International Core Journal of Engineering Volume 6 Issue 3, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

Diabetes Attention Analysis: A Network Analysis Approach

Panpan Lv[1], Yunxia Chen[2]*, Junping Su[2], Qian Gao[2], Ying Ma[2]

[1] Department of Endocrinology, Cangzhou people's Hospital, Chengde Medical College, Cangzhou 061000, Hebei Province, China;

[2]Cangzhou people's Hospital, Cangzhou 061000, Hebei Province, China.

[a]*1158743132@qq.com, [b]Conf_18031798228@126.com

Abstract

Diabetes is a group of metabolic diseases characterized by hyperglycemia. However, some people have some deficiencies in diabetes now, but how to improve everyone's understanding of diabetes needs further research, so in this article we pass the national The analysis of diabetes concerns was analyzed, and the Baidu index related to diabetes was analyzed through the principle of time series data visualization. We found that the attention of diabetes became an exponential distribution, and through the analysis of its centrality, it was found that the Baidu index can find the key At the time point, we found that the World Diabetes Day is an important time through the Sosso index of the past two years, which also confirms the effectiveness of our method.

Keywords

Diabetes; Attention; Visibility graph.

1. Introduction

All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper. Use *italic* for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below).

Do not number your paper: All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question.

In order to mine more information about time series, it is possible to study the time series from the view angle of complex networks. This transformation has been studied a lot. Zhang Xiaolian studies time series by constructing complex networks with pseudo-periodic time series, and studies The relationship between the topology of constructing the network and the dynamics of the original time series [3]. Xu et al. Constructed a method to map the time series to the nearest neighbor network [4]. Marvan et al. Introduced the concept of recursive networks [5]. Lucasa Introduced the visibility graph (VG) algorithm for time series [6]. The visibility graph algorithm not only retains the characteristics of time series, but also relates the time series to the characteristics of complex networks [7], using Visibility graph algorithms, According to the characteristics of the complex network constructed, whether the system under study is deterministic, random, etc. [8]. At present, time series analysis based on the VG method has been applied to different fields. Xinghua Fan uses the VG method to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

analyze seven carbons in China Similarity and heterogeneity of pilot market price time series [9] .Cheng Zhou addresses the characteristics of shield tunneling parameters by introducinga visibility graph model implemented in a complex network of shield tunneling in metro construction [10]. Peng-Fei Dai maps the four economic policy uncertainty indices of the United States and China into a complex network and studies the topology of the network through the VG method [11].

From the perspective of time series, this article uses the principle of vg to convert time series data into a complex network.From the basic parameters, it analyzes the degree distribution, centrality value, and community division of the complex network, thereby mining more hidden information in time series data. The information of Baidu search index data from January 1, 2011 to December 31, 2018 can be more comprehensively understood through the principle of vg into a complex network. Through the analysis of degree distribution, we find that The degree distribution of the comprehensive Baidu search on the PC and mobile terminals is a power-law distribution, which means that there are very few abnormal peaks and troughs, and 80% of the search times are in the normal range;

2. model

2.1 Time series visualization principle

This paper uses the time series visualization method proposed by Lacasa et al. (2008) for network construction [16]. To build a network for each subsystem in the value chain transformation system, first, the subsystem$x(t)$The discrete time series data corresponds to the nodes of the network, and the connected edges of the network are constructed according to visual criteria: in the time series $(x(t)$Any two points of data in$)(t^a, x^a)$with(t^c, x^c)You can create connecting edges by seeing between them, and any point between two points(t^b, x^b),when$t^a < t^b < t^c$All meet

$$x^b < x^a + (x^a - x^c)\frac{t^a - t^b}{t^c - t^a} \tag{1}$$

As shown in Figure 1, the height of the histogram bar in Figure 1 represents the data value at each time point.If the tops of the two histogram bars are visible to each other, the corresponding two points are connected in the network in the figure.

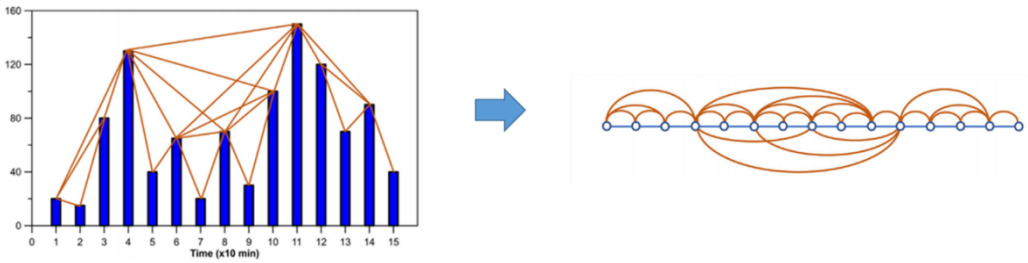

Figure 1 visibility graph

Second, an adjacency matrix is constructed based on the time series nodes and edges, and a network graph is formed, as shown in Figure 1.

2.2 Centrality analysis

In this paper, starting from the degree centrality, intermediary center, and eigenvector centrality of a complex network, a quantitative analysis of the search index network of keywords for diabetes in China from January 1, 2018 to November 25, 2019.

2.2.1 Degree Centrality

Degree centrality is the most direct measure of centrality in network analysis. The greater the degree of a node, the higher the degree of centrality of the node, and the more important the node is in the network. A graph with n nodes G = (V, E), and the degree centrality $C_D(v)$ of the node v is:

$$C_D(v) = \frac{\deg(v)}{n-1} \tag{2}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

For graph g, the complexity of the computational centrality is Θ (v^2) in the dense adjacency matrix representation and Θ (e) in the sparse matrix representation, where v is all the points and e is all the edges .

The definition of centrality can be extended (from the nodes) to the graph. Let v * be the node with the highest degree of centrality in G. Define X: = (Y, Z) to maximize the amount of n nodes connected to the graph (H) (Let y * be the node with the highest degree of centrality in X):

$$H = \sum_{j=1}^{|y|}[C_D(y^*) - C_D(y_j)] \tag{3}$$

The degree centrality of graph g is defined as follows:

$$C_D(v) = \frac{\sum_{j=1}^{|Y|}[C_D(y*) - C_D(y_j)]}{H} \tag{4}$$

When graph G has one node connected to all other nodes, and all other nodes only connect to this one central node, H is the largest (star graph). In this case H = (n − 1) (n − 2), the degree centrality of graph G can be simplified as:

$$C_D(G) = \frac{\sum_{j=1}^{|V|}[C_D(v*) - C_D(v_j)]}{H} \tag{5}$$

2.2.2 Closeness Centrality

In topology and related mathematical neighborhoods, compactness is a basic concept in topological space. Intuitively, when two sets are arbitrarily close, we say that they are compact. This concept is within a defined space Element distance is easily defined in the metric space, but it can be generalized to a topological space without a specific metric distance.

In graph theory, compactness is a measure of the centrality of a node in the graph. Nodes that are "shallower" (that is, have shorter geodesic distances) than other nodes have higher compactness. In network analysis, compactness tends to represent the minimum path length, because this gives higher values to more central nodes, and it is usually associated with other metrics (e.g., degree). In network theory, compactness Is a complex metric of centrality. It is defined as the average geodesic distance of the node v to other reachable nodes (such as the shortest path):

$$\frac{\sum_{t \in V \setminus v} d_{G(v,t)}}{n-1} \tag{6}$$

Where n≥2 is the size of the connected part V in the network starting from v. Tightness can be regarded as a measure of the length of time it takes to propagate information from a given node to other reachable nodes in the network.

Some people define the closeness as the reciprocal of this amount, but the two ways of transmitting information are the same (here, the speed is evaluated instead of time). The tightness C of the closeness node v $_C$ (v) is to all other The inverse of the geodesic distance sum of node V:

$$C_C = \frac{1}{\sum_{t \in V \setminus v} d_G(v,t)} \tag{7}$$

Tightness can be obtained by different methods and algorithms.Noh and Rieger (2003) proposed random-walk centrality, which is a measure of the speed of randomly propagating information from other nodes in the network to a (given) node-random. -Walk version for tight center.

Another kind of tightness measure is Stephenson and Zelen's (1989) information centrality, which is somewhat similar to Noh and Rieger's method. In essence, it is the harmonic average length of the path ending in node i. When i has many This length is smaller when the short path connects other nodes.

To measure the vulnerability of the network, Dangalchev (2006) modified the definition of compactness so that it can be applied to non-connected graphs, and the overall compactness is easier to calculate:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

$$C_c(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)} \tag{8}$$

Opsahl (2010) proposed an extension to disconnected networks.

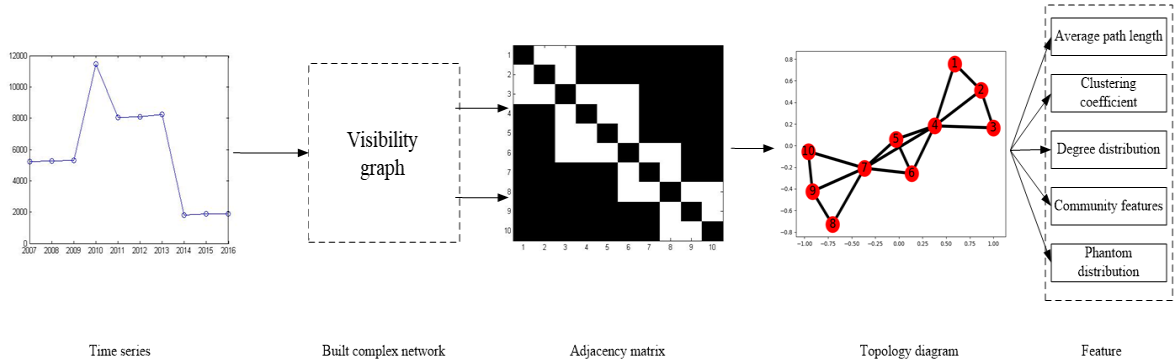

Figure 2 Time series network construction process and network feature extraction

2.3 Community division

Since Barabasi and Albert [13] pioneered research methods for complex networks, researchers in various fields have begun to look at the current world from the perspective of complex networks. According to network theory, we build actual data into complex networks. Describe these data and divide them into different categories. In the field of complex networks, community structure is a very interesting phenomenon in complex networks. How to find the community structure in the network has become an important research topic. In general, compared to nodes outside a group, a group of nodes in the network has significantly more connections within the group than connections outside the group, which means that the association within the group is significantly higher than the group External relevance [16]-[17]. It is still a very important scientific problem to find the community structure from a large-scale network. At present, it is generally possible to optimize the specific community by optimizing some specific indicators, of which The modularity Q proposed by Newman [18] and other methods is the most widely used method at present, and its specific form is shown in formula (9):

$$Q = \sum_{i=1}^{K} \left[\frac{l_i^{in}}{L} - \left(\frac{d_i}{2L} \right)^2 \right] = 1 - \frac{L_{inter}}{L} - \frac{1}{K} - \frac{1}{K} \sum_{j=2}^{K} \sum_{k=1}^{j-1} \left(\frac{d_j - d_k}{2L} \right) \tag{9}$$

Where k is the number of clusters and l is the total number of edges in the network, l_i^{in} with $d_i = l_i^{in} + l_i^{inter}$ Represents the number of edges in the cluster of cluster i and the total number of edges, L_{inter} Represents the total number of edges between clusters. The degree of modularity q is defined as "the number of edges located inside the cluster, minus the expected value of the same number of edges that fall in a random network without considering the cluster structure". The value of q can indicate cluster The quality of the structure, the larger the value of q, the more obvious the clustering structure in the network.

3. Data Sources

The data comes from Baidu Search Index.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

4. Data analysis

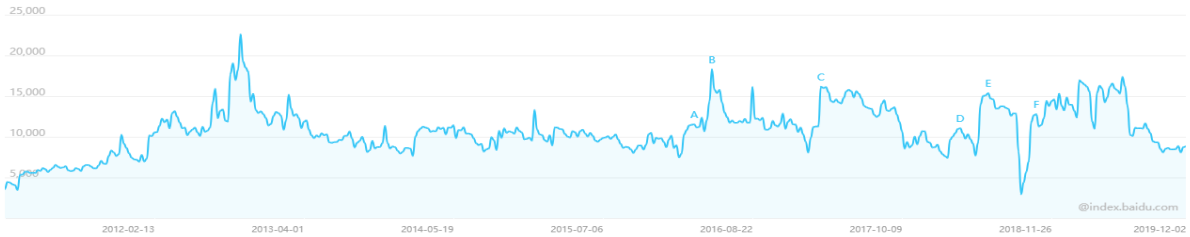

Figure 3 The relationship between the integration time and the number of searches

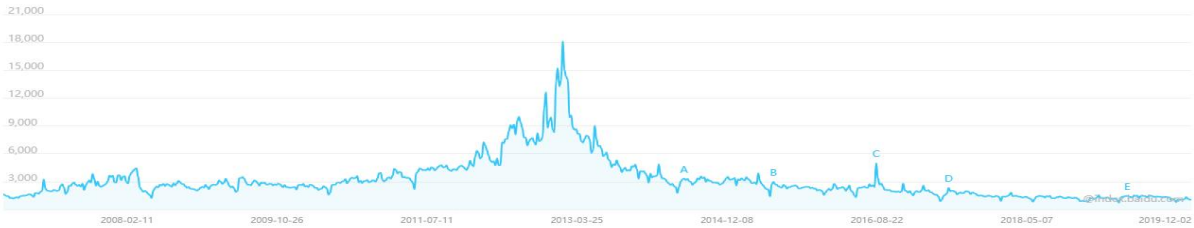

Figure 4 Relationship between computer time and search times

Figure 5 Relationship between mobile time and search times

5. Result analysis:

As shown in the section 2.1, we aim to convert time series data into a complex network form. The parameters of the network are shown in Table 2 below. In the table, we find that the number of edges and the average degree of the mobile end in the three networks are large. The side shows that there are more peaks and valleys at the mobile end. The diameters of the three networks are 9, 8, 10, which means that the shortest number of sides between the two time points with the longest distance are 9, 10, and 9. The path lengths are 4.721, 4.265, and 4.971, which indicates that the average between any two time points must pass 4.721, 4.265, and 4.971 edges to reach each other. The average distance can also indicate that the degree of attention passes through 4.721, 4.265, and 4.971 Connections can be made at the point of time.

Table 2 Data analysis

	综合	PC	移动
Edge	2978	2474	3017
Average degree	8.424	6.999	8.535
Diameter	9	8	10
Average path length	4.721	4.265	4.971
Density	0.012	0.01	0.012
Modularity	0.788	0.789	0.795
Number of communities	10	13	12
Cluster coefficient	0.794	0.789	0.754

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

First of all, we focus on the degree distribution of these three complex network, Figure 5 illustrates the cumulative distribution *P(k)* of networks mapped from the ALL, PC and Mobile terminal series. Clearly, the VG graph complies with some kind of power law tail $P(k) \sim k^{-\alpha}$. And the tendency of these three data is very similar to each others, which also exhibit the fat-tail phenomenon.

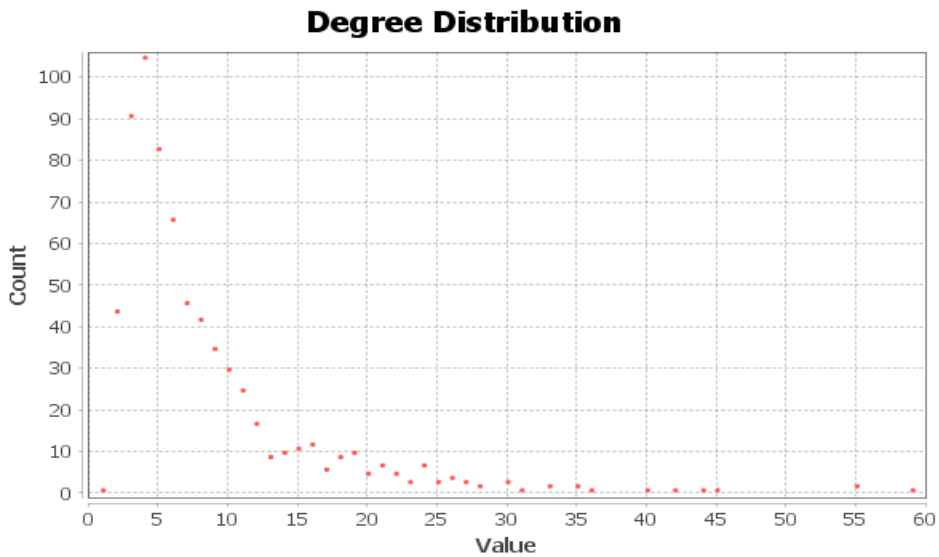

Figure 6 degree distribution of all

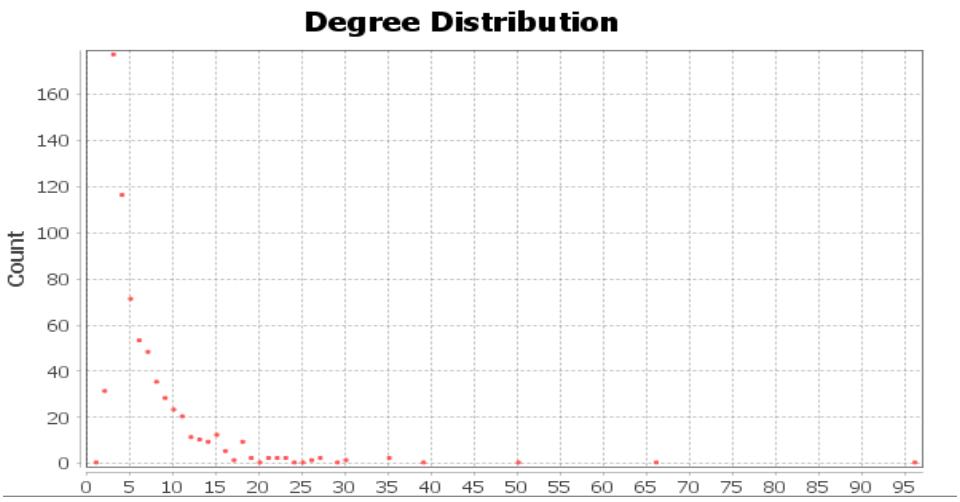

Figure 7 degree distribution of computer

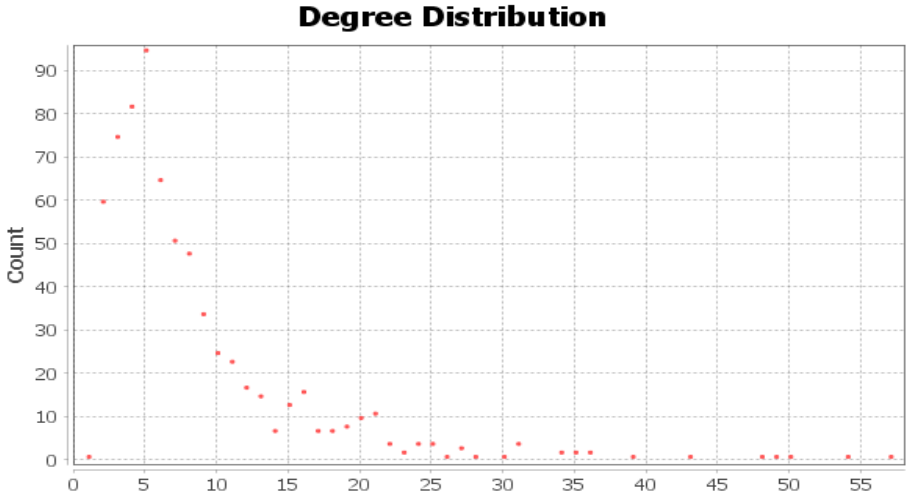

Figure 8 degree distribution of Mobile phone

ISSN: 2414-1895
DOI: 10.6919/ICJE.202003_6(3).0014

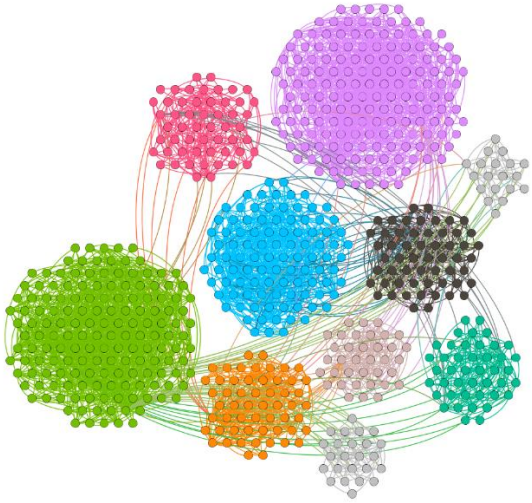

Figure 9 community of all

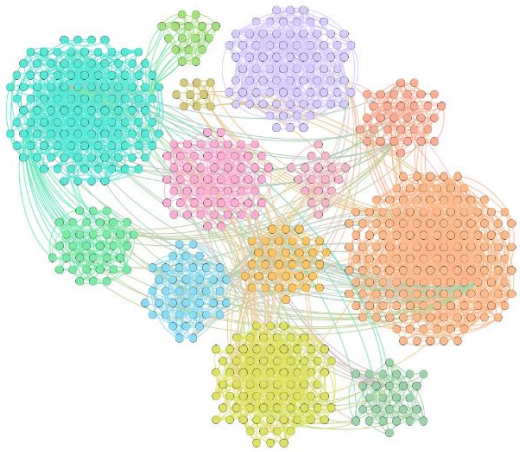

Figure 10 community of computer

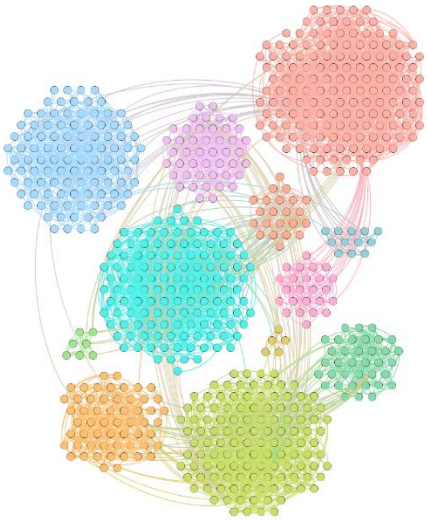

Figure 11 community of Mobile phone

ISSN: 2414-1895 DOI: 10.6919/ICJE.202003_6(3).0014

6. Conclusion

The results of this work show that the multifractal analysis of VGs from diabetes time series is a suitable tool to describe the nonlinear dynamics of diabetes. VGs have proven to have advantages such as: i) their topology inherits the features of the associated time series, which ends up resulting on supplementary information through the degree distribution; ii) and also, this novel connection between time series and complex networks opens a broad range of possibilities within the study of complex signals.

References

[1] Yueh L . Patent laws and innovation in China[J]. International Review of Law & Economics, 2009, 29(4):304-313.

[2] Wu J, Hu D, Wang Q. Role of education for spillover effects of FDI on innovation in China: Evidence from the 1998–2008 provincial panel data[C]//2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management. IEEE, 2010: 454-457.

[3] J. Zhang, M. Small, Complex network from pseudoperiodic time series: Topology versus dynamics, Phys. Rev. Lett. 96 (23) (2006) 238701, http://dx.doi.org/10.1103/PhysRevLett.96.238701.

[4] J. Zhang, J.-F. Sun, X.-D. Luo, K. Zhang, T. Nakamura, M. Small, Characterizing pseudoperiodic time series through the complex network approach, Physica D 237 (2008) 2856–2865, http://dx.doi.org/10.1016/j.physd.2008.05.008.

[5] X.-K. Xu, J. Zhang, M. Small, Superfamily phenomena and motifs of networks induced from time series, Proc. Natl. Acad. Sci. USA 105 (50) (2008) 19601–19605, http://dx.doi.org/10.1073/pnas.0806082105.

[6] N. Marwan, J.F. Donges, Y. Zou, R.V. Donner, J. Kurths, Complex network approach for recurrence analysis of time series, Phys. Lett. A 373 (46) (2009) 4246–4254, http://dx. doi.org/ 10.1016/j. physleta.2009.09.042.

[7] R.V. Donner, Y. Zou, J.F. Donges, N. Marwan, J. Kurths, Recurrence networks: A novel paradigm for nonlinear time series analysis, New J. Phys. 12 (2010) 033025, http://dx.doi.org/10.1088/1367-2630/12/3/033025.

[8] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuño, From time series to complex networks: The visibility graph, Proc. Natl. Acad. Sci. USA 105 (13) (2008) 4972–4975, http://dx.doi.org /10.1073/ pnas.0709247105.

[9] Li, X., Yang, D., Liu, X., Wu, X.M., 2012. Bridging time series dynamics and complex network theory with application to electrocardiogram analysis. IEEE Circuits Syst. Mag. 12, 33–46.

[10] Gao, Z.K., Cai, Q., Yang, Y.X., Dang, W.D., Zhang, S.S., 2016. Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series. Sci. Rep. 6

[11] Zhang, J., & Small, M. (2006). Complex network from Pseudo periodic time series : Topology versus dynamics. Physical Review Letters, 96, 238701.

[12] Xu, X., Zhang, J., & Small, M. (2008). Superfamily phenomena and motifs of networks induced from timeseries. PNAS, 105, 19601–19605.

[13] Xiao, D., & Wang, J. (2014). Graph based and multifractal analysis of financial time series model bycontinuum percolation. International Journal of Nonlinear Sciences and Numerical, 15, 265–277.

[14] Lacasa, L., Luque,B., Ballesteros, F., Luque, J.,&Nuno, J. C. (2008). From time series to complex networks: The visibility graph. PNAS, 105, 4972–4975.

[15] Lacasa, L., Luque, B., Luque, J., & Nu, J. C. (2009). The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. Europhysics Letters, 86, 30001.

[16] Pierini, J. O., Lovallo, M., & Telesca, L. (2012). Visibility graph analysis of wind speed records measured in central Argentina. Physica A, 391, 5041–5048.