

R/S Analysis in Mobile Social Networks

Wei Zheng

School of Software
Nanchang Hangkong University
Nanchang 330063, China
School of Mathematics and Statistics
Xidian University, Xi'an 710071

Kang Zhao

School of Software
Nanchang Hangkong University
Nanchang 330063, China
E-mail: kangzhao_email@126.com

Abstract-In this paper, a R/S analysis method based on clustering coefficient time series is presented to testify the mobile social networks (MSNs) is fractal and organized. It is found that MSNs are fractal in the scale of clustering coefficient. Then use the Hurst exponent to make some simple forecasts if the clustering coefficient of the network is breaking out in the next period. The result of the experiment verifies that the clustering coefficient time series of complex networks (MSNs) have the characteristics of fractal, and R/S analysis method proves that the cluster coefficient time series is bias random walk, and the results on forecasting the outburst of the clustering coefficient in MSNs meet the expectations.

Keywords-MSNs; R/S analysis; clustering coefficient; fractal theory; forecast

I. INTRODUCTION

Rescaled range analysis method(R/S analysis) is just a method which uses time series to calculate data and forecast how data moves [1]. It is proposed by English hydrologist Hurst on the basis of a large number of empirical studies, and it is gradually improved by Mandelbrot (1972,1975), Mandelbrot & Wallis (1969), Lo (1991) [2]. Rescaled range analysis method(R/S analysis) can tell random sequence and nonrandom sequence apart, and it can also explore nonlinear system in the process of long-term memory [3].

While studying natural phenomena, it usually simplify the model by regarding the events as points without size or mass during a time process, and thus can get a time series composed of those events. Researchers suggest that, although many natural phenomena seem rather complicated at the first sight, they do share some consistent and simple features [4]. A Rescaled range analysis method is used to simplify such complex question into an easier one. It can also help us to calculate the long memory of the time series, and knowing about if the clustering coefficient has the memory can help to know the trend and fluctuations in internal correlation of MSNs.

Rescaled range analysis method(R/S analysis) plays a critical role in many fields, such as flood forecasting, volume of runoff of rivers, the change of water level in lakes, analysis in groundwater regime, stock analysis, dam security, disaster forecasting, climatic variation forecasting. For examples, Wang et al. do research on hydrological time-series trend analysis and the diagnostic approach of dissimilarity [5]. Men et al. use rescaled range analysis

method to analyze some hydrometric station near 6 branches of the Yangtze River [6]. Zhang et al. do a research on the trend of underground water level of Xing Ping city [7]. Wang et al. find that Chinese Security Markets is nonlinear, and it also has the character of persistence and anti-persistence [8]. Fen et al. try to forecast the possibility of the typhoon landing in Zhejiang province [9]. Zhao et al. use R/S analysis method to analyze some basic characters of climate in Lanzhou city, and they find persistence in urbanization heating [10].

In recent years, climate has become a hot issue all over the world with the development of industry. Liu et al. take advantages of some surface meteorological observation data of Xian city, and use R/S analysis method to forecast the trend of the climate [11].

Mobile Social Network (MSN) is a social network which aims to study the activity routines of social groups by collecting the information of mobile terminals. For example, through excavating the data of people traveling and taxis track, it can help taxi drivers to find passengers quickly and conveniently. People use e-mail, BBS, blogs by handed mobile devices, and they form a social group.

At present, the applications of Mobile Social Network (MSN) are wide. For example, Wang et al. studied the context sensing mechanism of Mobile Social Network (MSN), and the aim is to improve the accuracy and reliability of the service selection, as well as to avoid blindness and randomness [12].

Watts and Strogatz in 1998 put forward the clustering coefficient for the first time [13]. And this time, a research will be done on the fractal of the complex network in the scale of clustering coefficient.

The research is that using rescaled range analysis method(R/S analysis) to analyze cluster coefficient time series in complex networks. Our aim is to prove that cluster coefficient time series is fractal, and use the conclusions to forecast the internal correlation and the size of the complex network.

Because there were little researchers who have done such research before, this research is a new way to study complex networks in a way. The data used to carry out the research is one of the time series of MSNs. It is proved that the cluster coefficient time series of complex networks are fractal and have long memory, through some of the series of

experiments, and could forecast the internal correlation and the size of complex networks.

In order to make it clearly, the very work will be put forward in three aspects, they are Preliminary, Results and Discussion and Conclusions.

II. PRELIMINARY

The cluster coefficient of complex networks is one of the import standards for measuring the size and internal correlation for studying complex networks [14]. Focus on the clustering coefficient of complex network, so it may find out the size of the specific network and the general rule of development trend.

A. Cluster

Cluster coefficient is put forward by Watts and Strogatz, and it is also an important standard to describe topology of complex networks. The aim is to investigate the character of mass organization in complex networks. The definition of cluster coefficient is the average value of real connections of neighbors divided the possible connections of neighbors[15].

The cluster coefficient of a single node shows that the compact degree of neighbor nodes which form a complete graph[16]. So let us define the local cluster coefficient as follow,

$G = (V, E)$, Graph G contains a series of points V and the edge E connected to them.

$e_{i,j}$, Edge connected with node i and node j ;

E_i , The total edges connected to node i . k_i , The number of neighbor nodes of V_i .

So, it defines cluster coefficient like this,

$$C_i = \frac{2E_i}{k_i(k_i-1)} \quad (1)$$

It can calculate the average of all of the nodes' cluster coefficient to observe the aggregation level of the network's nodes, because of the single node's affected by migration path strongly[17].

Next, the average cluster coefficient of the network is defined.

The cluster coefficient of the whole network is defined as the average local cluster coefficient of all nodes, that is,

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (2)$$

Forecast the internal correlation and the size of complex network by analyzing the changes and trends of average clustering coefficient with time.

B. R/S Analysis

Rescaled Range analysis method(R/S analysis), which is based on oceans of empirical researches, is discovered by English hydrologist Hurst. It plays a critical role in fractal theory[18]. It can forecast the size trend and calculate the internal correlation trend of complex networks. Here are the basic principles of R/S analysis,

For a time series $\{X_t\}$, cutting it up into A of length N intervals with equal length, as for every intervals, and suppose that,

$$M_n = 1/N \sum_{i=1}^N X_i \quad (3)$$

M_n is the average value of the n th section, and it supposed that,

$$X_{t,n} = \sum_{u=1}^t (x_u - M_n) \quad (4)$$

$X_{t,n}$ is the cumulative deviation of interval n , making it that,

$$R = \max(X_{t,n}) - \min(X_{t,n}) \quad (5)$$

If S stands for the standard deviation of the time series $\{X_t\}$, then,

$$s = [\frac{1}{n} \sum_{i=1}^t (x_i - \bar{x}_t)^2]^{1/2} \quad (6)$$

It can define Rescaled Range(R/S), it increases with time. Through long time practice and summarization, Hurst built a relation as follow (K is a constant and H stands for Hurst exponent),

$$\frac{R}{S} = K(n)^H \quad (7)$$

Taking logarithm on both sides of the equation, and get a new equation as follow,

$$\log(\frac{R}{S})_n = H \log(n) + \log(K) \quad (8)$$

Finally, it can estimate the Hurst exponent by making $\log(R/S)$ and $\log(n)$ do least squares regression.

The value of Hurst exponent has three kinds as follow,

(1) According to the fractal theory, $H = 0.5$ would imply an independent process.

(2) $0.5 < H < 1.0$ implies a persistent time series, and a persistent time series is characterized by long memory effects. Theoretically, what happens today impacts the future forever. This long memory occurs regardless of time scale. All daily changes are correlated with all future daily changes; all weekly changes are correlated with all future weekly changes.

(3) $0 < H < 0.5$ signifies anti-persistence. An anti-persistence system covers less distance than a random one. For the system to covers less distance, it must reverse itself more frequently than a random process.

Hurst investigated diverse natural phenomena-rainfall, sunspots, mud sentiments, tree rings, anything with a long time series. In all cases, Hurst found that H greater than 0.5. Generally, persistent time series are the most common type in nature. So it can also try to guess that persistent time series are in the cluster coefficient of complex networks.

III. RESULTS AND DISCUSSION

In this part the data resource will be shown, the results, some discussions and some conclusions which have made through the experiment.

A. Basic Data

The data of this experiment come from MSNs, a part of social networking services, which are called Inf 05 and Roller. In this paper, Inf 05 will be mainly discussed. The network contains every connection messages tell the beginning time and the ending time of each connection. Then deal with the data by setting them in a sparse matrix, and counting out the connecting time of each node. According to the length of each connecting time, then they can be put into the right groups. The data are dealt, and finally formed into a line.

B. Data Analysis & Bias Random Walk

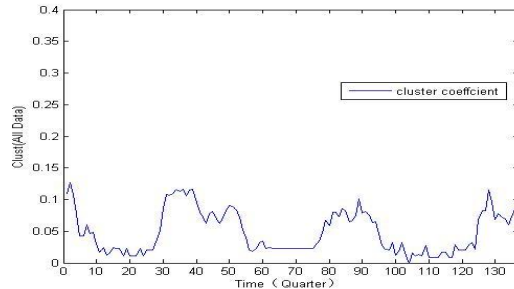


Figure 1. The cluster coefficient time series of Inf 05.

The data is the average cluster coefficient of MSNs. Every value on X axis stands for the average cluster coefficient in an hour. Then deal with the data by using the method mentioned about.

Treating cluster coefficient as random variables, that is $\{X_t\}$, and using R/S method to calculate R (increase over time), S, $\log(R/S)$ and $\log(N)$. Then use $\log(R/S)$ and $\log(N)$ to do least squares, and some of the results display as follow,

group	$\log(R/S)$	$\log N$	group	$\log(R/S)$	$\log N$
2	4.3988	0.6931	15	5.8505	2.7081
3	4.1766	1.0986	20	5.8831	2.9957
4	5.2311	1.3863	25	5.6851	3.2189
5	5.4231	1.6094	30	5.8768	3.4012
6	5.2702	1.7918	35	5.7977	3.5553
7	5.5794	1.9459	40	5.9313	3.6889
8	5.6976	2.0794	45	5.6524	3.8067
9	5.6175	2.1972	50	5.7578	3.9120
10	5.5042	2.3026	55	5.9063	4.0073
12	5.8285	2.4849	60	5.9933	4.0943

Figure 2. The $\log(R/S)$ and $\log(N)$ of each group of Inf 05.

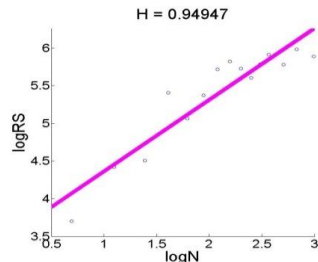


Figure 3. Use $\log(R/S)$ and $\log(N)$ of Inf 05 to do least squares.

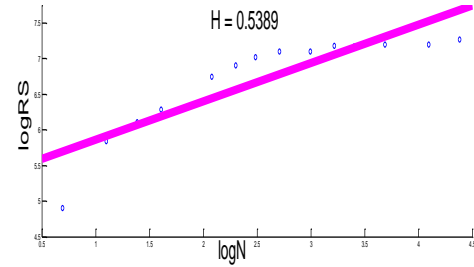


Figure 4. The $\log(R/S)$ and $\log(N)$ least squares of Roller.

From the figure1 it can be seen that there exists some unpredictable noise that makes the dispersed points fluctuated periodically. If there is no noise exist, the discrete points may be a straight line. The noise may be the white noise or some spike noise. Thinking about the reason of the noise, it might be some unexpected changes or the sudden hot spots of some social topics or etc. If collect much more data, make every group of data large enough, it could reduce the influence of unnecessary noise, then the point in the figure will be in a line.

The figure above(Figure.3) shows that the cluster coefficient time series of MSNs are fractal, and it can calculate the fractal dimension by using the equation,

$$D = 2 - H \quad (9)$$

From the Hurst exponents, it can get that the cluster coefficient time series is bias random walk, the Figure 3 shows that it can't tell the exact trend that the data is raise or drop in the future, but it can probably tells something in the next period.

C. Data Forecast

Take Inf 05 for example, all of the calculation is displayed in Table 1, ΔH stands for the D-value of this period of Hurst exponent minus the last period [19].

From Table.1 it can be seen that even though some of the data is cut, the H is not changes so fast, so it is called the long memory, and it is ubiquitous in the time series of nature.

Here, talk about the period of 5, 6, 8, 15, 16, because these periods are booming. If the17 data explains what has happened. In order to forecast the burst, it might investigate the period before instead of the right period, which is the period of 4, 5, 7, 14, 15. It finds out that the H of these periods is larger than the periods before, but the ΔH is smaller. For example the H of period 4 is larger than period 3, but the ΔH is smaller. Although some other periods which are not booming have such phenomenon, they do not appear so much. So from this view, it may predict the next period whether it is booming. From the result, the accuracy of forecasting will be 100 percent, and this method can ensure around 70 percent accuracy.

Here, use this discovery to predict the relationship among nodes in the network specifically. And with this method, a strategy may be built, which can distribute networks resources more reasonable and convenient.

No.	Clustering coefficient	H	ΔH
1	0.199	0.682443	--
2	0.028607	0.610847	-0.0716
3	0.032338	0.630467	0.01962
4	0.021349	0.642547	0.01208
5	0.06856	0.650706	0.008159
6	0.08257	0.652339	0.001633
7	0.081897	0.653258	0.000919
8	0.115671	0.639258	-0.014
9	0.096321	0.60943	-0.02983
10	0.068617	0.704121	0.094691
11	0.077963	0.697081	-0.00704
12	0.072178	0.787081	0.09
13	0.06901	0.821426	0.034345
14	0.059939	0.831703	0.010277
15	0.073265	0.834649	0.002946
16	0.085093	0.846472	0.011823
17	0.079859	0.829471	-0.017

TABLE 1. CUT THE DATA FOR DIFFERENT LENGTH IN ORDER TO DO THE FORECAST. CUT 1 TO 17 FROM THE END TO THE BEGINNING

IV. CONCLUSION

The result illustrates that the generative process of time series is not a random walk, they shows a fractal character, and it is a kind of biased random walk. And through using the Rescaled Range Rescaled Range analysis method, it is found a significant conclusion of fractional dynamics with long-memory in the cluster coefficient time series of complex networks in MSNs. The fractal dimension of the cluster coefficient time series is got, and it finds them fractal. So use the principle to forecast and calculate the trend of internal correlation and size of the network.

Therefore, being inspired by the research above, use this method to analyze time series and make some forecasts in complex networks, for example, it can analyze the networks traffic to forecast the trend and as the same time prepare for the congestion, also can predict some network attacks by using R/S analysis and prepare some strategies for it. As for computer operating system, R/S analysis can be used for memory scheduling and memory allocating.

But there are some obstacles in doing such researches. If the Hurst exponent is very approaching to 0.5, it may be hard to forecast the trend. And the time period may be a little large, but if reduces the length of the period the forecast.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant Nos 61501217, 61363015, 61501218 and 61262020.

REFERENCES

- [1] Xu Jun. Effectiveness of the Securities Market Analysis Efficient Market Hypothesis Testing of Shanghai Stock Exchange by the Method of Rescale Range. Wuhan University. May 2004
- [2] Zhang Xinhai. The Research on Fractal Dimension of Complex Networks. Southwest University. April 2014
- [3] Deng Pingping. Characteristics Test and Prediction of China's Security Market Based on the R/S Analysis Method. Central China Normal University. May 2012
- [4] Li Juan, Chen Yong. Rescaled Range (R/S) Analysis on Seismic Activity Parameters. ACTA SEISMOLOGICA SINICA. Vol.14 No.2 March 2001
- [5] Wang Xiaoli, Hu Baoqing, Xia Jun. R/S Analysis of Hydrological Time-Series Trend and Variation Points. Wuhan University. 35(2):10-12. 2002
- [6] Men Baohui, Liu Changming, Xia Jun. Calculation and Analysis of Runoff at Dam Sites in the First Stage Construction of the South-to-North Water Diversion Project Along the Western Line. Earth Science Frontiers. 91(91):052301-052301,2006
- [7] Zhang Yinqin, Hu Wei, Liu Junmin. The Analysis of Underground Water Level Dynamic Change Trend Based on R/S Analysis. Journal of Central South University, 2012, 43(12)
- [8] Wang Mingtao. The Nonlinear Characteristics of China's Stock Market Based on R/S Analysis. Forecasting, 2002, 21(3):42-45
- [9] Feng Lihua. The Possibility of the Typhoon Landing in Zhejiang Province Based on R/S Analysis. Acta Oceanologica Sinica, 2000(5):133-136 May 2000
- [10] Zhao Jing. The Analysis of Urbanization Effect on Climate in Lanzhou by the Method of R/S. Meteorological Monthly, 2001
- [11] Liu Yufeng, Yuan Zhihua, Sun Bo. Xi 'an Major Climate Change in Recent 15 Years R/S Analysis. Journal of Xi'an University of Arts & Science. 19(1):73-78, 2016
- [12] Wang Yuxiang, Qiao Xiuquan, Li Xiaofeng. Context Awareness of Mobile Social Network Service Selection Mechanism Research. Chinese Journal of Computers, 33(11):2126-2135, 2010
- [13] Watts, Strogatz. Collective Dynamics. Nature. 1998.
- [14] Han Hua, Liu Wanlu, Wu Lingyan. Based on Die Body Complex Network Measurement Metrics research. Chinese Journal of Physics. Vol.16, 2013
- [15] Cheng Tingting, Wang Hengshan, Liu Jianguo. General-Purpose Computing Network Graph Clustering Coefficient and Ring Method. SJCS. No.11, 2011
- [16] Pan Zaofeng, Wang Xiaofan. One Kind Can Adjust a Wide Range of Scale-Free Network Model Weighted Clustering Coefficient. Chinese Journal of Physics. Vol.55 August 2006
- [17] Zhou Mi, Cai Qingsong, Zhang Yingxin. RWP Model Based on Average Degree and Average Cluster Coefficient. JOURNAL OF BEIJING TECHNOLOGY AND BUSINESS UNIVERSITY. Vol.28, March 2010
- [18] Tan Yilong, Ying Yirong. The Build and Verification on Hurst Exponent Through R/S Analysis. Systems Engineering-Theory Methodology Application. Vol.20No.5, Sep 2011
- [19] Fu Yuhua, Fu Anjie. Using Fractal Method to Forecast the Exponent and Price of Oil Stocks. China offshore Platform. Vol 17.No. 6 May 1998