netizens respond while browsing these contents. Some volunteers would transship it to several other large social network websites (tianya.cn, mop.com, *etc*.) and would issue an open call to dig the truth out, which in turn intrigues more outraged reaction. Many anonymous participants would bid against the topic and several key clues to the issue were gathered and analyzed. Although most of the topics on the internet will vanish, several hot topics maintain a measureless social influence.

The characteristics of HFS could be summarized as follows:

- **Accessibility**: the online media can be easily reached by all, and the content can be shared among numerous netizens immediately after it was created. The content may be transshipped among several mainstream social network sites by volunteers.
- **Popularization**: the problem of HFS, such as looking for the right person through some pictures or other ambiguous clues, is often not too hard to need creative talent and can be easily solved once enough attention and therefore enough information was gathered according to the small-world phenomenon[11]. Analysis results show that a large percent of responds do not contribute to the problem directly, but just similar to the mutation in Biology or genetic algorithms, they are not absolutely useless. Rather, in some degree, the scale of the participant population plays an important role on whether the problem could be solved or not in tolerable time.
- **Centerlessness**: there is no central control on the topic in the process of HFS and some digressions are unavoidable.
- **Information Timeliness**: the timeliness is vital to the HFS life cycle. The topic can easily attract public attention, but also suffer quickly outdating.
- **Convergence**: Though under real-time constraint crowd perform worse than single-person decision, providing enough time and continuous attention, the HFS will somehow achieve the goal.

The HFS is the collaboration among the online mass towards a common goal. Underneath the bustle surface, there is an active virtual social network containing many individuals with frequently interactions.

Some related works have been conducted to study the online community activities. Yongqin Gao, *et al*.[8] focused on the Open Source Software developers and project network and proposed a modified scale-free network modeling approach. Nitin Agarwal, *et al*. [9] reviewed some key elements of research in Blogosphere. Other attractive researches includes social network, communication pattern [10], information propagation [12], online activities [7], and so on.

In the current literatures, there is still not much work devoted to the research of HFS phenomena with quantitative modeling and simulation techniques. Due to the intrinsic complexity, a moderate abstract and intuitionistic perspective will facilitate better understanding. In this work we combined the network forming up with information propagation, and thus propose a prototype HFS model to provide new insights into this rising phenomenon.

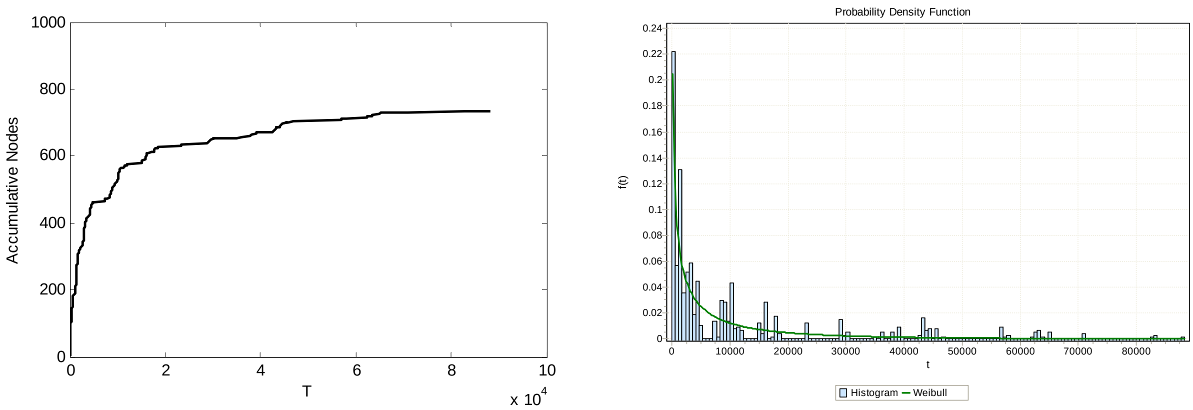
III. THE HFS PROCESS AS NETWORK EXPANSION

In this section, we start by collecting data from several HFS topics on some famous social network sites. Based on the analysis of the data, the activity pattern distribution of the participation in a HFS is shown.

A. Data Collection

Social network sites are selected as data sources for the research of HFS. Utilizing web crawling techniques, we collected a random set of hot HFS topics since 01/03/09 from online forums at www.tianya.cn, one of the largest online communities in China. Each topic contains a list of the originate message, replies, and iterative replies to replies. The dataset of each reply including the author ID, the date, and the text content are recorded for further analysis.

The statistics of these topics indicate that there exist differences in their shape scale and lifetime, but sharing a common evolving trend. Right after the topic is created, the number of responds increasing quickly. As time goes across the turning point, the growth slows down. Figure 1(a) depicts the accumulative participants during the life cycle of a typical hot topic. The according histogram of responds in each unit of time is shown in (b), indicating the burst-oriented and heavy-tail nature of HFS dynamic patterns.



(a) (b)

Figure 2. (a) The accumulative number of responds to a typical HFS evolving along with time. (b) The HFS's histogram of activity pattern and its probability density function with Weibull distribution fitness(*a*= 0.60122, *b*= 4.5126).

B. Input Modeling

As an essential part of the HFS network model, proper input modeling is a precondition for validity [16]. During the model construction, a new node is created and added to the network at each step. In literature, it is often assumed that the arrival of new nodes is a Poisson process [14], where the probability of a number of replies events occurring with a fix average rate and independently of the time since the last event. After an investigation on the topic posted and replied, the statistical results show that the probability of new replies and how many replies are strong related with the time eclipsed. As the burst-oriented nature of human behavior [5], the time intervals between consecutive online activities follow heavy-tailed process which decays slowly, in contrast to the exponential expected for Poisson processes. Further, Dezso, *et al*. [7]analyzed the dynamic access on web portals, found both the popularity and accessibility are equally important, and assert the news document visitation behavior is best approximated by a power law, $n(t) \sim t^{-\beta}$ with $\beta = 0.3 \pm 0.1$.

Since the news portals are different from most of the online forums, in which the most recently updated (either through create or reply) are always aligned as the first of topic list on the front page, while the deprecated topics will go down according to their rank. Yet, it's necessary to identify the distribution with time intervals between consecutive updates.

Previous observation in Figure 2 (b) has shown that the dynamic accesses quickly burst at the beginning, then slowly decay, heavy-tailed process, and the number of activities (n) in unit time is better approximated by the Weibulloid process in (1),

$$n(t) \sim \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} e^{-(t/\beta)^{\alpha}} \qquad (1)$$

Where α, β are related with the characteristics of the concrete topic, for example, in Figure 2, *a* equals 0.60122, and *b* equals 4.5126.

C. *Assumptions and HFS Network Model*

A wide range of social systems share some properties of complex networks. Traditionally, random network of ER model were widely used to describe these complex systems, which assumes the vertices are randomly connected with probability *p*. After investigating the topology of several large networks, as WWW and citation patterns, Barabási, *et al*. [2] found they are better described as the scale-free network, in which the vertex degrees follow a scale-free power-law distribution, which is $p(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$, in contrast to the exponential distribution expected in random networks. Afterwards, the instant messaging, email web subsequently have been empirically approved as scale-free network [6].

We model the participation and their interaction in the HFS as network, in which nodes represent user participation and links represent the interactions occurred between nodes. Once a message is posted, replies would be inspired and attached to the message. Even more, the reply may be followed by further replies, because the reply itself is also a message. As the HFS goes on, the network expands as new messages are added and new links formed.

For the first step, we ignore the sophisticated semantics among the textual messages. Instead, these messages are simply classified as approval, disapproval, and critical clue. Additionally, we apply James Surowiecki's asserts, in his book *Wisdom of the Crowds*, which declares that providing that the process is sound, the more people involve in solving a problem, the better the result will be [17]. Following this philosophy, we take the total number of participants and the maximum degree as the measurements for the successful accomplishment of a HFS. That is to say, in many situations, the more people involve in the HFS and the more convergent the participations are, the higher probability it will finally be solved.

The next question is how to construct the network in order to simulate the process of the HFS evolving process. There are several mature paradigms could serve as the skeleton, in many cases, the relationship between different participation shows scale-free properties. The Barabási–Albert (BA) [2] model gives an algorithm for generating random scale-free network using a preferential attachment mechanism. The process is shown as follows:

1) First, the network begins with an initial network of m_0 nodes and the degree of each node should be at least 1.

2) Second, at each time step, a new node is created and is connected to existing nodes according to the principle of preferential attachment, whereby each new node is connected to *m* of the existing with a probability that is proportional to the degree of the existing node. The probability p_i is given by the following formula:

$$p_i = k_i / \sum_j k_j$$, where k_i is the degree of node *i*.

The BA model was modified in our simulation to build the HFS network. Based on the observation that new nodes are not always added at constant frequency in reality, modification is made at the step 2 to capture the evolving nature of the HFS network.

IV. Improvement through Gossip with Feedback

The network based HFS model built in previous section has not taken the information propagation in actual situation into consideration. In the procedure of HFS, along with the expansion of the participation, the information propagation takes place. Being different from the original Gossip paradigm, the nodes in the HFS network not only carries out the propagation, but also contributes to the search progress, and provides feedback to predecessors. Based on a Gossip-like expansion of the participation network, the information processing, content contribution and feedback are introduced to enhance the credibility of the model.

In the improved model with Gossip, each node in the network is modeled as an autonomous agent with the capability of information processing. When new nodes are added to the network, they will absorb the information flowing from their neighbors. The information will be assimilated or may be mutated as the influence of different beliefs hold in each agent. Then, the most important is that the agents can provide feedback which is vital to the goal of HFS. The information will provide feedback in the cascade form until it reaches the hub. The information aggregation occurred as the result of cascade feedback. Once enough information is fed back and properly aggregated, we can say the problem of HFS is solved. The process of modified gossip model is illustrated by Figure 3. Before we get into the feedback gossip algorithm, we need to formally define the problem/solution, information propagation, mutation, and aggregation.

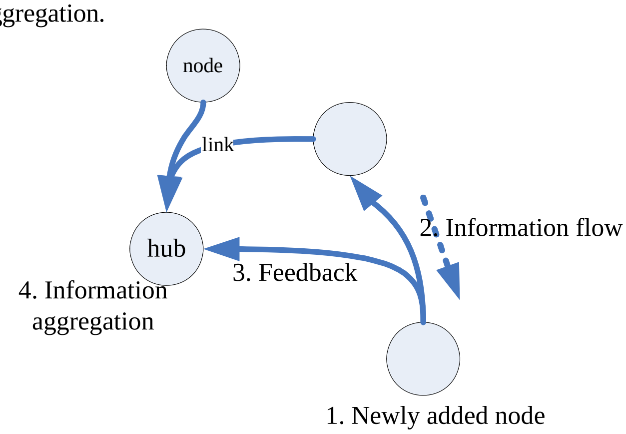


Figure 3. The cycle of information flow, feedback and aggregation for the HFS network.

The Problem under HFS is defined as an n-dimension vector, with each dimension as one factor in the Problem Space S_n , illustrated by the following formula. If certain dimensions are not clear yet, e is marked.

$$Q = [q_1, q_2, q_3, \cdots, q_n], q_i \in N \cup \{e\}$$
$$Q \in S_n \qquad (2)$$
$$S_n = \underbrace{N \cup \{e\} \times N \cup \{e\} \times \cdots \times N \cup \{e\}}_{n}$$

Then, we define the information propagation as:

$$P_j : \bigtimes_{1}^{m} S_n \rightarrow S_n, m = \# Set^{j}_{neighbour} \qquad (3)$$

Propagation happens when a new node is added to the network. The newly added node gains a new Problem instance according to the Problem instances in the neighbor nodes.

Once the agent got the problem, it will contribute to the problem solving progress with its own knowledge through the mutation process in (4)

$$M_j : S_n \rightarrow S_n \qquad (4)$$

Where M_j works as bellows,

1) Select Set_k s.t. $\# set_k = k, (k \in N) \cap (k \in [0, n])$ randomly.

2) For each q_i in set_k, $q_i = f(q_i)$, the mutation function $f : N \cup \{e\} \rightarrow N \cup \{e\}$.

3) Select a subset of the n dimensions Set_k.

4) For each dimension selected, utilize function f to reflect the contribution from a node in the Gossip-based information propagation.

As the information is fed back cascadingly, it will result in the problem instance ready for the aggregation on the predecessor's side. The information aggregation is defined in (5).

$$Aggregation_j : \bigtimes_{1}^{m} S_n \times S_n \rightarrow S_n, m = \# Set^{j}_{neighbour} \qquad (5)$$

When the number of solved problem dimensions has reached certain threshold, it can be declared that the HFS is convergent.

The modified GOSSIP model with information feedback is shown in pseudo code 1.

Algorithm 1. GOSSIP protocol with information feedback

```
while(network expanding){
   add a new node to the network;
   information propagation from neighbors;
   information processing and mutation;
   for each neighbor, trigger feedback:
      neighbor.aggregation;
      neighbor.Feedback(cascadingly);
   end
}
```

V. Simulation Experiments and Results

We have implemented a prototype simulation integrating the above model and conducted experiments with different parameters value as a preliminary demonstration. JUNG network modeling library was utilized to support visualization. The results reveal the structure of the evolving network, highlighting connectivity, clustering and strengths of relationships among user participations. Animated output allows viewers to see the evolution of the social network over time.

Figure 5 illustrates the simulation of the network structure during the evolving process of a typical HFS topic. At the initialization of the simulation, a new message is created as an initiator. After a relatively short period, a number of netizens are attracted and participated into the action by reviewing, commenting and copy the topic to another website. The information flows and feedbacks among these nodes. Also, the information is accumulated and aggregated during the expansion process. As time advances, more and more participants are intrigued to join in. Additionally, some nodes with large degree formed the hub of information aggregation. The increasing number of participants follows a decaying and long-tailed process. Panel (f) indicates the accumulative nodes and maximum degrees in the network.

With the formula (2) as dynamic input to the model, we got results with different pair of parameter values, as illustrated in figure 4. The solid lines indicate the accumulative participations occurred in the corresponding simulated HFS event, while dash lines mean the maximum degree the nodes in the network have. The relations between the line shape and parameters could be utilized to better identify the characteristics of HFS. Also, it is obvious that the HFS will converge from a tripping point as the number of participation increasing. The large accumulative participation is important on the convergence of the HFS, although a relatively small percent really contributing to the problem.

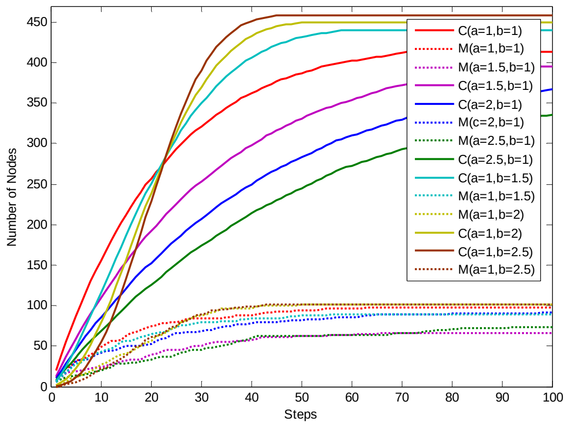
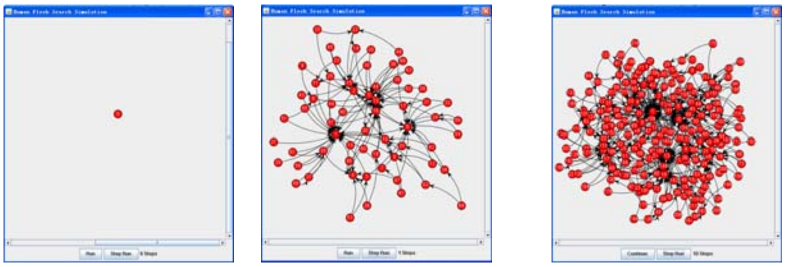
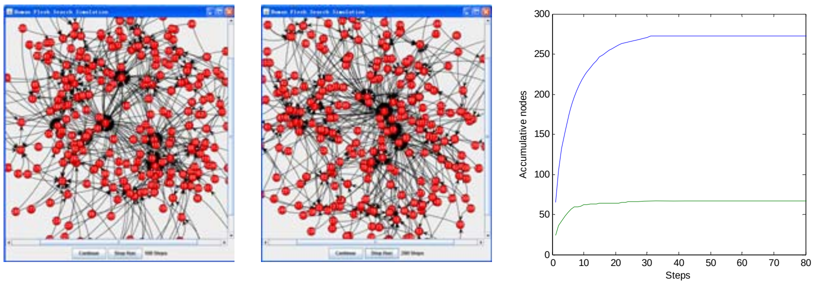


Figure 4. Different pair (a,b) of input parameter values and their corresponding accumulative nodes (real line) and maximum degree (dash line) evloving during the network expansion.



(a) 0 step (b) 1 step (c) 10 step



(d) 100 step (e) 200 step (f) Accumulative nodes

Figure 5. The Evolving Process of the HFS Network (*a*=0.5, *b*=2.0)

VI. CONCLUTION AND FUTURE WORK

The recently arising HFS phenomenon attracts attentions from sociologists, psychologists, legalists, and computer scientists. It poses great challenges as its complex structure and dramatic evolving process, and there is still little research has been devoted to studying how it comes into being and dynamically evolves. This study demonstrated a modeling and simulation approach to a detailed, quantitative understanding of the HFS. Based on the empirically investigation, the HFS model was built incorporating network expansion and GOSSIP with feedback. The netizen participation dynamics is properly modeled. It quickly bursts at the beginning, and then slowly decays as a heavy-tailed process. Additionally, a modified GOSSIP diffusion algorithm with feedback was proposed to better describe the information flow, which forms the HFS problem solving.

This paper facilitates the gain of new insight from the HFS phenomena. Still more problems are left wanting for further investigation. For example, the scale-free property of the HFS needs further verification; and the psychological motivation of the participants during a HFS process also deserved more attention.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (NSF) Grant (No.60773019, No. 60873120) and the Ph.D. Programs Foundation of Ministry of Education of China (No. 200899980004).

REFERENCES

[1] Gabriele D'Angelo, Ste-fano Ferretti. "Simulation of Scale-Free Networks," SIMUTools '09, Rome, Italy, 2009.

[2] A.-L. Barabási and R. Albert. "Emergence of scaling in random networks". Science, 286, 1999.

[3] Times ONLINE. Human flesh search engines: Chinese vigilantes that hunt victims on the web. [Online]. Available: http://technology.timesonline.co.uk/ tol/news/tech_and_web/article4213681.ece.2008

[4] RenRouSouSuo. 2008. [Online]. Available: http://zh.wikipedia.org/wiki/%E4%BA%BA%E8%82%89% E6%90%9C%E7%B4%A2.(in Chinese).

[5] A.-L. Barabási, "The origin of bursts and heavy tails in human dynamics," May 2005. [Online]. Available: http://arxiv.org/abs/cond-mat/0505371.

[6] Re´ka Albert* and Albert-La´szlo´Barabási. Statistical mechanics of complex networks. REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002.

[7] Z. Dezsö, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A. L. Barabási, "Dynamics of information access on the web," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 73, no. 6, pp. 066 132+, 2006.

[8] Y. Gao and G. Madey, "Towards understanding: a study of the sourceforge.net community using modeling and simulation," in *SpringSim '07: Proceedings of the 2007 spring simulation multiconference*. San Diego, CA, USA: Society for Computer Simulation International, 2007, pp. 145-150. [Online]. Available: http://portal.acm.org/citation.cfm?id=1404703

[9] N. Agarwal and H. Liu, "Blogosphere: research issues, tools, and applications," *SIGKDD Explor. Newsl.*, vol. 10, no. 1, pp. 18-31, 2008

[10] H. Wei, H. Xiao-Pu, Z. Tao, and W. Bing-Hong, "Heavy-tailed statistics in short-message communication," *Chinese Physics Letters*,

vol. 26, no. 2, pp. 028 902+, 2009. [Online]. Available: http://dx.doi.org/10.1088/0256-307X/26/2/028902.

[11] D. J. Watts and S. H. Strogatz, "Collective dynamics of `small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440-442, June 1998. [Online]. Available: http://dx.doi.org/10.1038/30918.

[12] Y. Fernandess and D. Malkhi, "On spreading recommendations via social gossip," in *SPAA '08: Proceedings of the twentieth annual symposium on Parallelism in algorithms and architectures.* New York, NY, USA: ACM, 2008, pp. 91-97. [Online]. Available: http://dx.doi.org/10.1145/1378533.1378547

[13] Reka Albert and Albert-L´aszl´o Barabási. Statistical mechanics of complex networks. Review of Modern Physics, 74(1):47–97, 2002.

[14] E. M. O'Grady, M. Rouleau, M. Tsvetovat. Network Fracture: How Conflict Cascades Regulate Network Density. *Agent 2007*, Chicago, IL, 2007.

[15] Barry Wellman, with Wenhong Chen and Dong Weizhen. Networking Guanxi. Social Networks in China: Institutions, Culture, and the Changing Nature of Guanxi. Cambridge University Press 2001.

[16] Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. "Discrete-event system simulation". 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc. 2005.

[17] Satnam Alag. Collective Intelligence in Action. Manning Publications Co. 2009.

[18] Brynn M. Evans, Ed H. Chi. Towards a Model of Understanding Social Search. CSCW'08, San Diego, California, USA. November 8–12, 2008,

[19] China Internet Network Information Center (CNNIC). 24rd Statistical Survey Report on the Internet Development in China. June. 2009. [Online]. Available:http://research.cnnic.cn/img/h000/h11/attach200907161306340.pdf.

[20] J. Howe, "Wired 14.06: The rise of crowdsourcing." [Online]. Available: http://www.wired.com/wired/archive/14.06/crowds.htm

[21] L.-H. Liu, F. Fu and L. Wang. Information propagation and collective consensus in blogosphere: a game-theoretical approach. eprint arXiv:physics/0701316.

[22] *Crowdsourcing*, online, August 2009. [Online]. Available: http://en.wikipedia.org/wiki/Crowdsourcing