A novel steady-state maintenance simulation framework for multi-information disseminations in crowd network

Zhong Wang, Hongbo Sun and Baode Fan
School of Computer and Control Engineering, Yantai University, Yantai, China

Received 13 February 2020
Revised 23 March 2020
Accepted 24 March 2020

Abstract

Purpose – The era of crowd network is coming and the research of its steady-state is of great importance. This paper aims to establish a crowd network simulation platform and maintaining the relative stability of multi-source dissemination systems.

Design/methodology/approach – With this simulation platform, this paper studies the characteristics of "emergence," monitors the state of the system and according to the fixed point judges the system of steady-state conditions, then uses three control conditions and control methods to control the system status to acquire general rules for maintain the stability of multi-source information dissemination systems.

Findings – This paper establishes a novel steady-state maintenance simulation framework. It will be useful for achieving controllability to the evolution of information dissemination and simulating the effectiveness of control conditions for multi-source information dissemination systems.

Originality/value – This paper will help researchers to solve problems of public opinion control in multi-source information dissemination in crowd network.

Keywords Emergence, Crowd network, Multi-source information dissemination, Steady-state

Paper type Technical paper

1. Introduction

The phenomenon of crowd intelligence (Chai *et al.*, 2017) can be easily observed in our society, such as "two heads are better than one" and "everybody's business is nobody's business." With the rapid development of information technology, crowd intelligence is more extensive and complex. People, enterprises, governments, equipment and articles become more and more smart with the ever expanding of intelligent depth, intelligent width and connection scale to form various crowd network systems (such as e-commerce platform, networked manufacturing supply chain, Wikipedia and network election).

In a crowd network system, intelligent individuals and consciousness are all mapped into information space to represent their selves in the form of digital-selves.

This work is supported by National Key R&D Program of China (Grant No. 2017YFB1400105).

International Journal of Crowd Science
Vol. 4 No. 3, 2020
pp. 273-282
Emerald Publishing Limited
2398-7294
DOI 10.1108/IJCS-02-2020-0003

Crowd science is constantly developing. More and more social phenomena can be searched for results by the crowd network. To maintain the effectiveness and stability of economic, social and government governance activities and avoid confusion, turbulence and mutation as much as possible, public opinion control of crowd network system becomes very important.

This paper aims to using behavior and method of controlling system steady-state, establishing a novel steady-state maintenance simulation framework and simulating stable operation under crowd network multi-source information system. To reducing the negative events or destructive change, track system status, monitor the "emergence" (Holland, 1998) in the system and identify the slow parameters of the system change with the depth, width, interconnection level and total resources, analyze the change, keep the system in a steady-state condition by fine-tune the monitoring level, spread speed. Finally, according to feedback of the simulation results, extract changes in dynamic simulation parameters and analyze simulation scale. Build stable and efficient crowd network multi-source information dissemination system in the future.

2. Related work

Simulation is an important way to study public opinion control in crowd science multi-source information system. In this process, a huge amount of information data are produced, and illegal information is inevitable in these data, such as rumors. These impact on the country and society stability cannot be underestimated, so this paper cannot guide such public opinion on the real network, that is, simulation has become an important method to study public opinion control. Information dissemination refers to the initial communicator spread information to other groups. The current social network is interactivity, timeliness and suddenness, its interaction process is the information dissemination. How to judge steady-state from data and to mining potential water army based on public opinion, so that the social network keep stable state have great significance. Therefore, the issue of govern water army in information dissemination has become the hot research by many domestic and foreign scholars.

At present, a large amount of research focuses on analyzing information content characteristics, analyzing users behavior characteristics and analyzing the network topology central indicators. The information content of water army has certain identifiable characteristics different from ordinary users. This direction focuses on analyzing the content of information by natural language processing and other techniques, methods to searching for the water army, such as reuse detection and text classification. Naud and Usui (2007) and others believe that in the process of network water army supervision, used text topic classification based on clustering algorithm to content briefing and trend analysis. Finally, achieving warning, guidance and monitoring of public opinion. Zhao *et al.* (2013) proposed to analyze the relevance of spammers' content and proposed a text reuse detection model and designed a corresponding filtering algorithm for spammers. Jindal and Liu (2008) first proposed the concept of "fake comments" in the e-commerce field, including fake comments, biased, unfair comments and irrelevant comments. Jindal proposed artificially labeled fake comments and classified the features in the fake comments to identification. This method has a very high recognition accuracy for false comments with high content repetition rate and high similarity, but information features are hidden is not identified.

The research on analyzing the user behavior characteristics has also achieved great results. The water army has the characteristics of concentration, high explosiveness and suddenness. Mukherjee *et al.* (2012) and others established a Bayes recognition model by studying water army suspiciousness. Lim *et al.* (2010) and others analyzed a large amount of

data from amazon's commodity reviews, summarizing several water army behaviors, such as high repetitiveness of text content, similar comments in same time. For a large structure network, the time spent in screening users will be enormous.

Considering the network topology central indicators refers to the node's attribute information, such as node degree and feature vector. Kitsak *et al.* (2010) proposed K-nuclear decomposition method. It forms a new perspective evaluates influence of nodes in the network and considers nodes location to determine the nodes influence. The water army have huge social influence and select greater influence nodes for attribute screening. This can adapt to large-scale social networks, under the single source information dissemination, this can sort the influence nodes very well and largest K-core node is the most influential node. However, in the case of multi-source information dissemination, this is limited by the dissemination source, which changes the individual's influence judgment and does not achieve a good effect.

Crowd network multi-source information dissemination topology is huge and has many nodes. The node is intelligent individual, its defense and strategy are more intelligent, and its hides obvious characteristics of the water army. The multi-source information dissemination system determines that each node is information originator. Judging steady-state, screening water army and applying the control conditions in time to reach the steady-state is an urgent problem to be solved.

3. Control system framework

3.1 Member mode

3.1.1 Primitive unit. As shown in Figure 1, a primitive unit consists of the following six parts: pattern, affector, decider, monitor, actuator and comparator. The input is direct information source or other members' information, according to actuator summarize each part comprehensive calculation results to determine the dissemination mode.

In Figure 1, pattern is a directed acyclic graph composed of decisions on time series. The arc represents the behavior, and the weight on arc represents the cost of the behavior; the nodes represent the results of the behavior; and the weight of node represents the benefits of the behavior. Affector is several advisers influence the primitive unit for next decision. The decider is comprehensively measures its own resource situation, compares the cost of next possible behavior and provides the executor with the optimal path. The actuator executes decision according to decider and affector and the execution result is also affected by the confidence level. Monitor correct deviations based on targets/commitments. Comparator connects other primitive units related to itself and learns behavioral results of others for next round of selection by negative feedback.

3.1.2 Collective unit. As shown in Figure 2, a collective unit consists of the following eight parts: pattern, affector, decompositor/selector, integrator, decider, monitor, actuator and comparator. The input is the target or commitment, and the output behavior's result.

In Figure 2, decompositor is ability for decompose the targets/commitments, selector selects the bottom primitive units based on decompositor's result. Integrator synthesizes decompositor's results for decider. Decider based on the integrator's execution results and comprehensively measures its own resource situation to make decision for actuator. Other parts is same to Figure 1.

3.1.3 Collective structure. The bottom layer of collective unit is composed of primitive units and its structure is shown in Figure 3.

3.2 Judge steady-state

In the crowd network system, the system state change is caused by preference of the simulation members and corresponding behavioral operations. When the information in different filed, the

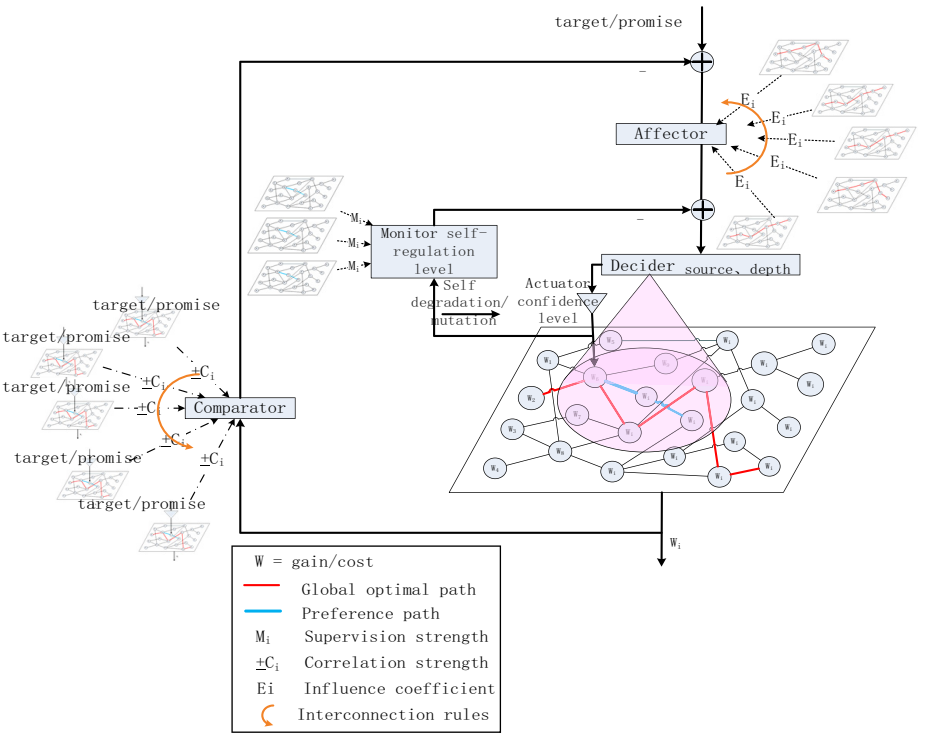

Figure 1.
Primitive unit

types of information are scattered and cannot cause intelligent group's resonance, the system will not in an emerging state. Even if there are some sensitive topics that cause a high attention degree, after the system's self-regulation will also turn into a steady-state.

The definition of system steady-state is mainly to define the preference consistency between members and information. In certain field, if the information points different directions, the members' preference for the information is not concentrated in a certain angle range, it is in a dispersed state. In this field, information vector sum is closer to 0, system state is more stable.

However, when there are two extreme people in different directions in a certain field, even if the finally vector sum may be close to 0, the system is not necessarily in a stable state at this time. Therefore, in the case information vector sum is obtained and is within the normal range, a preference threshold ρ can be set to verify it, and in any field, the vector and absolute within a certain range below the preference threshold ρ, the system can be considered steady-state.

3.3 Detection method

This paper based on establishing multi-source information dissemination model uses a variety of control conditions together to control the system state. By establish a negative feedback control system separately, monitor the system status to detect problems in time and use different control methods for the change of system slow parameters in different states, so that the system returns to the expected state accurately.

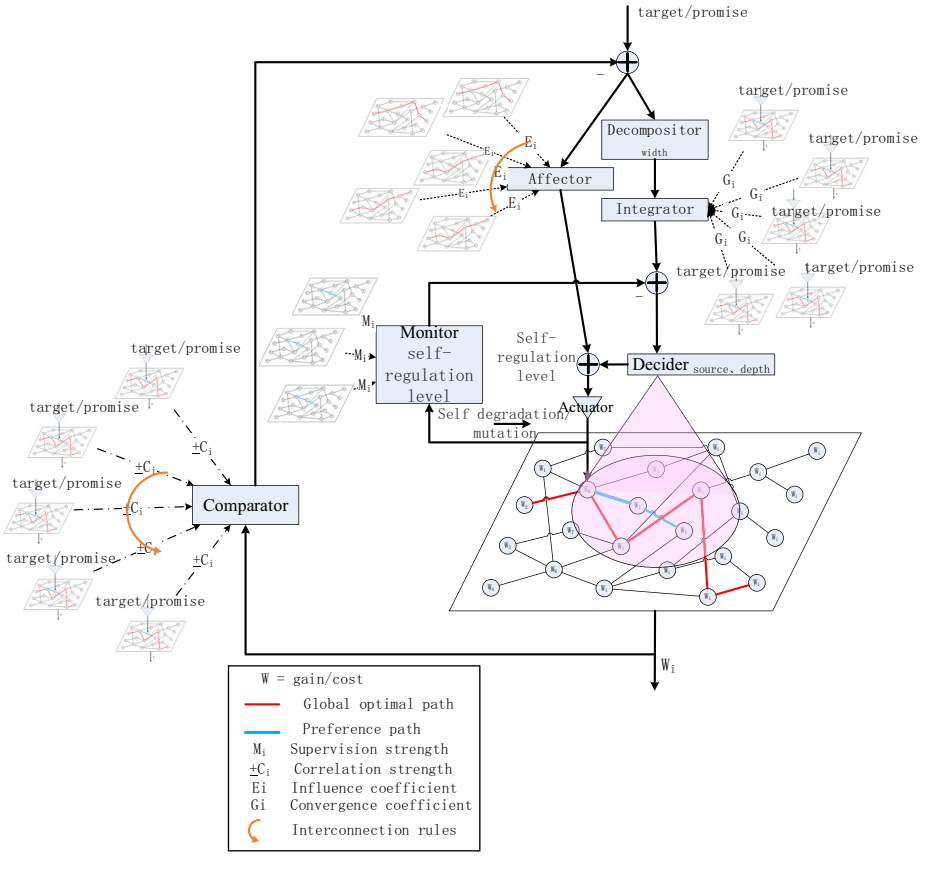

Figure 2.
Collective unit

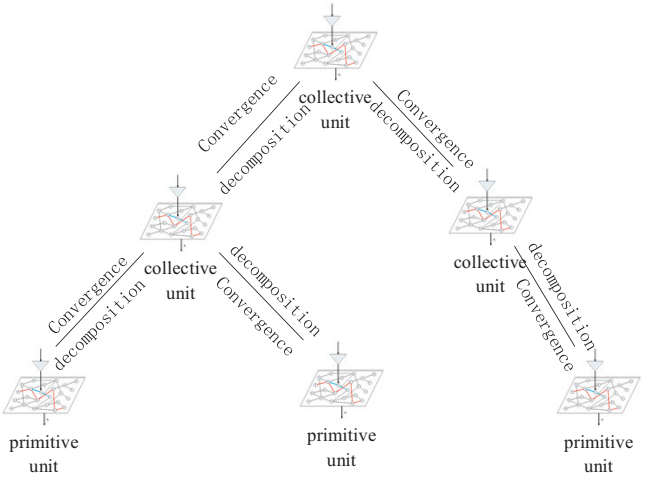

Figure 3.
Collective structure

In the crowd network control condition simulation system, the core algorithm includes monitoring system state; applying control conditions to the system. The key point of applying control conditions is find illegal messages, only in this way can throw a new information source for the negative information and forbidden negative information source. Searching for illegal information depends on the characteristics of illegal information and ordinary information. As the collective units' behavior is concentrated in a short time and shows a burst characteristic, the time interval between two information is short, and the interval distribution is in a relatively small range. However, ordinary intelligence individual is more affected by the influence of preference and influence intensity.

3.3.1 Time-based identification. In this section, based on intelligent individual's behavioral characteristics, propose average dissemination time to identify illegal information and identify collective units.

Average dissemination time refers a piece of information *Ii* began to spread in the crowd network system for the first time until the last time is forwarded, the total time is *Tfinal*, divide information forwarded times N. Use *Tavg* (*Ii*) to indicate that it can be written as:

$$T_{avg}(I_i) = \frac{T_{final}(I_i)}{N}, 1 \leq \mathrm{i} \leq \mathrm{N} \tag{1}$$

In crowd network, to complete the task of disturbing system status, once the certain information enhanced to a certain extent, then stop forwarding information. Average time interval *Tavg* (*Ii*) for collective unit is short, even if the information is forwarded by a small number of primitive units, the *Tavg* (*Ii*) cannot be changed; primitive units' forward needs to be slowly dissemination to others through the network and whether forward needs its own preference and other attributes to determine, average time interval is relatively large. Crowd network will continuously monitor the dissemination time of each piece of information. Information will be screened by abnormal time interval, then look for start of the information.

3.3.2 Community partitioning and degree distribution. The members of the crowd network are massive and collective units are cooperative to complete tasks and are organized. It is time-consuming and labor-intensive for the system to screening each member. Therefore, this paper uses the community partitioning algorithm to divide the whole network and then analyzes the internal members of the community through the degree distribution, reduce the system load.

Mark Newman (Newman, 2003) proposes a method to measure the community structure strength in the network, called modularity, which can also be called modular metrics. The formula is defined as:

$$\mathrm{Q} = \frac{1}{2m} * \sum_{ij} \left[A_{ij} - \frac{k_i * k_j}{2m} \right] \delta(C_i, C_j) \tag{2}$$

The size of the modular metric Q depends mainly on the partitioning of crowd network. Q can measure the quality of the network partition. Q's value is closer to 1, *t* quality of network partitioning is better. On the contrary, the quality of network partitioning is worse. Therefore, obtain the optimal network partition by maximizing the modularity Q value.

Mark Newman proposed a greedy algorithm FN that maximizes modularity. This method distinguishes each node in the network, treats each node as a community, then calculates the combined modular metrics of all communities, determine the optimal

combination and iterate the loop, finally get the division of the optimal community. However, the number of nodes in the crowd network is huge, this method finds the optimal solution among all the possibilities and is an N-P hard problem. It is not suitable for the community division of the crowd network.

Blondel *et al.* (2008) proposed a hierarchical greedy algorithm in 2008, which effectively reduces the time complexity. The algorithm steps are:

- consider each node in the network as a community and use the maximization module to determine the node merging;
- consider the divided community in 1 as the initial node of the new layer network and repeat 1 operation; and
- the operation cycle of 1, 2 is performed, until the Q is no longer increased after the community division and the optimal community division can be obtained.

The advantage of this algorithm is that the algorithm steps are simple and easy to implement; it can present a hierarchical community structure; compared with Mark Newman's method, the time complexity is reduced. In sparse networks, its time complexity increases linearly can handle billions of nodes in an acceptable range.

The degree is the number of edges connected to the node in the network. The connected between nodes and nodes in the crowd network is directed. Edges' number of the node is called the out-degree and in-degree.

To achieve the purpose of enhancing information intensity, according to its organizational structure, collective unit must be connected to a large number of primitive units in the internal structure, and many internal members are connected to a common upper-level collective unit. So under intelligent individual, the ratio of the out-degree O_i and in-degree Gi is very large. This paper first divides the overall network structure and statistics out-degree and in-degree of each community (*Ci*). The greater ratio degree D(*Ci*) is considered to be a collective unit.

$$\mathrm{D}(C_i) = \frac{\sum_{i=0}^{n} O_i}{\sum_{i=0}^{n} G_i} \tag{3}$$

n is the number of collective units in the community.

3.3.3 Deviation behavior. In the crowd network multi-source information disseminate system, the collective unit will release a large amount of illegal information with same filed and preference to disturb system status but other intelligent individual may not be interested in it. Then, the information and other members' information will be quite different and the collective unit will be identified according to the difference.

$$D_{ij} = r_{ij} - avg_m(r_{ij}) \tag{4}$$

Where D_{ij} is defined as deviation degree, indicating that the intelligent subjective *i*'s forwarding behavior makes a deviation degree with *m* for the information *j*, *m* is the intelligent subjective set, which received *j*; $r_{i\ j}$ represents times of *i* forward *j*; avg_m ($r_{i\ j}$) represents the average times of *m* forward *j*. If, $D_{i\ j} < 0$, it means that the forward times of it *i* is lower than the average value of m, it is not the collective unit; if $D_{i\ j} > 0$, two cases need to be considered, and set deviation threshold T^r, one is that the deviation is not large and the threshold value T^r is not exceeded, it cannot be determined whether *i* is a collective unit, and the second deviation is large, exceeding the threshold T^r, it is considered that *i* belongs to

collective unit. Therefore, based on the deviation behavior to determine whether it is a collective unit, the formula J_i summarized as:

$$J_i = \begin{cases} D_{ij} \leq T^r; \ \text{YES} \\ D_{ij} > T^r; \ \text{NO} \end{cases} \tag{5}$$

4. Simulation result

According to the steady-state maintenance simulation framework in this paper, the simulation experiment of multi-information disseminations is designed and simulation result is shown in Figure 4. Among it, the number of primitive unit is 430, the number of collective unit is 130 and the number of iteration is 1,000. Horizontal is the number of iteration, vertical is the sum of information intensity in the network, green is intensity change without any new information centralized delivery, blue is intensity change with random information delivery and red is put specific information according to unit's preference.

This experiment simulates multi-information disseminations' process in crowd network and shows the trend of information intensity in three cases. It can be seen from the figure that when there is no new information put to network, the total information intensity gradually weakens and when the iterations is great, it eventually tends to 0. When the information is put into the network and whether the information is specific or random, the total information intensity increases instantaneously and fluctuates greatly. However, through the steady-state control of system, the total information intensity can always reach a stable state. This simulation experiment gets expected effect, which shows that when

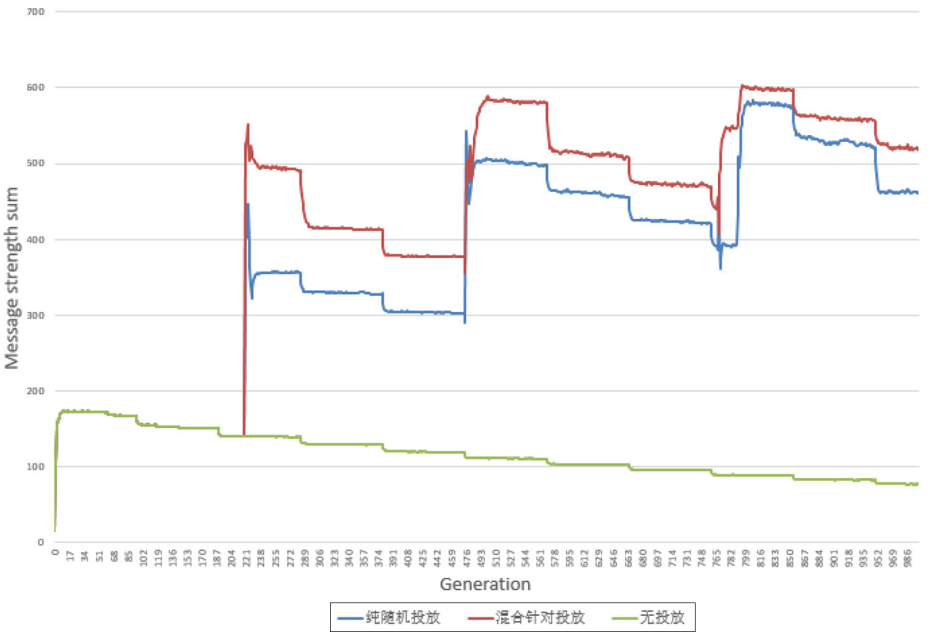

Figure 4.
Simulation result

network public opinion appears, system under steady-state maintenance framework can control the evolution of public opinion in many situations and proof the framework is reasonable and applicable.

5. Conclusion

Crowd network is rising and developing in our daily and economic lives, gradually replacing social networks. Compared with existing social networks, the structure of multi-source information dissemination systems is more changeable. Many intelligent subjective in crowd network system are self-centered subjects, it has consciousness and behavior's diversity and personalization. Its massive nodes are unmatched by traditional networks. Interaction relationship of many intelligent subjective constitutes the network structure. Their behavior is a combination of rationality and irrationality, which makes the intelligent subjective more in line with the word intelligent, single node's intelligence and the overall intelligence of the system work together, which is bound to bring greater challenges in maintaining system status.

This paper takes the crowd network multi-source information dissemination system as the research object and uses the mature information fusion model to describe the model, that according to crowd network simulation characteristics and establishes a novel steady-state maintenance simulation framework. Compared with other steady-state maintenance frameworks, it is specially designed for the crowd network simulation and that conforms to these new characteristics. In member mode, generate massive crowd science units becomes simple and easy to operate, users only need to change member attribute according to their requirements or set a reasonable range of attributes to generate crowd science units randomly. The control method is flexible and adaptable, which can deal with a variety of potential unsteady problems rather than a single problem. This framework can step by step in the whole information dissemination process, strive for control steady-state at every level of public opinion development adaptively, avoid one size fits all and achieve better effect in multi-source information dissemination. Analyzes the system framework and its members, simulates the relative stability of crowd network multi-source information disseminate system and control the information disseminate to influence intelligent subjective behavior in network system and physical world, for controllability of the evolution process provides a theoretical basis.

References

Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008 No. 10.

Chai, Y., Miao, C., Sun, B., Zheng, Y. and Li, Q. (2017), "Crowd science and engineering: concept and research framework", *International Journal of Crowd Science*, Vol. 1 No. 1, pp. 2-8, doi: 10.1108/IJCS-01-2017-0004.

Holland, J.H. (1998), "Emergence: from chaos to order".

Jindal, N. and Liu, B. (2008), "Opinion spam and analysis[C]", *International Conference on Web Search and Data Mining*.

Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E. and Makse, H.A. (2010), "Identification of influential spreaders in complex networks", *Nature Physics*, Vol. 6 No. 11, pp. 888-893.

Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B. and Lauw, H.W. (2010), "Detecting product review spammers using rating behaviors", *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, ACM, Toronto, Ontario*, October 26-30, 2010.

Mukherjee, A., Bing, L., (2012), and N. and Glance, "Spotting fake reviewer groups in consumer reviews", *International Conference on World Wide Web*.

Naud, A. and Usui, S. (2007), "Exploration of a text collection and identification of topics by clustering",

Newman, M.E.J. (2003), "Fast algorithm for detecting community structure in networks", *Physical Review E Stat Nonlin Soft Matter Phys*, Vol. 69 No. 6, p. 066133.

Zhao, B., Ji, G., Qu, W. and Gu, Y. (2013), "Microblog garbage user filtering algorithm based on reuse detection", *Journal of Nanjing University*, Vol. 49 No. 4.

Further readings

Barabási, A.L. and Albert, R. (1999), "Emergence of scaling in random networks", *Science*, Vol. 286 No. 5439, pp. 509-512.

Corresponding author

Hongbo Sun can be contacted at: hsun@ytu.edu.cn