A Complex Network of Well Log based on Visibility Graph

Xiao Li[1,2], Chen lixue[1] and Xiao jingzhong[2]

[1]School of Computer Science, Southwest Petroleum University
Chengdu 637001, P.R.China
[2]School of Computer Science and Technology, Southwest University for Nationalities
Chengdu 610041, P.R.China
x266@163.com

Abstract

In this paper, we applied the visibility graph algorithm to change the well log data into a complex network. The degree distribution of this complex network existed exponential distribution, which indicated the scale free property of our network. Compared the original curve with the corresponding degree and cluster coefficient, we found that the shape information of original well log has been inherited by the network. The hierarchical coefficient of the complex network is about 2, which indicates the hierarchical structure in the network. All the data show that it is feasible and significant to translate the well log data into a complex network.

Keywords: Visibility graph, Well log, Complex network, Cluster coefficient

1. Introduction

In this article we present a complex network which is transformed from the well logs by way of visibility graph [1]. In recent years, the research of complex networks has attracted attention of many researchers from different fields, ranging from social relations, biology, information technology, to genetics [2-6]. People apply the theories and technologies of complex networks to different fields actively. Data set collected from practical applications is abstracted to complex networks. The hidden rules and principles of the data will be found according to the methods and theories of complex networks.

1.1. Graph

The first problem in the research of complex networks is the way which abstracts a data set in the real world into a complex network. Normally, we represent complex networks with the graph which include some nodes and edges, as showing in Figure 1.

It is easy to abstract data set into graph with nodes and edges in some domains, but sometimes it is very hard, the abstraction of edges is pretty nerve-racking. An edge in the graph means that some relationships exist between the two connected nodes.

The second problem is to find the answer of the following questions. Does the complex network abstracted from a data set is meaningful and necessary? Does it characterize the data set?
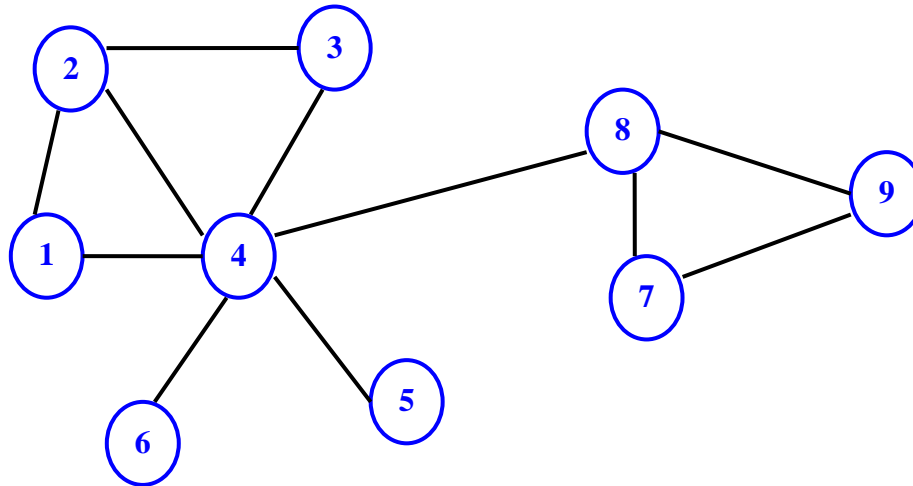
Figure 1. Graph with Nodes and Edges

The description of graph is come from graph theory originated from the famous Seven Bridges problem. The random network model(ER model) represented by Paul Erdos and Alfred Renyi in 1959 is an important development in graph theory. Aiming to construct a random network with N nodes, in every step, two nodes are connected with a fixed probability p. Many important topological properties are represented based on this ER model. The WS model was represented in 1998[24], WS model indicate that many real networks have shorter average shortest path compared with the regular networks, as well as larger cluster coefficient compared with the random networks, which is so called small-world property of networks. The scale-free network model was represented by Barabasi and Albert in 1999[25], which indicate the power low in the degree distribution.

1.2. Visibility Graph

In this article, we use the visibility graph [1, 7, 8] to translate time series data set into a complex network. Intuitively, time series is difficult to translate into a graph with nodes and edges. The visibility graph algorithm makes it feasible.

A node in the visibility graph is a point or location in Euclidean plane; an edge connected two inter-visible locations, which means there is no obstacle between the two locations.

The visibility graph algorithm is very simple and ingenious. Giving two points in a time series: A(x1, y1), B(x2, y2), point A and point B are visible, as long as they satisfy one of the following conditions:

1) Point A and point B are adjacent in time;

2) They are not adjacent, there exist intermediate point, for any intermediate point C(x3,y4), the following inequality is always satisfied

$$y3 < y2 + \frac{x2 - x3}{x2 - x1}(y1 - y2)$$

For example, in Figure 2, it shows the 20 random samples with the same time interval. The ordinate represents the sample and the abscissa represents time. In

visibility graph, the node is the sample value of each time interval, the edge is the visibility between nodes, which is showing in Figure 2 as red line. We call the visibility abiding the above expression as convex visibility. Meanwhile, the concave visibility is demanded sometimes. The concave is defined as follow, which is similar to the convex visibility.

$$y3 < y2 + \frac{x2 - x3}{x2 - x1}(y1 - y2)$$

There are many distinctive features in the visibility graph:

a) The time adjacent nodes are visible;

b) It is an undirected graph;

c) The visibility between nodes won't change by scaling the original time series;
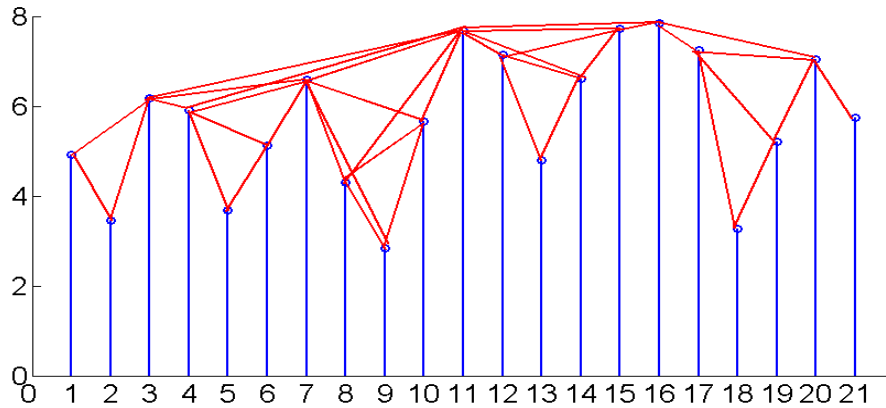
d) Nodes with large sample value have high visibility.

Figure 2. Visibility Graph Algorithm

1.3. Geophysical Logging

Typically, in geophysical logging [9-11], various logging tools are lowered into the hole by the logging cable. The tools are manufactured by many physical principles including: electrical, magnetic, acoustic, thermal, nuclear, etc. The electrical measuring instruments in the ground record various detail parameters continuously along the borehole. The well logs represented by such parameters can be used to identify underground rock formations, such as: oil, gas, water layer, coal [12-14].

The shape of well log is very important, which is the important evidence in analysis of well logs, such as the change of curve, the location of peak and trough etc. The main analysis of these curves include layer division and sedimentary facies [15]. Part data of the well log used in this paper is shown in Table1. At first we translate the gamma well log GR [16] into a complex network by visibility graph. We abstract each depth as a node since that the gamma ray log is sampled in same depth interval of 0.125m, the edge is the convex visibility between nodes according to the GR value of each node. Part of original GR log is shown in Table1, part curve of GR log is shown in Figure 3(a).

Table 1. Part Data of the Well Log

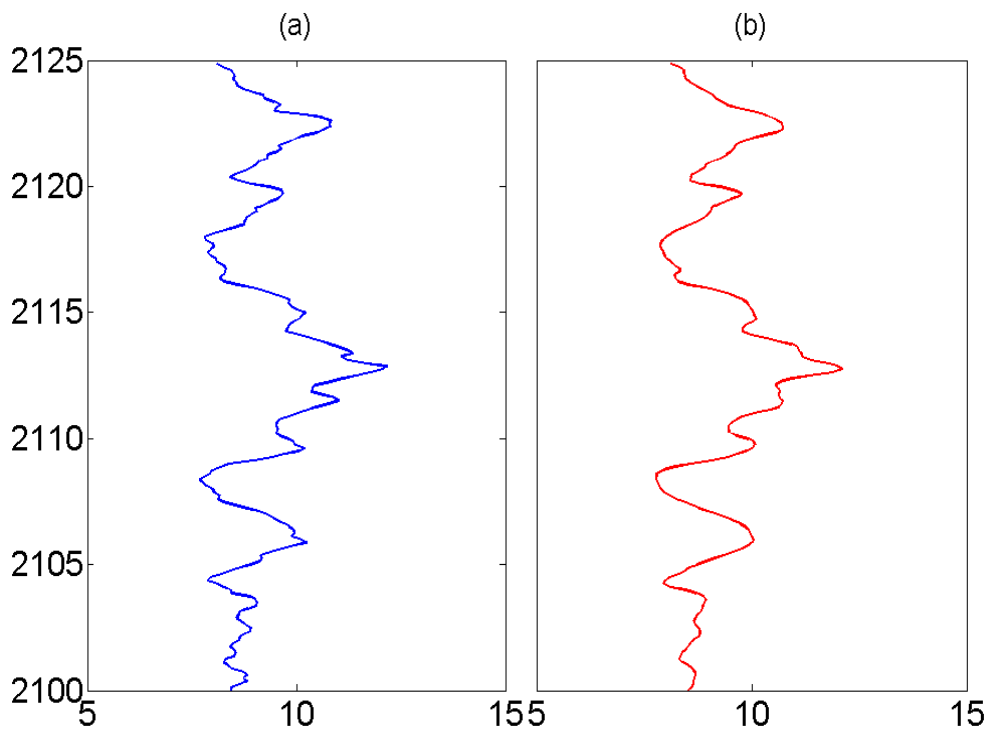| DEPTH	CALI	DT	GR	R4	RML	RMN	SP
…	…	…	…	…	…	…	…
2170.75	228.06	253.317	8.724	6.379	1.489	1.398	2.621
2170.875	227.33	296.262	8.776	6.186	1.343	1.293	2.602
2171	226.59	330.617	8.735	5.866	1.141	1.175	2.584
2171.125	226.57	357.054	8.911	5.425	0.846	1.059	2.566
2171.25	225.35	379.085	8.897	4.822	0.901	1.006	2.547
2171.375	224.33	401.115	9.057	4.275	1.133	1.174	2.529
…	…	…	…	…	…	…	…

Figure 3. Part Gamma Ray Curve

2. A Complex Network of Gamma Ray Log by Visibility Graph

Noises exist in the original curve in many reasons. So filtering which make the complex network characterize the curve more clearly is the first work that we need to do before we change it to a complex network. A low pass filter which reserved the low frequency approximate portion and got ride of the high frequency details is demanded [17]. The curve filtered is shown in Figure 3(b).

2.1. Wavelet Filter

We choose wavelet in the design of the low pass filter, which is defined in the following expression:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \varphi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j t - k)$$

Where ψ is wavelet function, $\varphi(t)$ is scaling function. In our application, we choose Daubechies as wavelet function, which is denoted as dbN, N represents the order of the wavelet. The time-domain and frequency-domain waveform of db4 wavelet is shown in Figure 4.
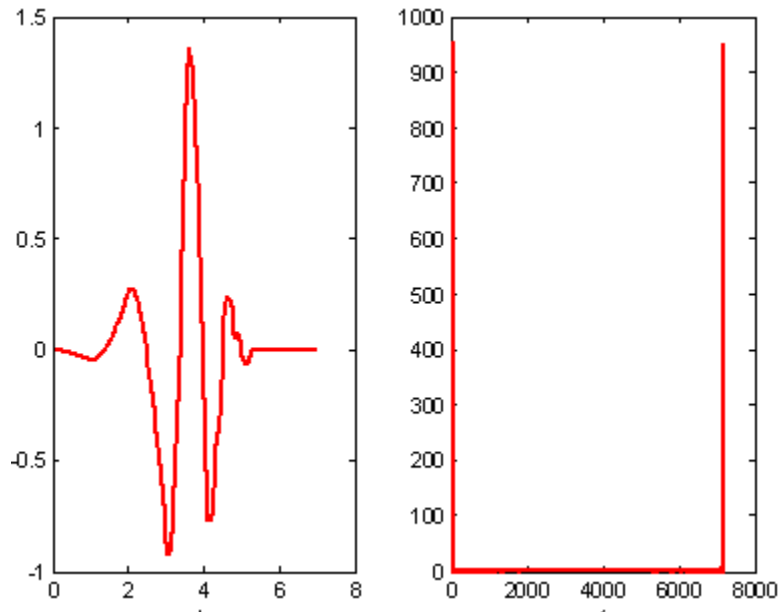

Figure 4. The Time-domain and Frequency-domain Waveform of db4 Wavelet

2.2. Degree Distribution and the Cluster Coefficient

The GR curve filtered translated into a complex network. For complex networks, degree distribution is the most important description of the statistical properties about relationships among the nodes, but also reflects the important evolution properties of the complex networks. Article of Barabasi and Albert indicates that the degree distribution of complex networks presents a power-law distribution [18-20].The degree distribution P(k) of the gamma ray log shown in Figure5. The ordinate represents the probability of the nodes with degree k; the abscissa represents the degree of the node.
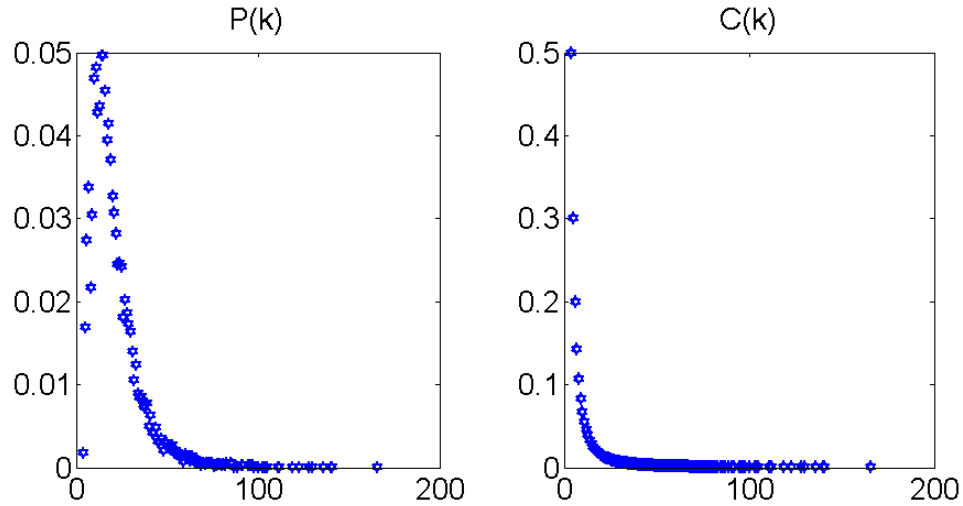
Figure 5. The Degree Distribution and Cluster Coefficient of the Complex Network

It is obviously clear that in Figure 5, degrees of most nodes are concentrated on 0-50, as degree of a small number of nodes are larger with the maximum of about 200. The probability density of smaller degree (about <20) exhibits a Poisson distribution, and the right part showed a significant exponential distribution.

Another important property of complex networks is the cluster coefficient [21-23]. Which is the proportion about that my friends are friend too. Cluster coefficient describes the proportion of the adjacent nodes are adjacent too. There are three kinds of cluster coefficient, one of which is the cluster coefficient of each node, another one is the average cluster coefficient C(k) of all nodes with the same degree value k(as shown in Figure 5), and the last one is the cluster coefficient of the whole complex network, which is the average of the cluster coefficient of each node. These three kinds of cluster coefficient indicate different properties of the complex networks. They are the important indicators in the research of complex network.

From the perspective of geometrical, the expression of the cluster coefficient of each node is shown as follow:

$$C_i = \frac{N_\Delta(i)}{N_\Lambda(i)}$$

Where $N_\Delta(i)$ represents the total number of the "triangle" contained node i in the complex network. In this expression, triangle means the adjacent nodes of node i are adjacent too.

$$N_\Delta(i) = \sum_{k>j} a_{ij}\, a_{jk}\, a_{ik}$$

Where a_{ij} is equal to 1 as node i and node j are adjacent, otherwise 0.

NΛ(i) is the total number of "triplet" of node i. in which node j and node k are both adjacent node of node i, it doesn't care about whether node j and node k are adjacent.

$$N_\Lambda(i) = \sum_{k>j} a_{ij} a_{ik}$$

The average cluster coefficient of all nodes with degree value of k is the average of the cluster coefficient of each node with degree value of k. Generally, the node with large degree value have lower cluster coefficient, which always are hub nodes; the average cluster coefficient indicate the cluster extent of the complex networks.

2.3. The Hierarchical Coefficient

Many works have been made for discovering the hierarchical structures (or named modules) in the complex networks. They found that there existed hierarchical structures in the networks as the average cluster coefficient of all nodes with degree k proportional to k-β, β is so called hierarchical coefficient. β=2 in our gamma ray log. The log-log-graph of C(k) and k-2 is shown in Figure 6, Which means the existence of hierarchical structure in the network, it is reasonable as the well log characterized geological layers indeed.
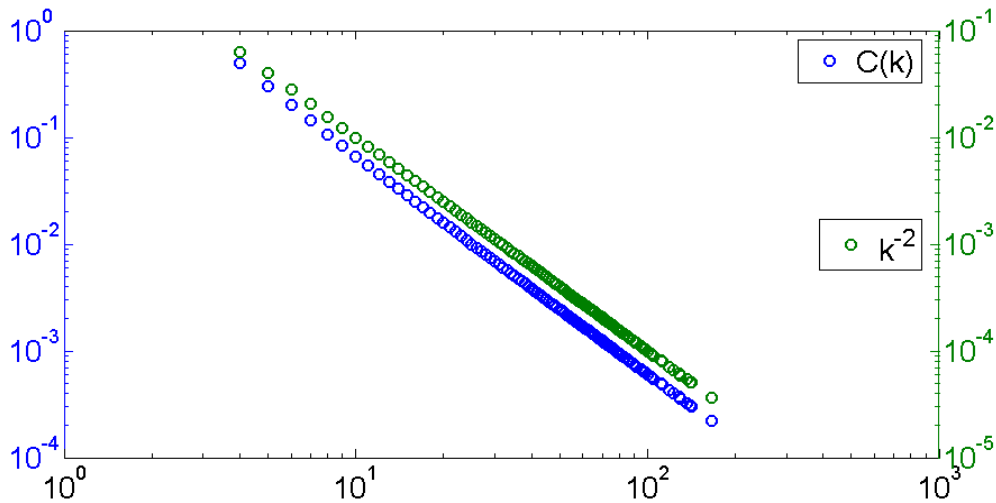

Figure 6. The log-log-graph of C(k) and K^{-2}

As the curve shape of the original well logs contain a lot of important content, did degree and clustering coefficient inherit these shape information? We compared the original curve and the degree and cluster coefficient corresponding, part of that shown in Figure 7. In Figure 7, there have 3 graphs, the left one is the GR value corresponding to depth, the middle one is the degree of each depth, the right one is the cluster coefficient of each depth, and the ordinates of three graphs are the same---depth.

The troughs in original curve always indicate lower degree and high cluster coefficient, the peak in original curve always indicate higher degree and lower cluster coefficient. The troughs and peaks are important gist in the research of well log. The complex network inherited the shape information of the well log.
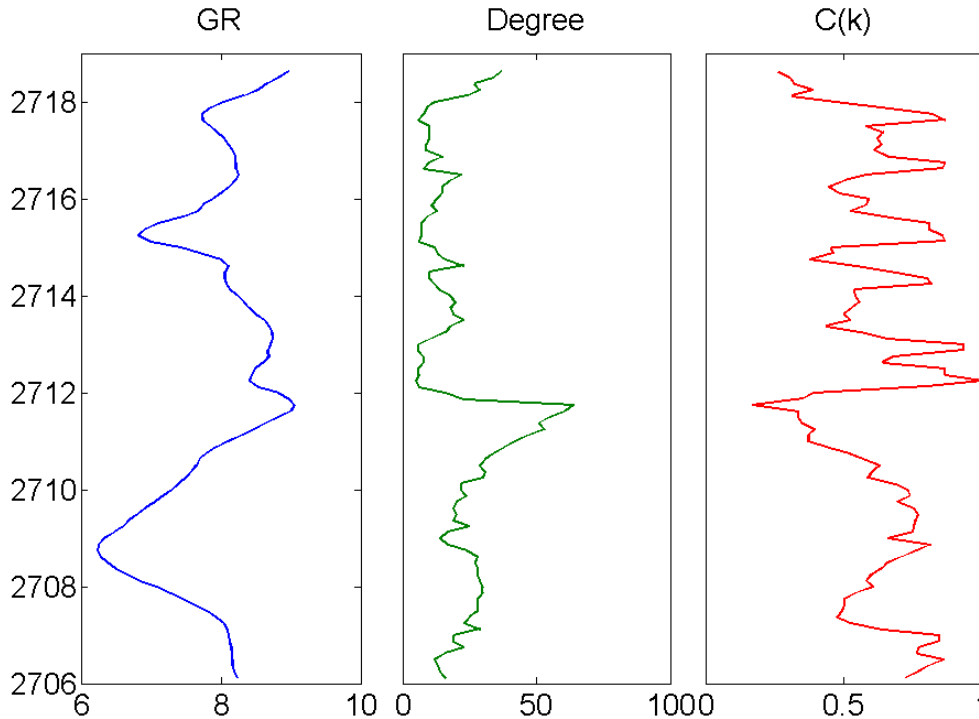
Figure 7. Comparing of the Original Curve, Degree and Cluster Coefficient

2.4. Multiple-edges Graph

In fact, in the analysis of the well logs, it is unreliable based on only a single curve; multiple curves are needed to combine and product a meaningful result. In this comprehensive analysis, the shape of each curve is still the key point.

Many works about complex networks are based on single-edge graph, a bunch of topological properties are made. However, the connection between nodes is not single in many complex networks in the real world. In traffic networks, a connection between nodes which represent the location indicate the reachable traffic routes between different place, which can be all kinds of traffic routes including highway, railroad, waterway, aircraft route. Similarly, the relationship between people is multiple in the social networks, a couple may be classmates or colleague simultaneously. Multiple-connection is common in many real complex networks.

The previous part in this paper has talked about complex network based on a single curve, every node is a depth point, and an edge is the visibility of the corresponding GR value of two depths. Aiming to simplify the problem, and also a better understanding of the node and edge, Figure 1 previous is a graph with single edge between nodes. The next problem is, can we construct the complex networks contained multiple curves? In this paper, our answer is multiple-edges graph. An example of the multiple-edges graph is shown in Figure 8. For convenience of description, assuming two kinds of edge marked by black and red existed in the graph.
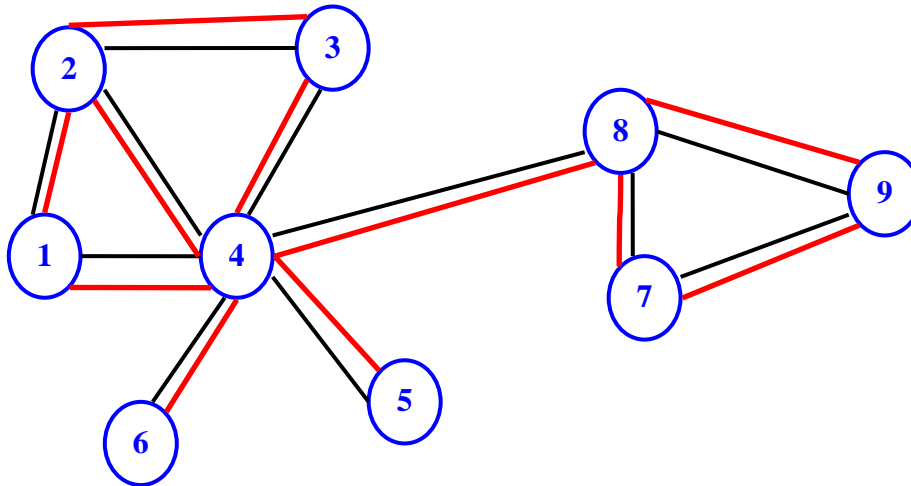
Figure 8. An Example of Multiple-edges Graph

The degree of each node in the complex networks represented by multiple-edges graph is so called comprehensive degree (CD), the definition of which is the following expression:

CD=f($\omega 1D1,\omega 2D2,\ldots,\omega iDi$)

Where Di indicate the degree value of a node corresponding the connection i. ω i indicate the weight of connection i. The expression showed that CD is a function about $\omega 1D1,\omega 2D2,\ldots,\omega iDi$.

Similarly, the definition of comprehensive cluster coefficient (CC) is the following expression, where Ci indicate the cluster coefficient of a node corresponding the connection i. ωi indicate the weight of connection i. CC is a function about $\omega 1C1,\omega 2C2,\ldots,\omega iCi$.

CC=f($\omega 1C1,\omega 2C2,\ldots,\omega iCi$)

The function and ω varies with different applications. In our case, we construct a complex network contained both GR curve and SP curve with node represented depth and edges represented the visibility of GR value and SP value. The function of CD and CC is defined as follow:

$$CD= \frac{D_{GR}}{\max(D_{GR}) - \min(D_{GR})} + \frac{D_{SP}}{\max(D_{SP}) - \min(D_{SP})}$$

$$CC= \frac{C_{GR}}{\max(C_{GR}) - \min(C_{GR})} + \frac{C_{SP}}{\max(C_{SP}) - \min(C_{SP})}$$

Parts of the comprehensive degree and comprehensive cluster coefficient curves are shown in Figure 9.
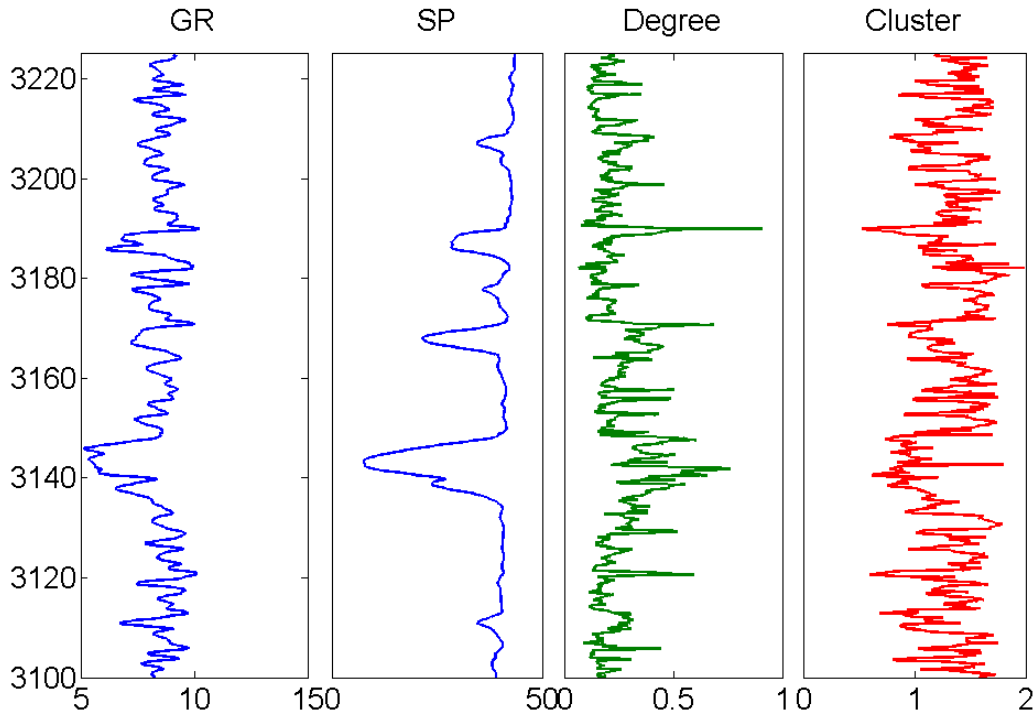
Figure 9. Part of the Comprehensive Degree and Cluster Coefficient

Similar to the degree and cluster coefficient of single curve, the comprehensive degree (CD) and cluster coefficient (CC) of two curves (GR and SP) has the same property. As we can see in Figure 9, there have a bunch of high comprehensive degree value and low comprehensive cluster coefficient around the trough of original curves, and a bunch of low comprehensive degree value and high comprehensive cluster coefficient in the peak of original curves.

3. Solution

We applied the visibility graph algorithm to change the well log data into a complex network. The degree distribution of this complex network existed exponential distribution, which indicated the scale free property of our network. The average cluster coefficient of each node is equal to 0.6475. Compared the original curve and the corresponding degree and cluster coefficient, we found that the shape information of original well log has been inherited by the network. We calculate the hierarchical coefficient of the complex network is about 2, which indicates the hierarchical structure in the network. All the data show that it is feasible and significant to change the well log data into a complex network.

4. Future Work

a) Analyzing the complex network combined much more logs.
b) Discover the deep relationship between the log and the cluster coefficient and degree and so.
c) Appling the complex network to stratify well log and divide sedimentary facies.

Acknowledgements

At the point of finishing this paper, I'd like to express my sincere thanks to all those who have lent me hands in the course of my writing this paper. First of all, I'd like to take this opportunity to show my sincere gratitude to my supervisor, Ms.chen lixue, who has given me so much useful advices on my writing, and has tried him best to improve my paper. Secondly, I'd like to express my gratitude to my classmates who offered me references and information on time. Last but not the least, I'd like to thank those leaders and working staff. Without their help, it would be much harder for me to finish my study and this paper.

References

[1] L. Lacasa, B. Luque and F. Ballesteros, "From time series to complex networks: The visibility graph", Proceedings of the National Academy of Sciences, vol. 105, no. 13, (2008), pp. 4972-4975.

[2] S. Boccaletti, V. Latora and Y. Moreno, "Complex networks: Structure and dynamics", Physics reports, vol. 424, no. 4, (2006), pp. 175-308.

[3] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems", Personal and ubiquitous computing, vol. 10, no. 4, (2006), pp. 255-268.

[4] E. Bullmore, "Economical Network Models of the Connectome in Schizophrenia", Schizophrenia Research, (2012), 136: S28.

[5] A. L. Barabási, "Network Science: Understanding the Internal Organization of Complex Systems", (Invited Talk)[C]//2012 AAAI Spring Symposium Series, (2012).

[6] Y. Chen, J. Zhu and P. Y. Lum, "Variations in DNA elucidate molecular networks that cause disease", Nature, vol. 452, no. 7186, (2008), pp. 429-435.

[7] E. Welzl, "Constructing the visibility graph for n-line segments in O (n< sup> 2</sup>) time", Information Processing Letters, vol. 20, no. 4, (1985), pp. 167-171.

[8] M. Pocchiola and G. Vegter, "Computing the visibility graph via pseudo-triangulations", Proceedings of the eleventh annual symposium on Computational geometry, ACM, (1995), pp. 248-257.

[9] E. Welzl, "Constructing the visibility graph for n-line segments in O (n< sup> 2</sup>) time", Information Processing Letters, vol. 20, no. 4, (1985), pp. 167-171.

[10] S. R. Dallimore, "Scientific results from the Mallik 2002 gas hydrate production research well program", Mackenzie Delta, Northwest Territories, Canada [M], Ottawa, Ontario, Canada: Geological Survey of Canada, (2005).

[11] M. Jin, M. Delshad and V. Dwarakanath, "Partitioning tracer test for detection, estimation, and remediation performance assessment of subsurface nonaqueous phase liquids", Water Resources Research, vol. 31, no. 5, (1995), pp. 1201-1211.

[12] G. B. Asquith, D. Krygowski and C. R. Gibson, "Basic well log analysis", American Association of Petroleum Geologists, (2004).

[13] O. Serra, "Fundamentals of well-log interpretation", Amsterdam: Elsevier, (1984).

[14] J. C. Davis and R. J. Sampson, "Statistics and data analysis in geology", (2002).

[15] H. Sakai, "Stratigraphic division and sedimentary facies of the Kathmandu Basin Group, central Nepal", J Nepal Geol Soc, vol. 25, (2001), pp. 19-32.

[16] M. Tzortzis, E. Svoukis and H. Tsertos, "A comprehensive study of natural gamma radioactivity levels and associated dose rates from surface soils in Cyprus", Radiation protection dosimetry, vol. 109, no. 3, (2004), pp. 217-224.

[17] H. Qiu, J. Lee and J. Lin, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics", Journal of Sound and Vibration, vol. 289, no. 4, (2006), pp. 1066-1090.

[18] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks", Reviews of modern physics, vol. 74, no. 1, (2002), pp. 47.

[19] A. L. Barabási and R. Albert, "Emergence of scaling in random networks", science, vol. 286, no. 5439, (1999), pp. 509-512.

[20] R. Albert, H. Jeong and A. L. Barabási, "Error and attack tolerance of complex networks", Nature, vol. 406, no. 6794, (2000), pp. 378-382.

[21] E. Ravasz and A. L. Barabási, "Hierarchical organization in complex networks", Physical Review E, vol. 67, no. 2, (2003), pp. 026112.

[22] C. Song, S. Havlin and H. A. Makse, "Self-similarity of complex networks", Nature, vol. 433, no. 7024, **(2005)**, pp. 392-395.

[23] G. Palla, I. Derényi and I. Farkas, "Uncovering the overlapping community structure of complex networks in nature and society", Nature, vol. 435, no. 7043, **(2005)**, pp. 814-818.

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks", Nature, vol. 393, no. 6684, **(1998)**, pp. 440-442.

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

Authors

Xiao Li, she received the BS degree in computer science from Southwest Petroleum University in 2010. She is presently employed as teacher at School of Computer Science and Technology, Southwest University for Nationalities. Her research interests are in the areas of network architecture, Algorithm Design and Analysis, and information systems. She is an active researcher coupled with the vast (13 years) teaching experience. She has published more than 10 research articles in reputed international journals of computer sciences.

Xiao Jingzhong, he received his BS degree in comuter science (2003) from University of Electronic Science and Technology of China . Now he is full associate professor of School of Computer Science and Technology, Southwest University for Nationalities. He is an active researcher coupled with the vast (17 years) teaching experience. His current research interests include different aspects of Parallel Algorithms and Cloud Computing.