Accepted Manuscript

Earthquake prediction based on community division

Yanjie Xu, Tao Ren, Yiyang Liu, Zhe Li

PII: S0378-4371(18)30572-7
DOI: https://doi.org/10.1016/j.physa.2018.05.035
Reference: PHYSA 19575

To appear in: *Physica A*

Received date : 17 November 2017
Revised date : 23 March 2018

Please cite this article as: Y. Xu, T. Ren, Y. Liu, Z. Li, Earthquake prediction based on community division, *Physica A* (2018), https://doi.org/10.1016/j.physa.2018.05.035

Highlights

- Two directed weighted earthquake network are structured by the earthquake influence number and the maximum magnitude.
- The earthquake is predicted based on the minimum edge weight.
- Two networks are divided into communities and their earthquake prediction accuracies are improved.

Earthquake Prediction Based on Community Division

Yanjie Xu, Tao Ren, Yiyang Liu and Zhe Li

***Abstract*—In time-space influence domain, two directed weighted earthquake network are structured based on the earthquake number and the maximum magnitude of Southern California. The earthquake prediction method is proposed based on the minimum edge weight. By CNM(community detection method) community division algorithm, the network is divided into several communities and the top 10 communities can be selected according to the number of nodes. Finally, we compare the accuracy of the divided network with the network without community division. The simulation results show that the community division can improve the accuracy of the earthquake prediction.**

***Index Terms*—Earthquake network, directed weighted network, prediction, community division**

I. Introduction

Earthquake is a worldwide problem since it has destructive power. Even if in a small earthquake belt, earthquake can happen thousands of times per year. Earthquake is one of the most important natural phenomena that hazards our life and property.

Fortunately, a number of researchers devote to revealing the regularities of the phenomena and make fruitful achievements in a long research history. In these achievements, the most famous one is the Gutenburg-Richter law[1] that reveals the relationship between earthquake magnitude and frequency, which is often used to study earthquake in geophysics[2-4], and the other one is Omori law[5] that describes the relationship between the frequency of the aftershocks and the time interval. For modeling the earthquakes, Baiesi and Paczuski[6,7] defined different tremor events as nodes of a network, where a pair of node is linked if the correlation between them exceeds a certain threshold. Abe and Suzuki[8-15] considered that each pair of successive earthquakes events are associated. While, He Xuan[16] proposed a different approach to build the network edge based on the space-time influence domain.

In order to build an earthquake network, Abe S and Suzuki N[8] proposed a research method of modeling a network of earthquake regions. Firstly, the earthquake region is divided into a number of cells one by one. If an earthquake happens in a cell, the cell is defined as a node in the earthquake network. Further, if two nodes are both affected by an identical earthquake, a link between the nodes is added into the network, and in an earthquake event, if two tremors occur in a node, self-loop is applied to the node. After studying the earthquake data of Southern California and Japan, it is found that earthquake networks of the two regions have scale-free characteristic by statistic on degree distribution of network node. Well, based on the study of Gardner J K[17], He Xuan[16] proposed a construction method of earthquake network by the space-time influence domain. Lin et al[18] established a earthquake

This work is partially supported by National Natural Science Foundation of China (61473073, 61104074), Fundamental Research Funds for the Central Universities (N161702001), and Program for Liaoning Excellent Talents in University (LJQ2014028).
Yanjie Xu is with the Software College, Northeastern University, Shenyang 110169, PR China (e-mail: nancy_xuyanjie@163.com).
Tao Ren is with the Software College, Northeastern University, Shenyang 110169, PR China (e-mail: chinarentao@163.com).
Yiyang Liu is with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, PR China (e-mail: 1285906739@qq.com).
Zhe Li is with the Software College, Northeastern University, Shenyang 110169, PR China (e-mail: gislzneu@163.com).

recurrence network based on the magnitude time series of California. The constructed networks are unweighted. However, in actual networks, the importance of different edge is different. Therefore, it is necessary to introduce the "weight" associated with the edge attributes, and then build a weighted earthquake network.

In current earthquake prediction studies, Kong Q[19] designs an intelligent earthquake data analysis software, Myshake, which can be used to collect data for early prediction of earthquakes and enhance EEW's ability of prediction(EEW, earthquake early-warning).Howell S et al[20] use statistical methods to alleviate earthquake hazards by pattern selection. Zhang Y et al[21] use an image segmentation method to establish a fast search engine for detecting the location of the source rapidly after an earthquake. Recently, the research of earthquake prediction based on complex network is increasing. He X et al[16,22] select the nodes in the California earthquake network by the K kernel theory, and predicate the earthquakes by the Bayesian network. Men K P et al[23,24]predicate M ≥ 7 Earthquakes in Xinjiang Region and M ≥ 8 Earthquakes in mainland of China by designing an ordered network. However, these studies only select a very small number of seismic data or a few nodes for research. In fact, when making prediction, we should study all regions and most of the seismic data.

In this paper, the seismic data of South California(114.0W-122.0W，32.0N -37.0N) is taken as the research object(1992-2014 seismic data is used to construct the network, and 2015 data is used to make earthquake prediction , data is from http://service.scedc.caltech.edu/ftp/catalogs/SCEC_DC/). Based on space-time influence domain, two kinds of directed weighted earthquake networks are respectively generated by the earthquake number and the maximum magnitude. Then, the modularity of weighted network is combined with the CNM algorithm[25] to divide two directed weighted earthquake networks. Next, on the basis of the community division, the top 10 communities are selected according to the number of nodes in the community. The earthquake is predicted for these communities based on minimum edge weight. Finally, the prediction accuracy of the network is normalized. It proves that the prediction accuracy of the earthquake network after the community division has been improved.

The outline of this paper is as follows. In Section 2, we construct two directed weighted earthquake network and predicate earthquake based on the minimum edge weight. In Section 3, we divide two directed weighted earthquake networks by combing the modularity with the CNM algorithm and select the top 10 communities. In Section 4, we make simulation on two directed weighted network. Finally, a summary and some conclusions are stated in Section 5.

II. Construction of Earthquake Network

California is in the south Pacific Rim seismic belt, where most shallow earthquakes occur, so in the division of the earthquake region, we only consider the latitude and longitude without depth. Furthermore, the magnitude of main seismic nodes is at least 2.5 and the magnitude of all nodes is at least 1.

For node i , its duration of the influence time and maximum influence distance are not linearly related to its magnitude. Their relation can be expressed as follows:

$$\log T_i = a_1 M_i + b_1 \quad (1)$$

$$\log L_i = a_2 M_i + b_2 \quad (2)$$

where M_i represents the magnitude of node i , T_i indicates the duration of the influence time, L_i indicates the maximum influence distance, and a_1、 a_2、 b_1 、 b_2 are constants obtained according to the statistics.

The time interval and distance between node i and node j are expressed as follows:

$$\Delta t_{ij} = t_i - t_j \quad (3)$$

$$d_{ij} = 2R\sin^{-1}\left(\sqrt{\sin^2\left(\frac{\pi\left(lat_i - lat_j\right)}{360}\right) + \cos\left(\frac{\pi lat_i}{180}\right)\cos\left(\frac{\pi lat_j}{180}\right)\sin^2\left(\frac{\pi\left(lon_i - lon_j\right)}{360}\right)}\right) \tag{4}$$

where t_i and t_j represent the seismic beginning of node i and j respectively, lat_i represents the latitude of the earthquake, lon_i indicates the longitude of the earthquake. In this paper, the magnitude of shock region is greater than the magnitude of the region affected by it.

Definition 1: In a seismic event, if there is an earthquake in both node i and j , $\Delta t \le T_i$ and $d_{ij} \le L_i$, node i has an edge pointing to j .The repeated edges and self-loop should be excluded.

In order to study the relationship between nodes in the earthquake network, this paper introduces the concept of weight and proposes two methods to define the edge weight.

A. Strategy 1

Strategy 1is based on the earthquake number. For strategy 1, the edge weight s_{ij} is defined as

$$S_{ij} = \frac{H_{ij}}{H_i} \tag{5}$$

where H_i represents the earthquake number occurring in node i , H_{ij} indicates the earthquake number of node j affected by node i .

B. Strategy 2

Strategy 2 is based on the maximum magnitude. For strategy 2, the edge weight s_{ijm} of the m -th earthquake event is defined as

$$s_{ijm} = \frac{M_{jm}}{M_{im}} \tag{6}$$

where M_{im} represents the magnitude of the node i of the m -th earthquake, M_{jm} represents the maximum magnitude of the node j affected by node i during the m -th earthquake event.

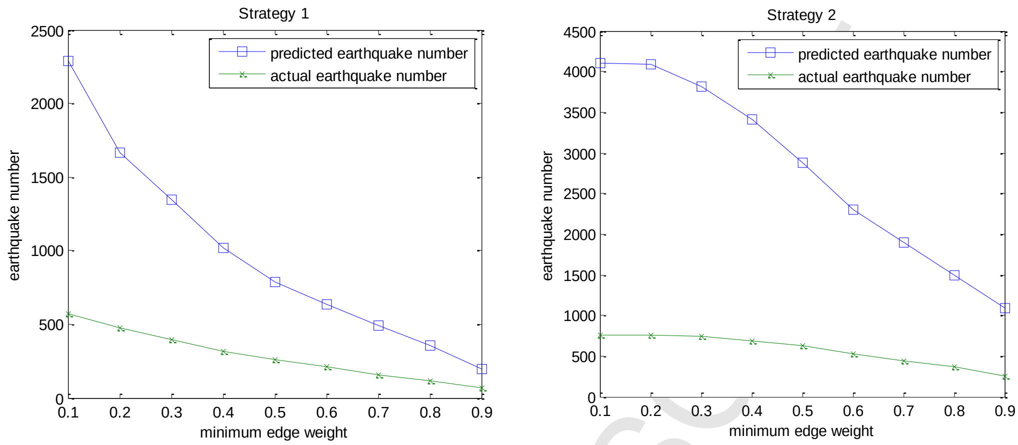
From 1992 to 2014, the number of earthquakes with magnitude greater than 1 is 228,393 in Southern California. So, there are many edge weight values based on the maximum magnitude. In this paper, the maximum value is defined as the edge weight value

$$S_{ij} = \max_{m \subseteq n_i} s_{ijm} \tag{7}$$

where n_i is the number of earthquake in node i .

The network has 1890 nodes and 16137 edges. Based on the statistical analysis of node degree and clustering coefficient distribution in the network, two directed weighted earthquake network generated in this paper have scale-free and small-world characteristics.

Based on two directed weighted earthquake networks, earthquake will be predicted for Southern California. In this paper, the prediction based on the minimum edge weight is used to study the influence of the edge number on prediction accuracy in this paper. The number of predicted and actual earthquakes in two weighted networks varies with the weight as shown in Fig1.



(a) (b)

Fig. 1 Prediction of directed weighted earthquake network, (a) strategy 1, (b) strategy 2

As shown in Fig. 1, the number of earthquake is correspondingly reduced with the minimum edge weight from 0.1 to 0.9 and the predicted number of earthquake is correspondingly reduced. With the increase of the minimum edge weight, the prediction accuracy of two weighted earthquake networks is shown in Table 1.

TABLE 1

PREDICTION ACCURACY OF TWO NETWORK BASED ON MINIMUM EDGE WEIGHT

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
strategy 1	0.248	0.284	0.291	0.313	0.326	0.3330	0.325	0.332	0.359
strategy 2	0.184	0.184	0.194	0.202	0.217	0.226	0.234	0.245	0.229

As Fig. 1 and Table 1 shown, the overall experimental result of earthquake prediction is not effective because the region where nodes are located is not considered when we predict the earthquake. Based on the analysis of California's 1992-2014 seismic data, regions with a magnitude greater than 5 are concentrated in small local regions (33N-34.5N, 116.3W-118.1W).Moreover, the influence of the main shock region on the other regions is mostly related to magnitude. Therefore, it is impossible to predict the earthquake based on earthquake influence number or the magnitude only. In order to improve the accuracy of earthquake prediction, the earthquake network should be reasonably divided and then predicted according to the structure of the network.

III. THE ALGORITHM OF COMMUNITY DIVISION

Based on the CNM algorithm proposed in [22], the modularity of the directed weighted network is considered, in order to divide the community structure of the network. The algorithm flow is as follows:

Step 1, Initialize: Assume that each node in the network is an independent community, modularity $Q=0$. The initial edge weight e_{ij}, the node in-edge weight a_i^{in} and out-edge weight a_i^{out} of directed weighted network are defined as follows:

$$e_{ij}=\begin{cases}\dfrac{s_{ij}}{\sum\limits_{v\neq z} s_{vz}}, & \text{if node } i \text{ has edge pointing to node } j\\ 0, & \text{others}\end{cases} \tag{8}$$

$$a_i^{in}=\frac{w_i^{in}}{\sum\limits_k w_k^{in}} \tag{9}$$

$$a_i^{out} = \frac{w_i^{out}}{\sum_k w_k^{out}} \tag{10}$$

where s_{ij} represents the edge weight between the community i and j, w_i^{in} and w_i^{out} represent the in-edge and out-edge weight of community i respectively.

The initial modularity increment matrix is defined as

$$\Delta Q_{ij} = \begin{cases} s_{ij} - a_i^{out} a_j^{in}, & \text{if node } i \text{ has edge pointing to node } j \\ 0, & \text{others} \end{cases} \tag{11}$$

The max-heap H can be obtained by calculating the maximum increment from each row of the modularity increment matrix.

Step 2: Select the ΔQ_{ij}, which is the largest value in the max-heap H. Find the corresponding communities i and j, and join them into new community j .Then update ΔQ_{ij}, H and auxiliary vector $\boldsymbol{a}_i$, ΔQ_{ij} and $\boldsymbol{a}_i$ can be updated by

$$\Delta Q'_{jk} = \begin{cases} \Delta Q_{ik} + \Delta Q_{jk}, & \text{if community } i, j \text{ has edge pointing to } k \\ \Delta Q_{ik} \text{-} 2a_j^{out} a_k^{in}, & \text{if community } i \text{ has edge pointing to } k, \text{ while community } j \text{ is not} \\ \Delta Q_{jk} \text{-} 2a_i^{out} a_k^{in}, & \text{if community } j \text{ has edge pointing to } k, \text{ while community } i \text{ is not} \end{cases} \tag{12}$$

$$\boldsymbol{a}_j^{'in} = \boldsymbol{a}_i^{in} + \boldsymbol{a}_j^{in}, \quad \boldsymbol{a}_i^{'in} = \mathbf{0} \tag{13}$$

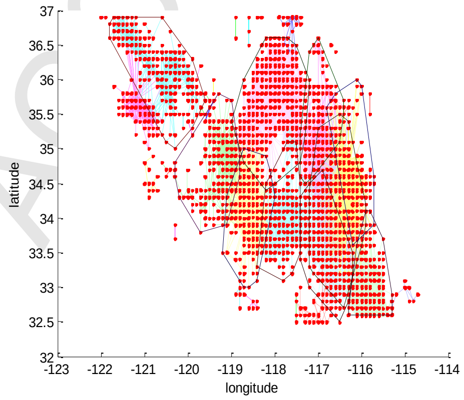
$$\boldsymbol{a}_j^{'out} = \boldsymbol{a}_i^{out} + \boldsymbol{a}_j^{out}, \quad \boldsymbol{a}_i^{'out} = \mathbf{0} \tag{14}$$

Then we can get the modularity value $Q = Q + \Delta Q_{ij}$.

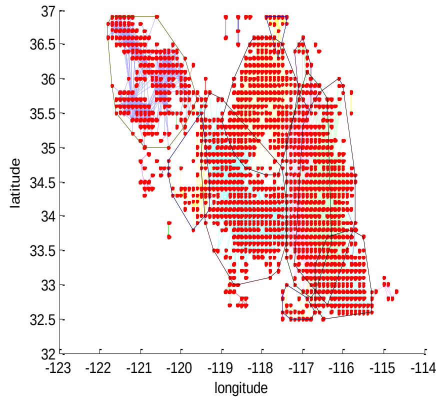
Step 3: Repeat step 2 until all the nodes are divided into their own appropriate community, that is, the largest element of the network modularity increment matrix changes from positive to negative. The modularity Q reaches the peak value.

According to the algorithm, the directed weighted earthquake network based on the earthquake number is divided into 41 communities, and the directed weighted earthquake network based on the maximum magnitude is divided into 38 communities.

The communities of two directed weighted earthquake networks are sorted by the number of nodes. In the directed weighted earthquake network based on the earthquake number, the first community with the largest number of nodes accounts for 11.5% of the total network, while the top 10 community nodes account for 63% of the network. In the directed weighted earthquake network based on the maximum magnitude, the first community with the largest number of nodes accounts for 15.2%of the total network, while the top 10 community nodes account for 66.6% of the network. The top 10 communities are shown in Fig 2.



(a)



(b)

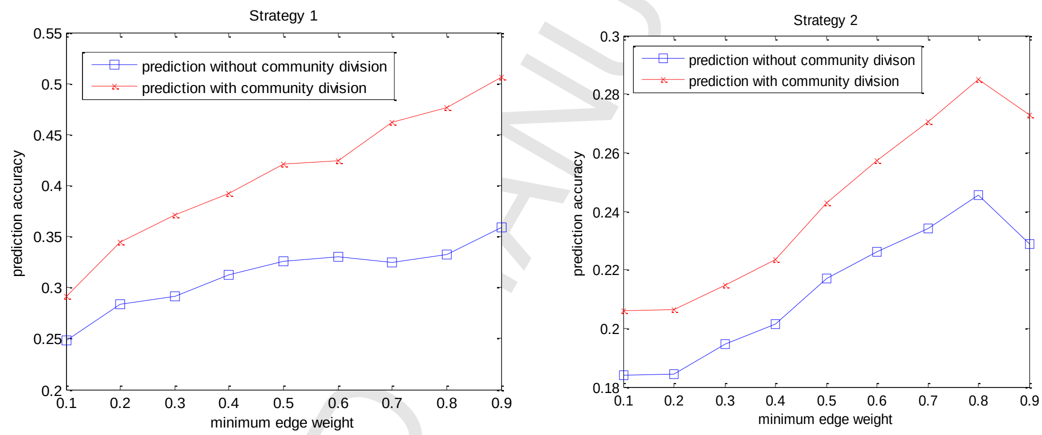
Fig. 2 The top 10 communities of the two constructed networks (a) strategy 1(b) strategy 2

IV. SIMULATION ANALYSIS

In the top 10 communities of two directed weighted earthquake networks, earthquake is predicted based on the minimum edge weight. In order to normalize the prediction accuracy of the network after the community division, the prediction accuracy of the network is calculated as follows:

$$Z_l = \frac{\sum_r n_r z_l^r}{\sum_r n_r} \tag{15}$$

where Z_l is the prediction accuracy for the minimum edge weight value l, n_r represents the earthquake number of the r-th community, z_l^r represents the prediction accuracy of the r-th community. In this paper, when the prediction number of earthquake is null, the corresponding data is discarded.



(a) (b)

Fig. 3 Comparison of prediction accuracy (a) strategy 1 (b) strategy 2

As shown in Fig. 3(a), with the increase of the minimum edge weight, the accuracy of the earthquake prediction increases with or without division, and the earthquake prediction accuracy with community division is significantly higher than the one without community division. For Fig. 3(b), with the increase of the minimum edge weight, the earthquake prediction accuracy with community division is higher than that one without community division, and their trend is same.

It is obvious that the prediction accuracy of two directed weighted earthquake network are improved with community division. The earthquake network is divided into communities, that is, the seismic area is divided into a certain region, so that the earthquake prediction can be carried out in each region, thus the prediction accuracy is improved.

V. CONCLUSION

In this paper, we generate two directed weighted earthquake networks based on time-space influence domain. Strategy 1 is based on the earthquake number. Strategy 2 is based on the maximum magnitude. Then combining with the CNM community division algorithm, the network is divided into several communities and we select the top 10 communities according to the number of nodes. It is obvious that the top 10 communities of the network contain most nodes. Because the smaller the minimum edge weight is, the more the edge of the earthquake network is. So in order to better study the earthquake prediction accuracy as the earthquake number changes, we make earthquake prediction based on the minimum edge weight. Although there are many community nodes, the number of earthquake is not proportional to the number of nodes. So in order to better study the influence of community division on the earthquake prediction accuracy, this paper normalizes the prediction accuracy and the earthquakes number, and compares with the prediction accuracy without community division. The simulations show that the network earthquake prediction accuracy can be improved by community division.

REFERENCES

[1] Gutenberg B, Richter C F. Frequency of earthquakes in California[J]. Bulletin of the Seismological Society of America, 1944, 34(4): 185-188.

[2] K. Schalcher, G. M. Dimirovski, A. Kugi and D. V. Jurukovski, "Vibration Control of High-Rise Buildings." In K. J. Astroem, P. Albertos, M. Blanke, A. Isdidori, W Schaueflberger and R. Sanz, Editors, Control of Complex Systems. Springer-Verlag, Ltd., London, 2001, Chapter 15, pp. 348-368 (ESF Reserch monograph).

[3] R. Stefansson, Advances in Erthquake Prediction, Research and Risk Mitgation. Springer Praxis Books, Berlin Heidelberg, 2011 (Seismic resercah monongraph).

[4] M. Hayakawa, Earthquake Prediction with Radio Tchniques. Springer Praxis Books, Berlin Heidelberg, 2015.

[5] Omori F. On the aftershocks of earthquakes ［J］． Journal ofthe College of Science, Imperial University of Tokyo, 1894, 7:111 － 119.

[6] Baiesi M, Paczuski M. Scale-free networks of earthquakes and aftershocks[J]. Physical review E, 2004, 69(6): 066106-1-066106-8.

[7] Davidsen J, Paczuski M. Analysis of the spatial distribution between successive earthquakes[J]. Physical Review Letters, 2005, 94(4): 048501.

[8] Abe S, Suzuki N. Scale-free network of earthquakes[J]. Europhysics Letters, 2004, 65(4): 581-586.

[9] Abe S, Suzuki N. Small-world structure of earthquake network[J]. Physica A Statistical Mechanics & Its Applications, 2004, 337(1–2):357-362.

[10] Abe S, Suzuki N. Complex earthquake networks: hierarchical organization and assortative mixing[J]. Physical Review E, 2006, 74(2):26113-1-26113-5.

[11] Abe S, Suzuki N. Dynamical evolution of clustering in complex network of earthquakes[J]. European Physical Journal B, 2007, 59(1):93-97.

[12] Abe S, Suzuki N. Violation of the scaling relation and non-Markovian nature of earthquake aftershocks[J]. Physica A Statistical Mechanics & Its Applications, 2009, 388(9):1917-1920.

[13] Abe S, Suzuki N. Scaling relation for earthquake networks[J]. Physica A Statistical Mechanics & Its Applications, 2009, 388(12):2511-2514.

[14] Abe S, Suzuki N. Determination of the scale of coarse graining in earthquake network[J]. Epl, 2009, 87(4):48008-48012.

[15] Abe S, Pastén D, Suzuki N. Finite data-size scaling of clustering in earthquake networks[J]. Physica A Statistical Mechanics & Its Applications, 2011, 390(7):1343-1349.

[16] He X, Zhao H, Cai W, et al. Analyzing the structure of earthquake network by k-core decomposition[J]. Physica A: Statistical Mechanics and its Applications, 2015, 421: 34-43.

[17] Gardner J K, Knopoff L. Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?[J]. Bulletin of the Seismological Society of America, 1974, 64(5): 1363-1367.

[18] Lin M, Fan X X, Wang G, et al. Network structure entropy and its dynamical evolution for recurrence networks from earthquake magnitude time series[J]. The European Physical Journal B, 2016, 89(5): 1-7.

[19] Kong Q, Allen R M, Schreier L, et al. MyShake: A smartphone earthquake network for earthquake early warning and beyond[J]. Science advances, 2016, 2(2): e1501055.

[20] Howell S, Smith-Konter B, Frazer N, et al. The vertical fingerprint of earthquake cycle loading in southern California[J]. Nature Geoscience, 2016.

[21] Zhang J, Zhang H, Chen E, et al. Real-time earthquake monitoring using a search engine method[J]. Nature communications, 2014, 5.

[22] Zhang Y, Zhao H, He X, et al. Bayesian prediction of earthquake network based on space–time influence domain[J]. Physica A: Statistical Mechanics and its Applications, 2016, 445: 138-149.

[23] Men K. The ordered network structure of M≥ 8 great earthquakes and their prediction in Mainland China[J]. Zeitschrift für Naturforschung A, 2012, 67(5): 308-316.

[24] Men K P, Zhao K. The Ordered Network Structure and Prediction Summary for M≥7 Earthquakes in Xinjiang Region of China[J]. Zeitschrift für Naturforschung A, 2014, 69(12): 635-644.

[25] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks[J]. Physical review E, 2004, 70(6): 066111.