

Fuzzy weighted recurrence networks of time series Chock

Tuan Pham

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-152587>

N.B.: When citing this work, cite the original publication.

Pham, T., (2019), Fuzzy weighted recurrence networks of time series Chock, *Physica A*, 513, 409-417.
<https://doi.org/10.1016/j.physa.2018.09.035>

Original publication available at:

<https://doi.org/10.1016/j.physa.2018.09.035>

Copyright: Elsevier

<http://www.elsevier.com/>



Fuzzy Weighted Recurrence Networks of Time Series

Tuan D. Pham

Department of Biomedical Engineering

Linköping University, Linköping 58183, Sweden

Phone: +46-13-286778

E-mail: tuan.pham@liu.se

ABSTRACT

The concept of networks in the context of graph theory delineates a wide variety of real-life complex systems. The theory of networks finds its applications very useful in many scientific and intellectual domains. Weighted networks can characterize complex statistical graph properties, particularly where node connections are heterogeneous. A framework of fuzzy weighted recurrence networks of time series is presented in this letter. Popular graph measures including the average clustering coefficient and characteristic path length of fuzzy weighted recurrence networks are shown to be more robust than those of unweighted recurrence networks derived from binary recurrence plots.

Keywords: Time series; Nonlinear dynamics; Fuzzy recurrence plots; Fuzzy weighted recurrence networks.

1 Introduction

Applications of complex networks are pervasive in many disciplines, including natural science, computer science, engineering, life science, medicine, health, sociology, economics, and finance. Stemming from the concept of recurrence plots [1, 2], research into recurrence networks of time series has opened a new direction for exploring and gaining insight into the behavior of complex systems [3, 4]. Recurrence networks constructed from recurrence plots are unweighted networks. However, the existence of weighted networks is widespread in many natural relationships [5, 6, 7], where weights represented in real-life networks are heterogeneous. Thus, the use of network weights is useful for recognizing links of varying importance and influence in complex systems [7]. Yet relatively little effort has been spent on the development of methods for weighted recurrence networks. It appears that there is only one method for constructing multivariate weighted recurrence networks [8], in which the edge weights are obtained using cross recurrence plots of multiple dynamical systems; but none for univariate weighted recurrence networks.

Based on the concept of fuzzy recurrence plots [9], the formulation of univariate fuzzy weighted recurrence networks of time series is introduced herein. It is pointed out herein that the univariate scalable recurrence networks reported in [10] are also derived from fuzzy recurrence plots, but these networks are unweighted and therefore a different development with respect to the work addressed herein. The rest of this letter is organized as follows. Section 2 briefly reviews the technique for constructing an unweighted recurrence network. Section 3 presents the formulation of a fuzzy weighted recurrence network. Two popular graph measures known as the clustering coefficient and characteristic path length for unweighted and

weighted networks are presented in Section 4. Results obtained from unweighted recurrence networks and fuzzy weighted recurrence networks are presented in Section 5. Comparisons and discussion of the graph measures of complex networks obtained from unweighted recurrence networks and fuzzy weighted recurrence networks are addressed in Section 6. Finally, Section 7 is the conclusion of the research findings.

2 Unweighted recurrence networks

An unweighted recurrence network (RN) represented by its adjacency matrix \mathbf{A} is defined as [11]

$$\mathbf{A} = \mathbf{R} - \mathbf{I}, \quad (1)$$

where \mathbf{R} and \mathbf{I} are the $N \times N$ recurrence matrix of the recurrence plot and $N \times N$ identity matrix, respectively.

A recurrence matrix is constructed by considering the recurrences of the phase-space states $\mathbf{X} = \{\mathbf{x}\}$. In other words, a recurrence matrix is a visualization of the number of times the phase space trajectory of the dynamical system visits the same location in the phase space it has visited before. Hence, a recurrence plot, denoted as $RP = [R_{ij}]$ is defined as [2]

$$R_{ij} = \Theta(\phi - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i, j = 1, \dots, N, \quad (2)$$

where ϕ is the recurrence threshold, and Θ is the Heaviside step function, that is $\Theta = 1$ if $\|\mathbf{x}_i - \mathbf{x}_j\| \leq \phi$, or $\Theta = 0$ if $\|\mathbf{x}_i - \mathbf{x}_j\| > \phi$.

3 Fuzzy weighted recurrence networks

Let $\mathbf{X} = \{\mathbf{x}\}$ be the set of phase-space states, N a given number of clusters of the states, and a set of N fuzzy clusters, $\mathbf{V} = \{\mathbf{v}_i : i = 1, \dots, N\}$. Fuzzy clusters can be defined as groups that contain data points, where each data point has a degree of fuzzy membership of belonging to each group (the reader is referred to [13] for detailed explanations about the concept and technical formulation of fuzzy clustering). By analogy with the inference for constructing a fuzzy recurrence plot and scalable network [9, 10], a fuzzy relation $\tilde{\mathbf{R}}$ between \mathbf{v}_i and \mathbf{v}_j , $i, j = 1, \dots, N$, is characterized by a fuzzy membership function $\mu \in [0, 1]$, which expresses the degree of similarity of each pair $(\mathbf{v}_i, \mathbf{v}_j)$ in $\tilde{\mathbf{R}}$, and has the following three properties [12]:

1. Reflexivity: $\mu(\mathbf{v}_i, \mathbf{v}_i) = 1, \forall \mathbf{v}_i \in \mathbf{V}$.
2. Symmetry: $\mu(\mathbf{v}_i, \mathbf{x}) = \mu(\mathbf{x}, \mathbf{v}_i), \forall \mathbf{x} \in \mathbf{X}, \forall \mathbf{v}_i \in \mathbf{V}$.
3. Transitivity: $\mu(\mathbf{v}_i, \mathbf{v}_j) = \vee_{\mathbf{x}} [\mu(\mathbf{v}_i, \mathbf{x}) \wedge \mu(\mathbf{v}_j, \mathbf{x})], \forall \mathbf{x} \in \mathbf{X}, \forall \mathbf{v}_i, \mathbf{v}_j \in \mathbf{V}$, where the symbols \vee and \wedge stand for max and min, respectively.

An $N \times N$ fuzzy weighted recurrence network (FWRN) can be constructed with an associated fuzzy weighted adjacency matrix as

$$\mathbf{W} = \tilde{\mathbf{R}} - \mathbf{I}, \quad (3)$$

where \mathbf{W} is an $N \times N$ adjacency matrix of edge weights, and \mathbf{I} is the $N \times N$ identity matrix.

The set of N fuzzy clusters, \mathbf{V} , can be obtained using the fuzzy c -means algorithm (FCM) [13] as follows. Let μ_{ij} denote a fuzzy membership grade of \mathbf{x}_i , $i = 1, \dots, M$, which belongs

to a cluster j , $j = 1, \dots, c$, whose center is \mathbf{v}_j . This fuzzy membership is calculated by the FCM as

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left[\frac{d(\mathbf{x}_i, \mathbf{v}_k)}{d(\mathbf{x}_i, \mathbf{v}_j)} \right]^{2/(m-1)}}, \quad (4)$$

where $1 \leq m < \infty$ is the weighting exponent, and $d(\mathbf{x}_i, \mathbf{v}_j)$ is used as a Euclidean distance between \mathbf{x}_i and \mathbf{v}_j .

Using the fuzzy membership grades, each cluster center \mathbf{v}_j is computed as

$$\mathbf{v}_j = \frac{\sum_{i=1}^M (\mu_{ij})^m \mathbf{x}_i}{\sum_{i=1}^M (\mu_{ij})^m}, \quad \forall j. \quad (5)$$

The iterative procedure of the FCM is outlined as follows.

1. Given c , m , step t , $t = 0, \dots, T$, initialize matrix $\mathbf{U}^{(t=0)} = [\mu_{ij}]^{(t=0)}$
2. Compute $\mathbf{v}_j^{(t)}$, $j = 1, \dots, c$, using Eq. (5).
3. Update $\mathbf{U}^{(t+1)}$ using Eq. (4).
4. If $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \epsilon$ or $t = T$, stop. Otherwise, set $\mathbf{U}^{(t)} = \mathbf{U}^{(t+1)}$ and return to step 2.

The predefined FCM parameters m , T and ϵ usually take the values of 2, 100, and 0.00001, respectively. The number of clusters can be estimated using a cluster validity measure such as the partition entropy, denoted by H , which is defined as [13]

$$H = \frac{1}{M} \sum_{j=1}^c \sum_{i=1}^M \mu_{ij} \log(\mu_{ij}). \quad (6)$$

Given a maximum number of clusters, the partition entropy H for each cluster size, $c \geq 2$, is computed. The number of clusters that has the minimum value of H is considered as an optimal c for the FCM algorithm.

4 Average clustering coefficient and characteristic path length

Two most well-known measures of the statistical characterization of a complex network are the average clustering coefficient and characteristic path length [14, 15, 16]. A clustering coefficient of a node in a network is a numerical indicator of a node that tends to cluster with other neighboring nodes. The average clustering coefficient expresses the average amount of connectivity around individual nodes of a network, whereas the characteristic path length is considered as a measure of the efficiency of transfer of information in a network.

The average clustering coefficient for an unweighted network represented with an $N \times N$ (binary) adjacency matrix $\mathbf{A} = [a_{ij}]$, $i, j = 1, \dots, N$, is defined as

$$C = \frac{1}{N} \sum_{i=1}^N C_i, \quad (7)$$

where C_i is the local unweighted clustering coefficient for node i , and defined as

$$C_i = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}, \quad k_i \neq 0, 1, \quad (8)$$

where k_i is the degree of node i , which is the number of links of node i .

The average clustering coefficient for a weighted network is defined as

$$C^w = \frac{1}{N} \sum_{i=1}^N C_i^w, \quad (9)$$

where C_i^w is the local weighted clustering coefficient for node i , and defined as [17]

$$C_i^w = \frac{\sum_{j,k} [w_{ij}w_{ik}w_{jk}]^{1/3}}{k_i(k_i - 1)}, k_i \neq 0, 1, \quad (10)$$

where $w_{ij}, w_{ik}, w_{jk} \in W$.

The characteristic path length of a network is defined as the average of all shortest path lengths:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j, i,j=1}^N d_{ij}, \quad (11)$$

where d_{ij} is the length of the shortest path between nodes i and j (the Dijkstra's algorithm was used for computing the shortest weighted path in this study).

5 Results

Three time series (x , y , and z components) of the well-known Lorenz attractor [18] were used in this study to construct RN and FWRN. Figure 1 shows the plots of the three Lorenz attractor components, each of which consists of 4000 time points, using standard parameters [18] $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$, which are proportional to the Prandtl number, Rayleigh number, and geometric factor, respectively [19]. Pink-noise ($1/f$) and white-Gaussian-noise time series of the same length were also generated for testing the validity of the RN and FWRN.

The embedding dimension and time delay chosen for both RN and FWRN were 3 and 1, respectively. The recurrence threshold for computing the RN was 10% of the standard deviation of the time series as recommended to be an appropriate value [2]. The partition entropy was used to determine an optimal number of clusters for the FCM algorithm. Based on the cluster validity, 22 clusters were selected as the number of nodes for constructing the FWRN. Figure 2 shows the plot of the numbers of clusters against the partition entropy of the x component of the Lorenz attractor, where the partition entropy sharply drops from 2 to 10 clusters and becomes more stable from the number of clusters = 20.

The clustering coefficients and characteristic path lengths were computed for the RN and FWRN of the three Lorenz attractor components, pink noise, and white Gaussian noise. The results are shown in Table 1. The times taken for computing the clustering coefficient and characteristic path length for each of the three Lorenz components and the two noise time series obtained from the FWRN were less than 3 seconds, running Matlab codes on a personal computer Probook 6570b with Core i7. Using the RN, on the same computer it took 69, 81, 1826, 525, and 289 seconds for computing the clustering coefficients and characteristic path lengths for the x , y , z Lorenz components, pink, and white-noise time series, respectively.

Another test was carried out to test and compare the performance of the proposed FWRN by using a publicly-available PhysioBank dataset that includes electromyograms (EMG) recordings from three subjects: one is without neuromuscular disease (healthy), one with myopathy, and one with neuropathy (<https://physionet.org/physiobank/database/emgdb>). The length of the three EMG time series is 2000, which are shown in Figure 3. The embedding dimension and time delay chosen for both RN and FWRN were 2 and 1, respectively. The

recurrence threshold for computing the RN was also 10% of the standard deviation of the time series. Based on the partition entropy for cluster validity, the number of clusters was selected as 35 that indicates the minimum value of the partition entropy (Figure 4). The clustering coefficients and characteristic path lengths were computed for the RN and FWRN of the three EMG time series. These results are shown in Table 2.

6 Discussion

The clustering coefficients of the RN of the three Lorenz attractor components are higher than those of the two random time series. The characteristic path lengths of the RN of the x and y Lorenz attractor components are higher than those of the two random time series, whereas the z component is lower than the two time series of noise. For the FWRN, the clustering coefficients and the characteristic path lengths of the three Lorenz attractor components are lower than those of the two random time series.

To visualize the relationship of the graph measures of the chaotic and random time series, phylogenetic trees of the two graph measures of the RN and FWRN were constructed using the hierarchical clustering with unweighted pair group method with arithmetic mean (UPGMA) [20], which are shown in Figure 5. For the RN, the clustering coefficients of the two random time series are located between the group of x and y components and the z component (Figure 5(a)); whereas for the FWRN, the tree well separates the three Lorenz attractor components from the two random time series (Figure 5(b)). The tree locates the characteristic path lengths of the z component in the same group with those of the two time series of noise obtained from the RN (Figure 5(c)). For the FWRN, once again the tree show

a clear separation between the three chaotic components and the two time series of noise (Figure 5(d)) .

Figure 6 shows the phylogenetic trees of the two graph measures of the RN and FWRN of the EMG time series recorded from healthy, myopathy, and neuropathy subjects. For both RN and FWRN, the clustering coefficients of the time series of the healthy subject are separated from those of the myopathy and neuropathy subjects (Figure 6(a) and (b)). The characteristic path lengths of the RN for the healthy and neuropathy are grouped together; whereas those of the FWRN well split the healthy subject from the group of myopathy and neuropathy subjects (Figure 6(c) and (d)).

Results shown from the phylogenetic trees of the graph measures demonstrate the higher performance of the FWRN than the RN. Furthermore, because an FWRN is constructed with fuzzy clusters as the number of the network nodes, which are much smaller than the number of the phase-space states, the computational time for computing the graph measures using the FWRN is therefore much less than the RN.

7 Conclusion

The concept and formulation of an FWRN, which appears to be the first kind of univariate weighted recurrence networks, has been presented. The results suggest the effectiveness of FWRN in the analysis of nonlinear time series. The utilization of hierarchical clustering to construct a phylogenetic tree of graph properties of fuzzy weighted recurrence networks can be useful from the standpoint of pattern classification of complex data of various categories. Extension of FWRN to higher dimensional data by modifying the formulation of fuzzy recur-

rence plots would result in wider applications of the proposed approach. Furthermore, the partition entropy is one of several methods for studying cluster validity, each of which may yield a different estimate for an optimal number of clusters that are accordingly expected to provide different values for the graph measures of an FWRN. In this study, the numbers of clusters were estimated based on a single time series. Therefore, investigation into other cluster validity measures and consideration of the number of clusters for each time series would result in more effective implementation of the FWRN.

Note: The Matlab code for constructing the fuzzy weighted recurrence network is available at the author's personal homepage: <https://sites.google.com/site/professortuanpham/>.

References

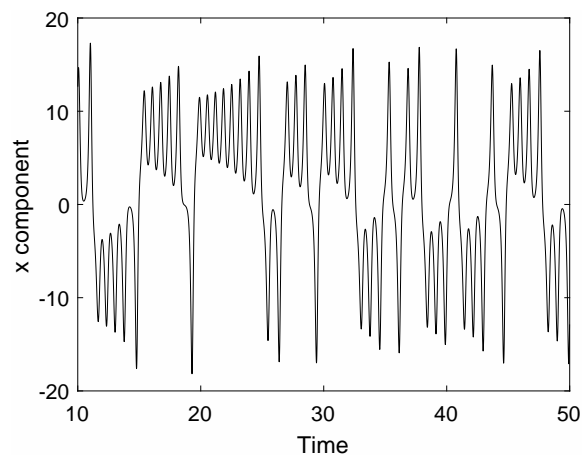
- [1] Eckmann JP, Kamphorst SO, Ruelle D. Recurrence plots of dynamical systems, *Euro-physics Letters* 1987, 5: 973-977.
- [2] MarwanN , Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems, *Physics Reports* 2007, 438: 237-329.
- [3] Marwan N, Donges JF, Zou Y, Donner RV, Kurths J. Complex network approach for recurrence analysis of time series, *Physics Letters A* 2009, 373: 4246-4254.
- [4] Donner RV, Y. Zou Y, Donges JF, Marwan N, Kurths J. Recurrence networks: a novel paradigm for non-linear timeseries analysis, *New Journal of Physics*, 12: 033025.
- [5] Horvath S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*. Springer, New York, 2011.

- [6] Zhu B, Xia Y. Link prediction in weighted networks: a weighted mutual information model, *PLoS ONE* 2016, 11: e0148265.
- [7] Unicomb S, Iniguez G, Karsai M. Threshold driven contagion on weighted networks, *Scientific Reports* 2018, 8: 3094.
- [8] Gao ZK, Yang YX, Cai Q, Zhang SS, Jin ND. Multivariate weighted recurrence network inference for uncovering oil-water transitional flow behavior in a vertical pipe, *Chaos* 2016, 26: 063117.
- [9] Pham TD, Fuzzy recurrence plots, *EPL* 2016, 116: 50008.
- [10] Pham TD, From fuzzy recurrence plots to scalable recurrence networks of time series, *EPL* 2017, 118: 20003.
- [11] Marwan N, Kurths J. Complex network based techniques to identify extreme events and (sudden) transitions in spatio-temporal systems, *Chaos* 2015, 25: 097609.
- [12] Zadeh LA. Similarity relations and fuzzy orderings, *Information Sciences* 1971, 3: 177-200.
- [13] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [14] Watts DJ, Strogatz S. Collective dynamics of “small-world” networks, *Nature* 1998, 393: 440-442.

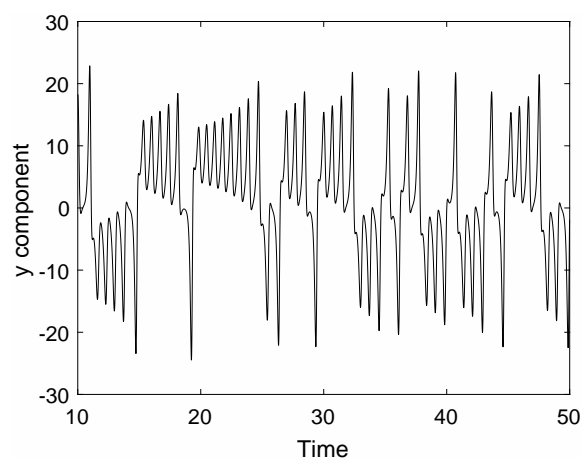
- [15] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks, *Proceedings of the National Academy of Sciences* 2004, 101: 3747-3752.
- [16] Albert R, Barabasi AL. Statistical mechanics of complex networks, *Rev. Mod. Phys.* 2002, 74: 47-97.
- [17] Fagiolo G. Clustering in complex directed networks, *Phys Rev. E.* 2007, 76: 026107.
- [18] Lorenz EN. Deterministic nonperiodic flow, *J. Atmos. Sci.* 1963, 20: 130-141.
- [19] Sparrow C. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.* Springer, New York, 1982.
- [20] Weib M, Goker M. Chapter 12 – Molecular Phylogenetic Reconstruction, In *The Yeasts* (Fifth Edition), edited by Kurtzman CP, Fell JW, Boekhout T. Elsevier, London, 2011, pp. 159-174.

Table 1: Average clustering coefficients (C) and characteristic path lengths (L) obtained from FWRNs and RNs for x , y , and z components of Lorenz system, pink noise (PN), and white Gaussian noise (WN).

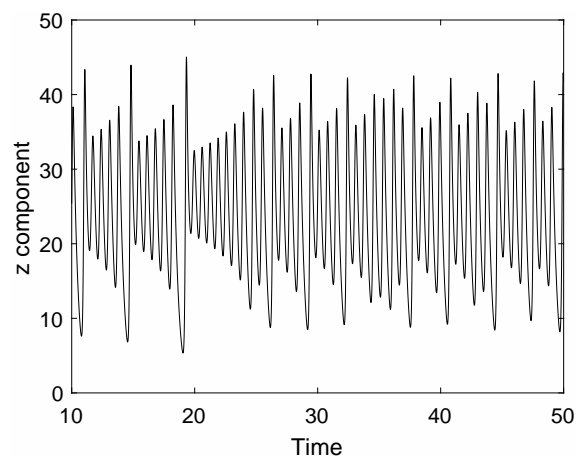
Time series	C	L
RN		
x component	0.6845	13.0608
y component	0.6890	14.0846
z component	0.9313	1.1147
PN	0.6618	2.1332
WN	0.5538	2.3935
FWRN		
x component	0.0152	0.0071
y component	0.0226	0.0133
z component	0.0255	0.0143
PN	0.0440	0.0524
WN	0.0483	0.0761



(a)



(b)



(c)

Figure 1: Three components of Lorenz attractor.

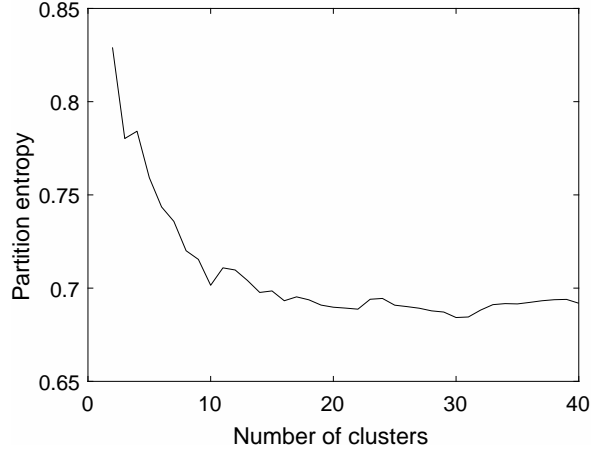
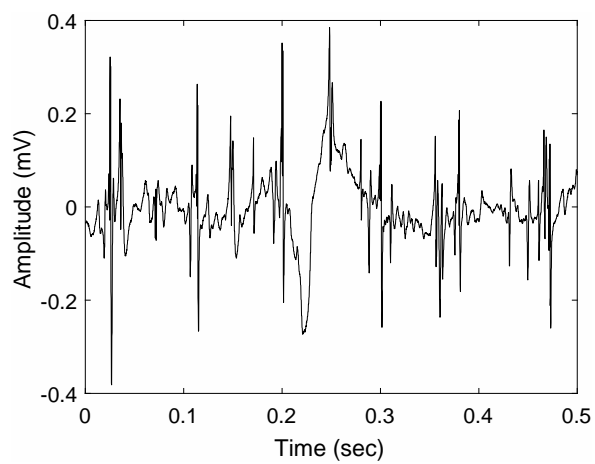


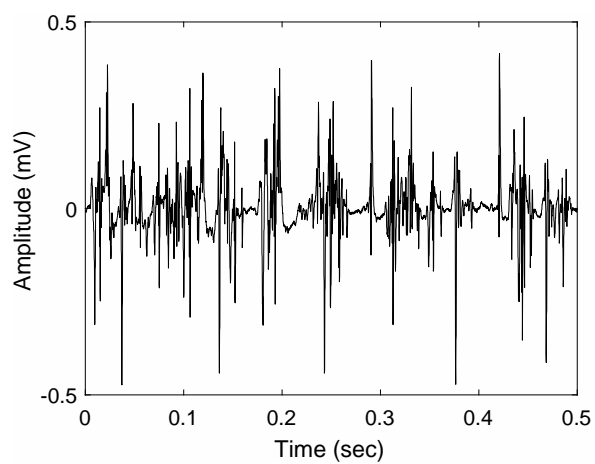
Figure 2: Cluster validity of x component of Lorenz attractor.

Table 2: Average clustering coefficients (C) and characteristic path lengths (L) obtained FWRNs and RNs for EMG recordings of healthy, myopathy, and neuropathy subjects.

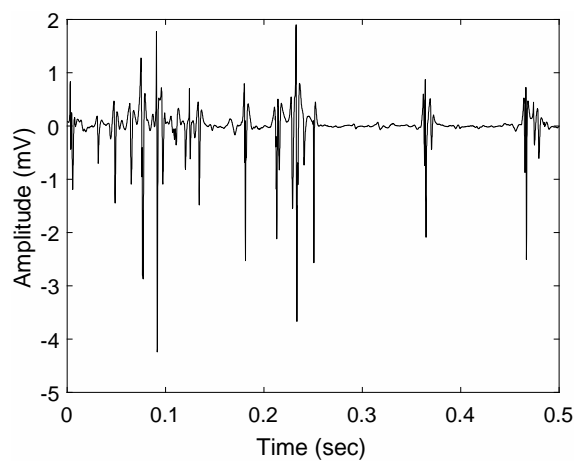
Subject	C	L
RN		
Healthy	0.8680	1.6977
Myopathy	1	1
Neuropathy	0.9442	1.3595
FWRN		
Healthy	0.0258	0.0334
Myopathy	0.0271	0.0370
Neuropathy	0.0272	0.0366



(a)



(b)



(c)

Figure 3: EMG recordings from three subjects: (a) healthy, (b) myopathy, and (c) neuropathy.

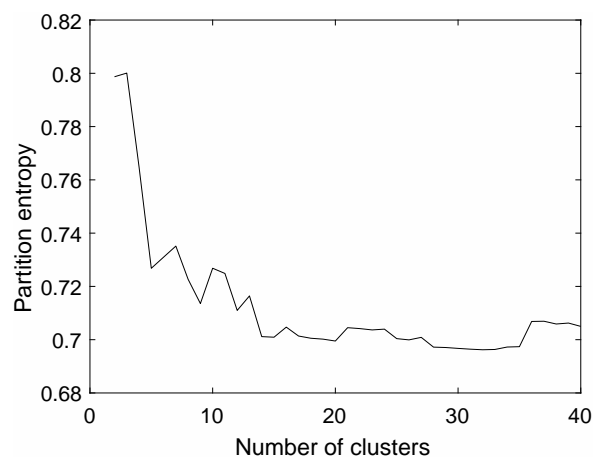
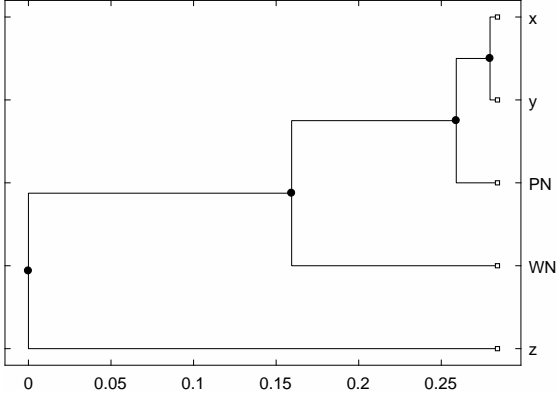
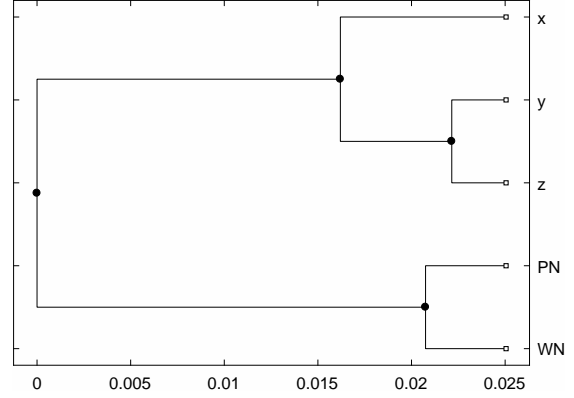


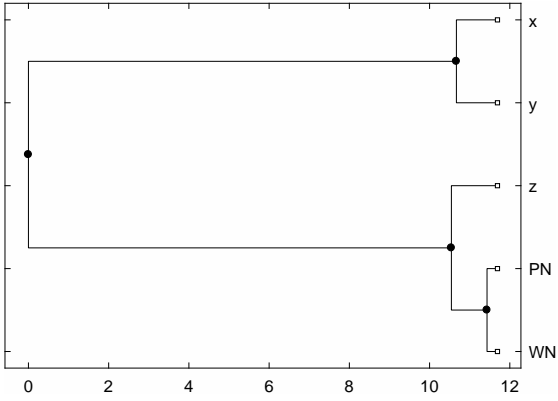
Figure 4: Cluster validity of EMG time series of healthy subject.



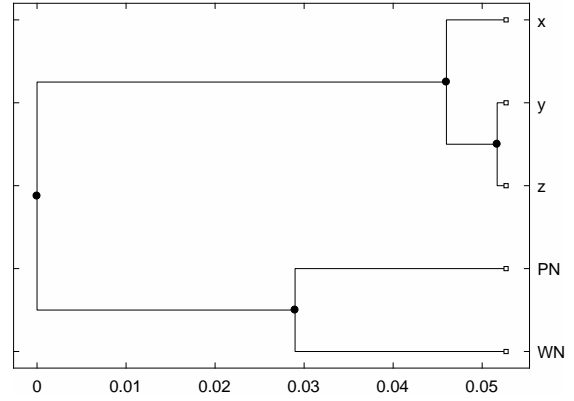
(a) RN



(b) FWRN



(c) RN



(d) FWRN

Figure 5: Hierarchical clustering of average clustering coefficients (a) and (b), and characteristic path lengths (c) and (d) for x , y , z Lorenz components, pink noise (PN), and white Gaussian noise (WN) obtained from recurrence networks (RN) and fuzzy weighted recurrence networks (FWRN).

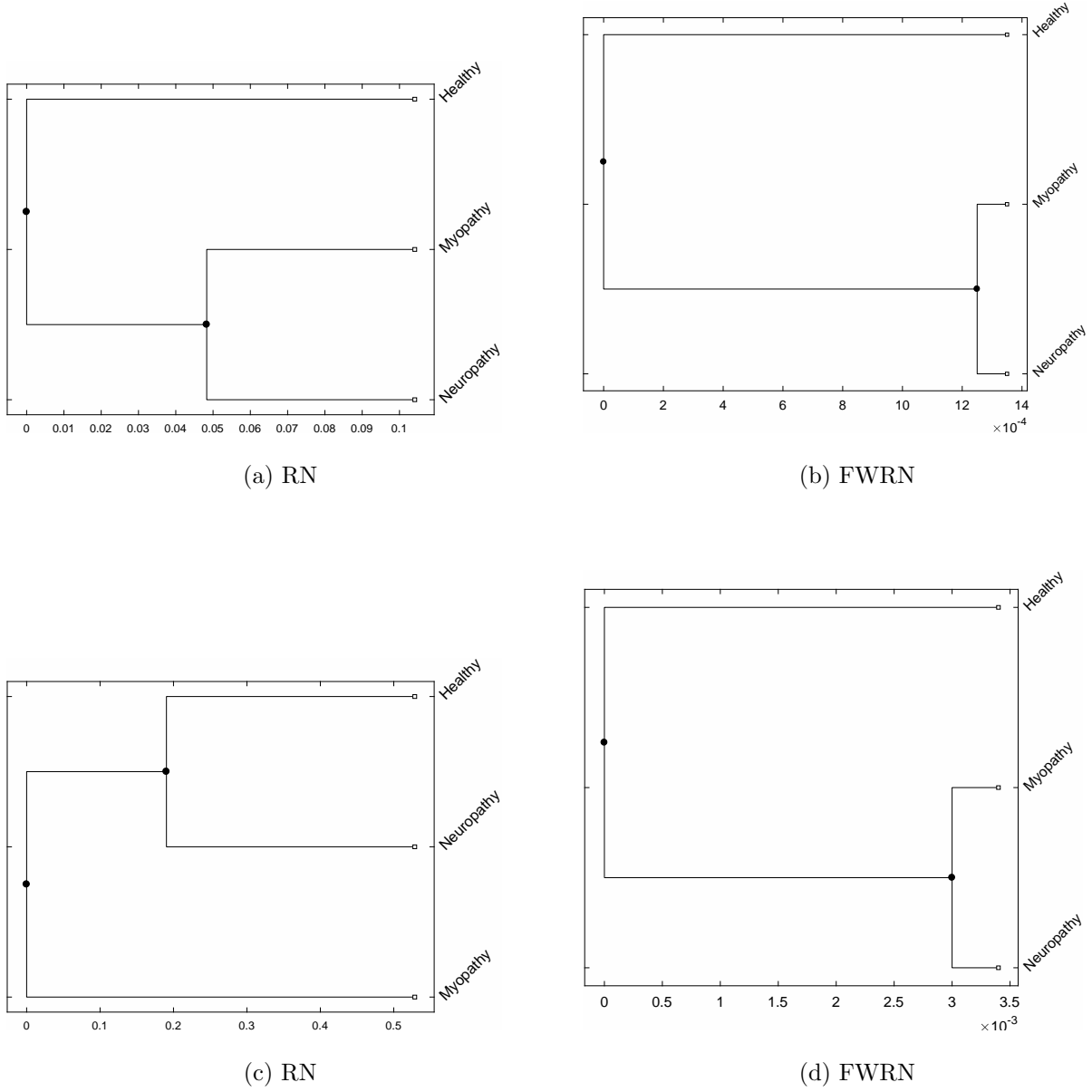


Figure 6: Hierarchical clustering of average clustering coefficients (a) and (b), and characteristic path lengths (c) and (d) for EMG recordings of healthy, myopathy, and neuropathy subjects obtained from recurrence networks (RN) and fuzzy weighted recurrence networks (FWRN).