Parametric Construction of Episode Networks from Pseudoperiodic Time Series Based on Mutual Information

Frank Emmert-Streib*

Computational Biology and Machine Learning Lab, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, United Kingdom

Abstract

Recently, the construction of networks from time series data has gained widespread interest. In this paper, we develop this area further by introducing a network construction procedure for pseudoperiodic time series. We call such networks *episode networks*, in which an episode corresponds to a temporal interval of a time series, and which defines a node in the network. Our model includes a number of features which distinguish it from current methods. First, the proposed construction procedure is a parametric model which allows it to adapt to the characteristics of the data; the length of an episode being the parameter. As a direct consequence, networks of minimal size containing the maximal information about the time series can be obtained. In this paper, we provide an algorithm to determine the optimal value of this parameter. Second, we employ estimates of mutual information values to define the connectivity structure among the nodes in the network to exploit efficiently the nonlinearities in the time series. Finally, we apply our method to data from electroencephalogram (EEG) experiments and demonstrate that the constructed episode networks capture discriminative information from the underlying time series that may be useful for diagnostic purposes.

Introduction

The definition and the study of discrete objects in the form of graphs or networks with specific properties is a topic that reaches back over two hundred years [1–3], eventually leading to the founding of graph theory [4–6]. Despite the mathematical origin of this field, its contemporary form, sometimes called network analysis [7–10], attracts interdisciplinary interest as networks can be found pervasively in nature. For this reason many new methods have been developed in recent years to provide quantitative approaches for the structural analysis of complex networks [11].

One specific aspect of network analysis that is currently of great interest is the inference, reconstruction and construction of networks from data. For example, in molecular biology and the biomedical sciences powerful experimental assays allow the measurement of the activity of genes or gene products on a genome-scale. Several methods have been introduced to infer various forms of gene networks [12] from such high-throughput data [13–20]. Similarly, in neuroscience one tries to infer neural networks that capture the interactions among neurons or neuronal regions [21–23]. The ultimate goal of these methods is to infer causal networks [24,25]. That means in the above networks an interaction in the inferred networks corresponds to a predicted physical interaction among system variables that can be verified experimentally [26]. For example, in gene networks this could correspond to the binding of two proteins or for neural networks this could be the synaptic connection of two neurons. In other

words, it is assumed that there exists a network that underlies the data which shall be estimated or reconstructed from the data. Recently, a fundamentally different way to construct networks, using time series data, has been introduced [27–34]. The principle difference to the methods discussed above is that the networks constructed this way are merely a *representation* of the data. That means it is not assumed that there exists a network that is behind the data which should be reconstructed, but the network is *constructed* from the data to form a formal representation thereof, which serves as a means for further analysis. In the remainder of this paper we are concerned with the latter type of networks.

Specifically, there are two network construction methods for time series data that have gained considerable popularity since their introduction. The first network construction method generates so-called *cycle networks* [34]. For this method a node in the constructed network corresponds to a cycle in the time series and two nodes are connected if the corresponding cycles are similar to each other as measured by a correlation coefficient. That means the correlation coefficients between pairs of cycles give the components of a similarity matrix which is used to obtain the connectivity of the cycle network. The resulting network is undirected, because the correlation coefficient does not provide information about a directionality. It is noted in [34] that the similarity matrix can be either filtered by applying a global threshold parameter transforming it into a binary network or the similarity matrix can be used unfiltered in which case it is a fully connected network. The key of the above method is that each

cycle corresponds to a well defined part of the time series, which can be seen as a profile vector of a certain length. The second of these methods constructs so-called *visibility graphs* from a time series [30]. In a visibility graph, nodes correspond to time points of a time series and two nodes are connected if there is a certain criterion met that involves the values of the time series. A visualization of this criterion shows that two nodes are connected if one time point is *visible* from the second one, hence the name of these networks. This leads to an undirected and unweighted network.

The major purpose of this paper is to introduce a construction procedure for networks from pseudoperiodic time series. Here, by pseudoperiodic we mean a time series that exhibits oscillatory or even chaotic behavior. Our construction method adds on previous methods, and includes several key features that makes it distinct. The first feature of our method is to *estimate* the connectivity structure of the constructed network from the underlying time series. This is different to the construction of visibility graphs [30] that establish the connectivity among nodes by testing a *geometric* criterion, instead of a statistical one. However, this is similar to cycle networks [34] that estimate the correlation coefficients between cycles. Also financial networks have been constructed based on the estimation of correlation coefficients [29]. Considering the fact that we assume a time series to behave oscillatory or even chaotic one can expect that this time series is strongly nonlinear. For this reason we estimate mutual information values, instead of correlation coefficients, because the mutual information is capable of capturing nonlinear effects in a time series [35,36] and, hence, provides more accurate estimates of the similarity of nonlinear time series intervals. Also, this builds directly on results obtained from other fields in which estimates of mutual information values have been used to infer causal gene networks [14,37,38]. Second, we define a node in the constructed network as an *episode*. An episode is a temporal interval of the time series that consists of $n_e \geq 1$ consecutive cycles. That means an episode is n_e times longer than a cycle. The extended length of an episode, compared to a cycle, has the advantage of increasing the accuracy of the statistical estimates of the mutual information value. The reason for this is that a cycle does not need to have a certain minimal length to qualify as a cycle. However, it is clear that very short cycles convey less information about the time series than long cycles. Due to the fact that the notion of a 'cycle' is parameter free, one cannot adjust for this shortcoming. For this reason we extend the principle idea behind the usage of a cycle in the construction of a network [34] by means of an episode. Third, our network construction model is a parametric method because an episode is a function of n_e, the number of consecutive cycles. This gives us a parameter that can be optimized to result in the 'best' network for a given time series. We call the optimal value of n_e the *effective length* of an episode and provide a procedure to estimate its value. None of the previous methods introduced to construct networks from time series data is parametric. Fourth, the size of the constructed network, which corresponds to the number of nodes, is adjustable in our model. Again, this is related to the length of an episode. For the *effective length* of an episode, this results in networks of minimal size, which means that it consists of the least number of nodes.

This paper is organized as follows. In the next section we introduce episode networks and their construction. Then we present results studying the influence and the distribution of the mutual information values of episode networks. Next, we compare properties of episode and cycle networks. We then show several examples to demonstrate how the effective length of an episode can be estimated. Finally, we apply our method to EEG data. By this analysis we demonstrate that the constructed episode networks capture discriminative information from the underlying time series that might be useful for diagnostic purposes. We finish this article with a summary and conclusions.

Methods

In this section we introduce and discuss the construction procedure for episode networks. Further, we introduce an algorithm to estimate the *efficient length* of an episode to construct such networks.

Construction of episode networks

Episode networks, defined below, are based on mutual information values [39,40]. The mutual information is a measure for the nonlinear dependency of two random variables X and Y, defined by

$$I(X,Y) = \sum_{x_i \in X} \sum_{y_i \in Y} P(X=x_i, Y=y_i) \cdot \log \frac{P(X=x_i, Y=y_j)}{P(X=x_i) \cdot P(Y=y_i)} \quad (1)$$

Here by $\log$ we mean the logarithm to the base 2. $I(X,Y)$ is always ≥ 0. If the two random variables are independent from each other the mutual information becomes zero, because $P(x,y) = P(y)P(x)$.

Before we define the construction procedure for episode networks formally, we provide a brief depiction of it. The basic idea that underlies the networks we want to construct is as follows. For a given pseudoperiodic time series that consists of $\{C_i\}_{i=1}^{N_c}$ cycles we define an episode as n_e consecutive cycles. This results in $N_e = N_c / n_e$ different episodes $\{E_j\}_{j=1}^{N_e}$. That means an episode is an interval of a time series that contains n_e consecutive cycles. This is visualized in Fig. 1. We use these episodes as the nodes in a network. The connection of this network is based on the similarity between these episodes. Here we measure the similarity between pairs of episodes by their mutual information value [39,41], $I(E_i, E_j)$. That means we estimate a similarity matrix W, whose components correspond to the mutual information values between pairs of episodes, i.e., $W_{ij} = I(E_i, E_j)$. From the similarity matrix W we construct an episode network as the *maximal connected component*, which is an undirected, unweighted network. Here we define the *maximal connected component* as the network G obtained from W which is (1) a connected network and (2) the edges used to construct G have maximum mutual information values. The first property means that in an episode network each node is connected via an undirected path to any other node. The second property defines its construction procedure from W which is as follows: First, we initialize the adjacency matrix A of G as a zero matrix. Then we identify the largest edge weight (mutual information value) in W and its corresponding episode pair (i,j) and add an undirected, unweighted edge in A, i.e., $A_{ij} = A_{ji} = 1$. If this results in a connected component of the vertices in A we stop, otherwise we proceed to the next largest edge weight in W and continue until we obtain a connected component of the vertices in A. Formally, our construction corresponds to a *greedy optimization* [42] of the mutual information values used to construct G.

The construction of the maximal connected component is visualized in Fig. 2. Assuming the black edges have already been added and all other edge weights are zeros, except w_a and w_b, for which holds $w_a > w_b$. In the next step we face a decision which of the two edges w_a and w_b to include. According to our construction procedure, the edge w_a will be added because its mutual information value is larger than w_b. This is in contrast to the *minimum spanning tree* (MST) algorithm [42] which would add the
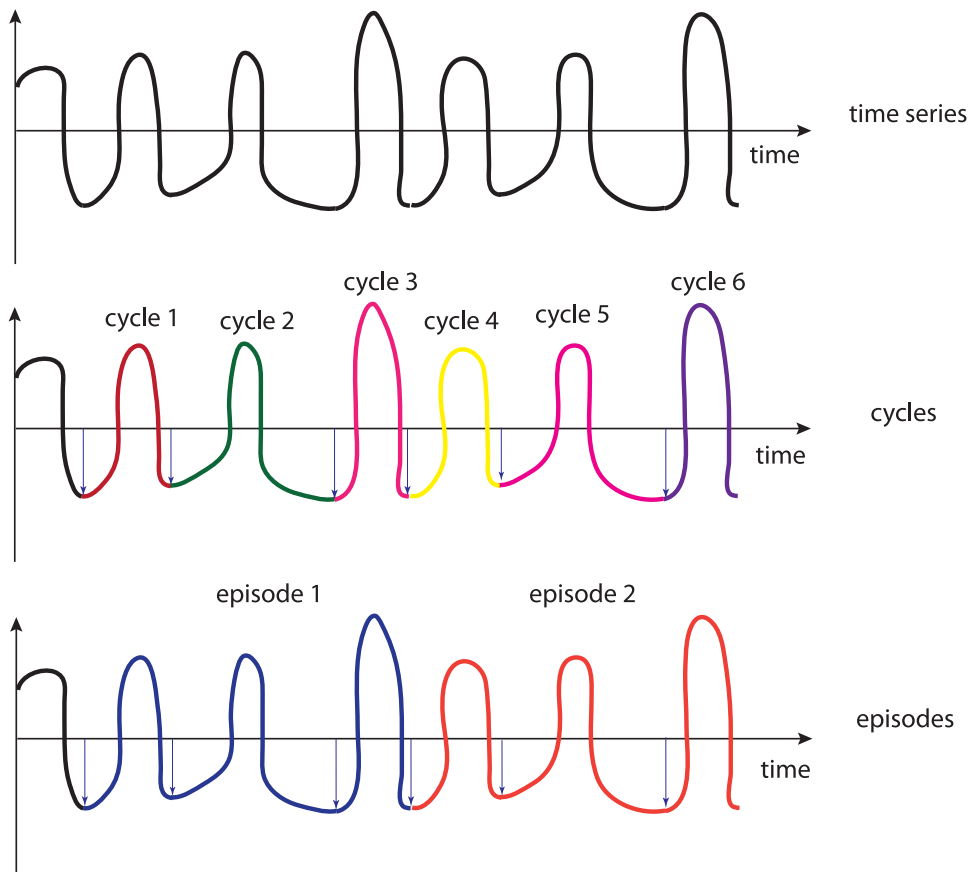
Figure 1. Top: Hypothetical time series. Middle: Identification of all cycles in the time series. Bottom: Definition of episodes. In this case, an episode consists of three consecutive cycles.
doi:10.1371/journal.pone.0027733.g001

edge w_b instead of w_a. That implies that the resulting network we obtain from our procedure does not need to be a tree, but the network can have an arbitrary complex structure.

The construction procedure of an episode network from a time series can be summarized by the following four steps.

Algorithm 1 construction of an episode network

    1: Given: pseudoperiodic time series
    2: Initialize: adjacency matrix A as zero matrix
    3:. Initialize: episode length n_e as a natural number
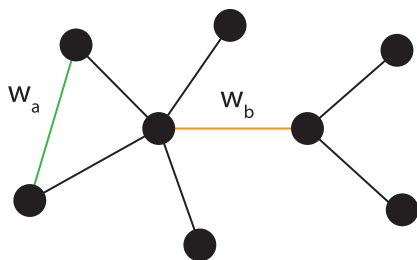   4: identify all cycles in the time series, $\{C_i\}_{i=1}^{N_c}$

Figure 2. Construction principle based on the maximal connected component. If $w_a > w_b$, edge w_a will be added, otherwise w_b. This is in contrast to the MST algorithm which would always add w_b.
doi:10.1371/journal.pone.0027733.g002

5: construct $N_e = N_c/n_e$ episodes from the cycles, $\{E_j\}_{j=1}^{N_e}$
6: **for** $i \in \{1, \ldots N_e\}$ **do**
7: **for** $j \in \{(i+1), \ldots, N_e\}$ **do**
8: $W_{ij} = I(E_i, E_j)$ - estimate mutual information
9: **end for**
10: **end for**
11: **while** A is not connected **do**
12: $(i^*, j^*) = \text{argmax} \{W_{ij}\}$
13: $W_{i^*,j^*} = W_{j^*,i^*} = 0$
14: $A_{i^*,j^*} = A_{j^*,i^*} = 1$
15: **end while**
16: Return: adjacency matrix A of the episode network

1. Identify the cycles $\{C_i\}_{i=1}^{N_c}$ in the time series.
2. Define the episodes $\{E_j\}_{j=1}^{N_e}$ in the time series. An episode corresponds to a node in the constructed network G.
3. Estimate the similarity between pairs of episodes $W_{ij} = I(E_i, E_j)$ by their mutual information value.
4. Connect nodes (episodes) in the network G if they are part of the maximal connected component in W.

Formally, the construction procedure for episode networks is defined in algorithm 1. We want to point out that, usually, the length of different cycles C_i will be different, hence, the length of

different episodes E_i will be different too. In order to estimate the mutual information value between episode pairs of different length, we employ a similar strategy as in [34,43] used for correlation coefficients. That means if episode E_i is longer than E_j we estimate all possible mutual information values one can obtain by shifting the start position s of E_j with respect to E_i and select from these values the maximum mutual information value, i.e.,

$$I(E_i, E_j) = \max_s \{I(E_i, E_j(s))\}. \tag{2}$$

This allows to circumvent the problem of unequal episode lengths.

Our network model to construct episode networks is similar to the construction of cycle networks [34] but has the following benefitial features. First, due to the fact that the model is intended to convert a pseudoperiodic time series into a network, one can assume that the signal in the time series is strongly nonlinear. For this reason it appears sensible to estimate the similarity between episodes with a measure that is capable of capturing such nonlinearities. The mutual information is a nonlinear extension of the correlation coefficient between two random variables and, hence, possesses this property [35,36]. Further, it has been demonstrated for large-scale high-throughput data from gene expression experiments that mutual information based inference methods are able to reconstruct molecular interactions among genes or gene products reliably [14,37,38,44,45]. This demonstrates that theoretical properties of the mutual information translate to data from experiments making this measure a favorable choice over correlation coefficients.

Second, a cycle is a well defined entity within a pseudoperiodic time series [46] and as such is appropriate to represent a node of a network. However, this bears an implicit limitation with respect to the length of the profile vectors that are compared. The problem is that from simulation studies we found that the estimated correlation coefficients between two cycles is in general very high, and frequently even close to 1. One reason for this is that for pseudoperiodic time series the length distribution of the profile vectors is of similar order. Another reason is the 'periodic shape' of the cycles, which they naturally represent. Both effects do not prevent, but hamper that the full range of possible correlation coefficients from -1 to $+1$ is actually observed. The usage of an episode as basic building block of a node, and of the profile vectors, leads naturally to larger profile vectors and enables by this a larger diversity of observed similarity values among different episodes.

Third, an episode network grows proportional to the number of episodes $\#\{E_i\}_{i=1}^{N_e}$ in the time series. That means, the number of nodes in an episode network grows with the number of episodes, $N_e = \#\{E_i\}_{i=1}^{N_e}$, and not with the number of cycles $N_c = \#\{C_i\}_{i=1}^{N_c}$ as the cycle networks. Due to the fact that an episode consists of n_e cycles we obtain $N_e = N_c/n_e$. That means the size of an episode network is directly controllable by the number of cycles n_e that define an episode. Further practical implications of this relate back to point 2 discussed above as well as to a more efficient computational complexity of an analysis of smaller compared to larger networks.

The last point raises the question how to choose the length of an episode to construct episode networks, which are best suited for a given time series. This point is addressed in the next section.

Effective length of an episode

First, we would like to note that in order to define a procedure to determine the effective length of an episode we need to specify a measure c. This measure will serve as a reference that allows us to

quantify what we mean by *effective*. Because the length of an episode has an influence on the structural properties of the resulting episode network G, we are looking for a measure to quantify the structural properties of a network, i.e., $c(G)$. Potentially, there are several choices for such a measure. For example, the mean path length or the edge density of a network are possible measures. Due to the fact that we will study the value of this measure for different values of the episode length, $c_{n_e}(G)$, we require it to be largely independent of the size of the network, because n_e effects directly the size of the constructed episode network. For our analysis we use the global clustering coefficient as measure. The global clustering coefficient, also called global transitivity, is a well-known measure that captures an important property of complex networks [47]. Briefly, it measures the probability that adjacent nodes of a vertex are connected with each other. For our analysis we use the global clustering coefficient of a network, $c(G)$, which is defined as the average clustering coefficient of all individual nodes in the network. In the following definition we specify the notion of the *effective length* of an episode. In this definition, the function h is directly related to the measure c as will be explained in more detail below.

Definition 1 (*effective length*) *As effective length of an episode we define the maximal length of an episode, denoted as $n_e^{el}(h)$, for which the structural properties of a population of episode networks is maximal, as measured by h. Quantitatively, we define $n_e^{el}(h)$ as*

$$n_e^{el}(h) = \max\{\underset{n_e}{\mathrm{argmax}}\{h_{n_e}\}\} \tag{3}$$

with respect to the function h.

In order to point out that the effective length $n_e^{el}(h)$ of an episode network is defined with respect to a network measure, h, we included this dependency in the above definition explicitly for reasons of clarity. The efficient length is the maximum of $\mathrm{argmax}_{n_e}\{h_{n_e}\}$ because in case there are several elements that maximize h_{n_e}, we want to chose the largest episode length because this results in the smallest networks.

Algorithm 2 procedure to estimate the effective length n_e^{el} of an episode

1: Initialize:
2: $n_e^{max} > 1$
3: $m_{n_e}(G) = \sum_{j=1}^{E} c_j(G)$ for $n_e \in \{1, \ldots, n_e^{max}\}$
4: $\sigma_{n_e}^2 = Var(m_{n_e}(G))$ for $n_e \in \{1, \ldots, n_e^{max}\}$
5: $m_{max} = m_1(G)$ - mean global clustering coefficient for $n_e = 1$
6: $\theta = m_{max} - \sigma_1(G)$
7: $h_{n_e} = 0$ for $n_e \in \{1, \ldots, n_e^{max}\}$
8: $h_1 = m_{max}$
9: **for** $n_e \in \{2, \ldots, n_e^{max}\}$ **do**
10:     **if** $m_{n_e}(G) - \sigma_{n_e}(G) \times f \geq \theta$ **then**
11:        $h_{n_e} = m_{max}$
12:     **else**
13:        $h_{n_e} = m_{n_e}(G)$
14:        $m_{max} = h_{n_e}$
15:        $\theta = m_{max} - \sigma_{n_e}(G)$
16:     **end if**
17: **end for**
18: Return:
19: $n_e^{el}(h) = \max\{\mathrm{argmax}_{n_e}\{h_{n_e}\}\}$

The rational behind the above definition of the optimization criterion is to select the length of an episode in a way that provides us with the *maximal* information from an episode network of *minimal* size. The information is maximal because for $n_e = 1$ an episode network consists of the maximal number of nodes, N_c, and, hence, can exhibit its largest structural diversity. This can be seen as the highest resolution achievable. Increasing the value of n_e, leads to a reduction of the number of nodes in an episode network and, potentially, to a restriction in the diversity of the network structure. However, by searching the maximum of h and n_e, we obtain smaller episode networks that represent approximately the same structural information as larger networks. Hence, using the effective length n_e^{el} leads to the smallest size of an episode network with similar structural properties as larger episode networks, constructed from the same time series.

The discrete function h in the above definition is obtained from estimates of structural properties of G in dependence on n_e. Formally, we define h in algorithm 2. In this procedure, $m_{n_e}(G)$ corresponds to estimates of the mean global clustering coefficients of a population of episode networks and σ_{n_e} to the standard deviations. Practically, we approximate this population of networks by an ensemble of networks of size E. The basic idea underlying the definition of the function h is to utilize information about the variability of the structural properties of networks from the same population. It is necessary to formulate this with respect to a population, because every structural property of an episode network, e.g., the clustering coefficient $c(G)$ is a random variable, due to the fact that the time series used to construct G is just a sample from a dynamical system. In order to assess such a random variable one needs to consider its inherent variability. As long as $m_{n_e}(G) - \sigma_{n_e}(G) \times f \geq \theta$ holds, the values of the constructed episode networks are within one standard deviation of the threshold θ. Here the factor f allows to adjust this range, but for our simulations we used $f = 1.0$. According to the central limit theorem, for a sufficiently large episode network the values of the global clustering coefficients $c(G)$ are approximately normal distributed ($P_{n_e} = N(m_{n_e}(G), \sigma_{n_e}^2(G)/\sqrt{n_e})$) with mean $m_{n_e}(G)$ and variance $\sigma_{n_e}^2(G)/\sqrt{n_e}$ [48], because the global clustering coefficient is the average of the clustering coefficients of the individual nodes in the network. Considered from this perspective, the criterion $m_{n_e}(G) - \sigma_{n_e}(G) \times f \geq \theta$ means that the probability to observe a value of $c(G)$ larger or equal to θ is $1 - F_{P_{n_e}}(\theta)$. Here $F_{P_{n_e}}$ is the cumulative distribution function of P_{n_e}. Hence, our procedure identifies the maximal length of an episode n_e^{el} for which the structural properties of episode networks are still within the variability range of the population of episode networks.

Results

In the following we, first, study the influence and the distribution of the mutual information values of episode networks and compare properties of episode and cycle networks [34]. Then we show how to determine the effective length of an episode to construct the networks. Finally, we apply our method to EEG data to demonstrate that the constructed episode networks capture discriminative information from the underlying time series.

Influence and distribution of the mutual information

For the following analysis we use time series data generated with a Rössler system [49] given by

$$\dot{x}(t) = -y(t) - z(t) \tag{4}$$

$$\dot{y}(t) = x(t) + a \cdot y(t) \tag{5}$$

$$\dot{z}(t) = b + z(t) \cdot (x(t) - c) \tag{6}$$

For parameter values of $a = 0.398, b = 2.0, c = 4.0$ the Rössler system exhibits a chaotic behavior. From this system, we use the x-component to generate a time series with 800 cycles. For $n_e = 5$ we obtain 160 episodes. From this we estimate an episode and a cycle network G_E and G_C. First, we want to note that the episode network contains 3090 edges and the cycle network contains 2582 edges. Already from these numbers one gets the impression that the usage of the mutual information as estimator of episode similarity has a profound impact on the inferred network structure. In order to demonstrate this more clearly, we show in Fig. 3 the histograms of the mutual information values (left figure) and the correlation coefficients (right figure) of all none-zero edge weights of the mutual information matrix and the correlation matrix. From this follow two interesting observations. First, the distribution of mutual information (MI) values appears vertically mirrored compared to the distribution of the correlation coefficients. That means, if one goes from high to small values of the MI values, one enters first the long tail of the distribution and then reaches the center of mass of the majority of values. For the correlation coefficients this behavior is reversed. Due to the fact that our network construction procedure adds successively edges starting with high edge weights (mutual information values) and working toward lower values, the distributional shape of the mutual information values is beneficial because it allows for a more selective procedure.

The second related observation refers to the covered range of selected values, colored in red in Fig. 3. In this figure, the mutual information (MI) values and the correlation coefficients that were actually used to construct G_E and G_C are colored in red. For the episode network the range covered by the selected values is $0.875 = $ (maximal value of the selected MI values - minimal value of the selected MI values)/(maximal value of all MI values - minimal value of all MI values) whereas for the cycle network the covered range is only 0.094. This is intimately connected with the distributional shape of both networks, as discussed above, and the location of its tail. To quantify the distributional shape of the tail of the mutual information values we conduct a statistical test suggested in [55] to test if the tail follows a power law $\sim I^{-\alpha}$. A maximum-likelihood fit results in an exponent of $\alpha = 3.02$ and a goodness-of-fit test (Kolmogorov-Smirnov) gives a p-value of 1.3×10^{-4}, indicating that the tail of the mutual information values is unlikely to follow a strict power law. That means the distribution has a long tail but does not exactly decay as $\sim I^{-\alpha}$.

The next dynamical system we study is a Duffing map [50],

$$x_{n+1} = x_n \tag{7}$$

$$y_{n+1} = -bx_n + ay_n - y_n^3 \tag{8}$$

For the parameters $a = 2.77$ and $b = 0.1$ the obtained time series is chaotic. Using this parameter configuration we generate a time series with 1811 episodes ($n_e = 5$). A maximum-likelihood fit of the tail of the mutual information values gives an exponent of $\alpha = 7.06$ and the goodness-of-fit test gives a p-value of 0.0022 [55]. This test shows that the tail of mutual information values for the Duffing
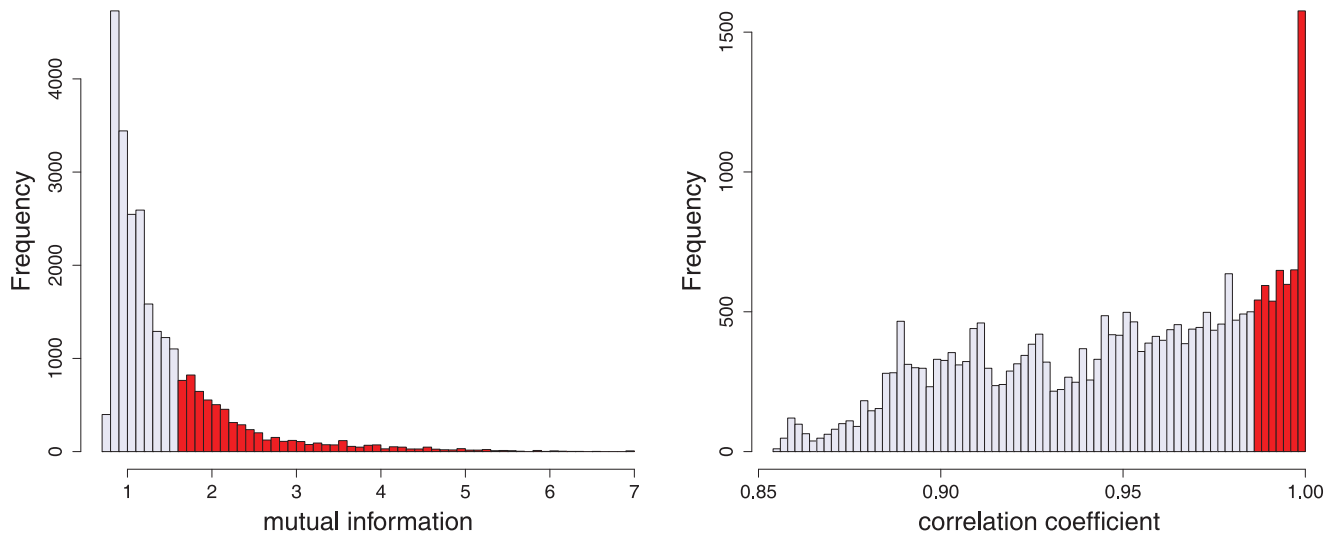
Figure 3. Histogram of mutual information values (left) and correlation coefficients (right) obtained for a time series with 160 episodes and 800 cycles generated with a Rössler system. Only the values colored in red are used to construct the corresponding episode and cycle network.
doi:10.1371/journal.pone.0027733.g003

map is closer to a power law than the Rössler system, but is also unlikely to be an exact power law, $\sim I^{-\alpha}$.

To demonstrate that the resulting episode networks for a Rössler system and a Duffing map have a different network connectivity, we show in Fig. 4 the degree distribution of the episode networks (left: Rössler system 1607 episodes; right: Duffing map with 1811 episodes). Despite the fact that we used parameters of the Rössler system and the Duffing map leading to a chaotic behavior of both time series, the resulting degree distributions of the episode networks are quite different from each other. This means that not every chaotic system maps to a network with the same connectivity structure. Despite the fact that the mutual information distribution of an episode network of a Rössler system has a long tail, its degree distribution has not. This is similar to the results obtain for cycle networks [34].

In Fig. 5 we show two visualizations of episode networks. The top network is obtained from a Rössler system consisting of 321 nodes and 4208 edges. Its average path length is 9.06, the clustering coefficient is 0.822 and the maximal degree is 40. The second network is obtained from a Lorenz system,

$$\dot{x}(t) = \sigma \cdot (y(t) - x(t)) \tag{9}$$

$$\dot{y}(t) = x \cdot (\tau - z(t)) - y(t) \tag{10}$$

$$\dot{z}(t) = x(t) \cdot y(t) - \beta \cdot z(t) \tag{11}$$

with the parameters $\sigma = 10$, $\tau = 28$ and $\beta = 8/3$, which lead to a chaotic time series. The episode network was constructed using the x-component of the Lorenz equations and consists of 1663 nodes and 31430 edges. Its average path length is 6.89, the clustering coefficient is 0.756 and the maximal degree is 143. Both networks are obtained for $n_e = 5$. It is interesting to see that despite the fact that both time series are chaotic the resulting episode networks 'look' quite different, which reflects also in their structural properties. We want to emphasize that the shown episode

networks are obtained by the application of algorithm 1. That means, there is no manual adjustment of any parameter necessary.

In Fig. 6 we show the degree distribution for the episode network constructed from the Lorenz equations, shown in Fig. 5. Comparing this degree distribution with the ones for the Rössler system and the Duffing map, shown in Fig. 4, one can see that also the episode network constructed from the Lorenz equations has a characteristic connectivity structure that is different to the other two dynamical systems. This confirms also the visual impression from the graphical representation of the episode networks shown in Fig. 5.

Estimation of the effective length of an episode

In the previous section we used an episode length of $n_e = 5$ to construct the episode networks. That means we defined an episode as 5 consecutive cycles as the nodes in our networks. In this section we use the quantitative procedure introduced in the methods section that allows to determine the *effective length* of an episode automatically. Our results will show for different systems that in general $n_e = 5$ is a good choice for the effective length of an episode.

In order to determine the effective length of an episode we start with a time series of a fixed length L and construct for various values of n_e the corresponding episode networks. According to the definition of *effective length*, we need to identify the value n_e from which on the characteristics of the networks change. Per definition, this is the point of the first decrease of the maximal global clustering coefficient.

The first time series we study is again from a Rössler systems with the same parameters as in Eqn. 4–6. The left Fig. 7 at the top shows our results averaged over $E = 25$ independent time series. Here the black dots correspond to the mean value of the clustering coefficient and the error bars give its standard deviation. Interestingly, the mean clustering coefficient does not only increase in the first step from $n_e = 1$ to $n_e = 2$ but its standard deviation decreases considerably. This indicates a stabilizing effect of longer episodes on the constructed episode networks. That means despite the fact that 25 different time series have been used, the resulting networks become more similar to each other for $n_e \in \{2, \ldots, 15\}$
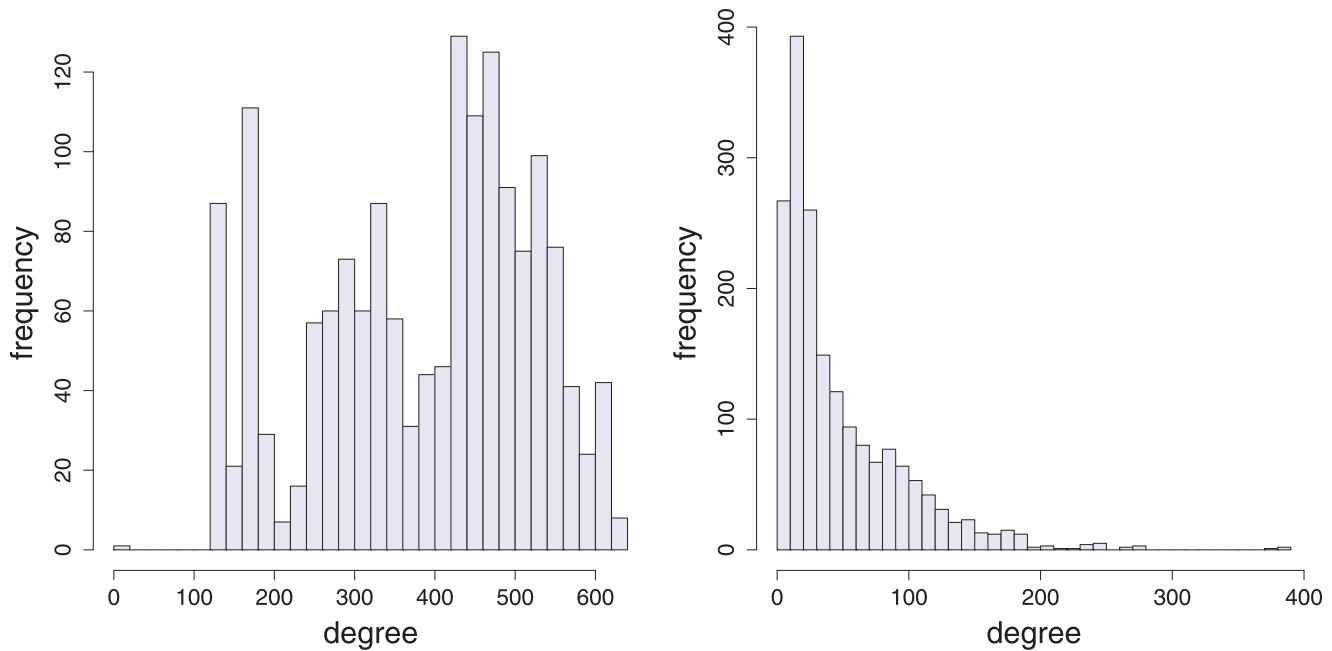
Figure 4. The degree distribution of episode networks. Left: Rössler system with 1607 episodes. Right: Duffing map with 1811 episodes.
doi:10.1371/journal.pone.0027733.g004

than for other values of n_e. This kind of robustness is desirable because the networks should reflect characteristics of the underlying dynamical system, rather than only of the individual time series used, which provides merely a sample thereof. In order to provide a quantitative cut-off value of the effective episode length, we use algorithm 1 to calculate the function h. The result of this is included in Fig. 7 showing the function h in red. The maximal value of n_e for which h is maximal is indicated as a vertical line, corresponding to an effective length of $n_e^{el} = 15$.

For the above analysis we used time series of a fixed length L, and constructed different episode networks for different values of n_e. Due to the fact that the time series is of fixed length L, the networks constructed for larger values of n_e are smaller (consist of fewer nodes) than the networks for smaller n_e values. In order to study if this effects the obtained results we repeated the above analysis, however, this time we keep the size (number of nodes) of each constructed episode network fixed. That means we need to generate larger time series for larger values of n_e. More precisely, in order to maintain a constant size of the episode networks, the length of the time series used to construct an episode network for n_e needs to be of length $L \times n_e$. The right Fig. 7 at the top shows the results of this analysis, again averaged over 25 independent simulations. Due to the larger length of the time series used for this new analysis the variances are in general smaller. This leads to a more conservative estimate of n_e^{el} which is in this case 5.

As a second example we determine the effective episode length for time series from Lorenz equations, see Eqn. 9 to 11. As time series we use again the x-component of the Lorenz equations. The results of our analysis are shown in the second row in Fig. 7. The obtained results are similar to the Rössler system with the difference that for the Lorenz equations the determined effective length is 5 for the fixed length time series (left figure) and the fixed size episode networks (right figure). As a general observation from our studies we note that there is always an increase in the clustering coefficient at the first step regardless of the considered time series. That means it is always beneficial to use an episode length $n_e > 1$ to construct episode networks.

The above analysis shows that the efficient episode length n_e^{el} resulting from a 'fixed length time series' can be the same as from a 'fixed size episode networks', but it does not have to. Due to the fact that the variability the in latter analysis is usually smaller the resulting estimates for n_e^{el} are more conservative and, hence, preferred. From this follows that it is advisable to use no episode lengths longer than 5 because this will lead to changing structural characteristics of the constructed networks. Using smaller values than n_e^{el} is in principle possible, however, the resulting networks are larger. That means these networks contain a similar amount of information as the episode network constructed for n_e^{el}, but they are larger because they consist of more nodes due to smaller episode lengths. Usually, larger networks consume more time for their analysis. For this reason, it is desirable to have the smallest networks possible that contain the same information.

Application: EEG data

Finally, we demonstrate that episode networks are useful in the practical analysis of time series data. We use electroencephalographic (EEG) time series data which measure the electrical activity of the brain [51,52]. From the total dataset available for a study that contains recordings from a 128-channel amplifier system [53], we select three different types of data which come from extracranial and intracranial recordings. The first type corresponds to surface EEG recordings of control patients. The second and third type are from intracranial recordings from presurgical patients measured in the hippocampus formation. Type 3 represent only seizure free intervals, whereas type 2 measures seizure activity.

For these three data types, we use a total of 96 single-channel recordings of 23.6 seconds duration. From these data we create 24 time series of length $94.4 = 4 \times 23.6$ seconds to obtain time series of a sufficient length. That means in total we construct for each of the three different data types 8 individual episode networks and average over the obtained results. This mimics three (small) populations of patients in order to estimate the variability within
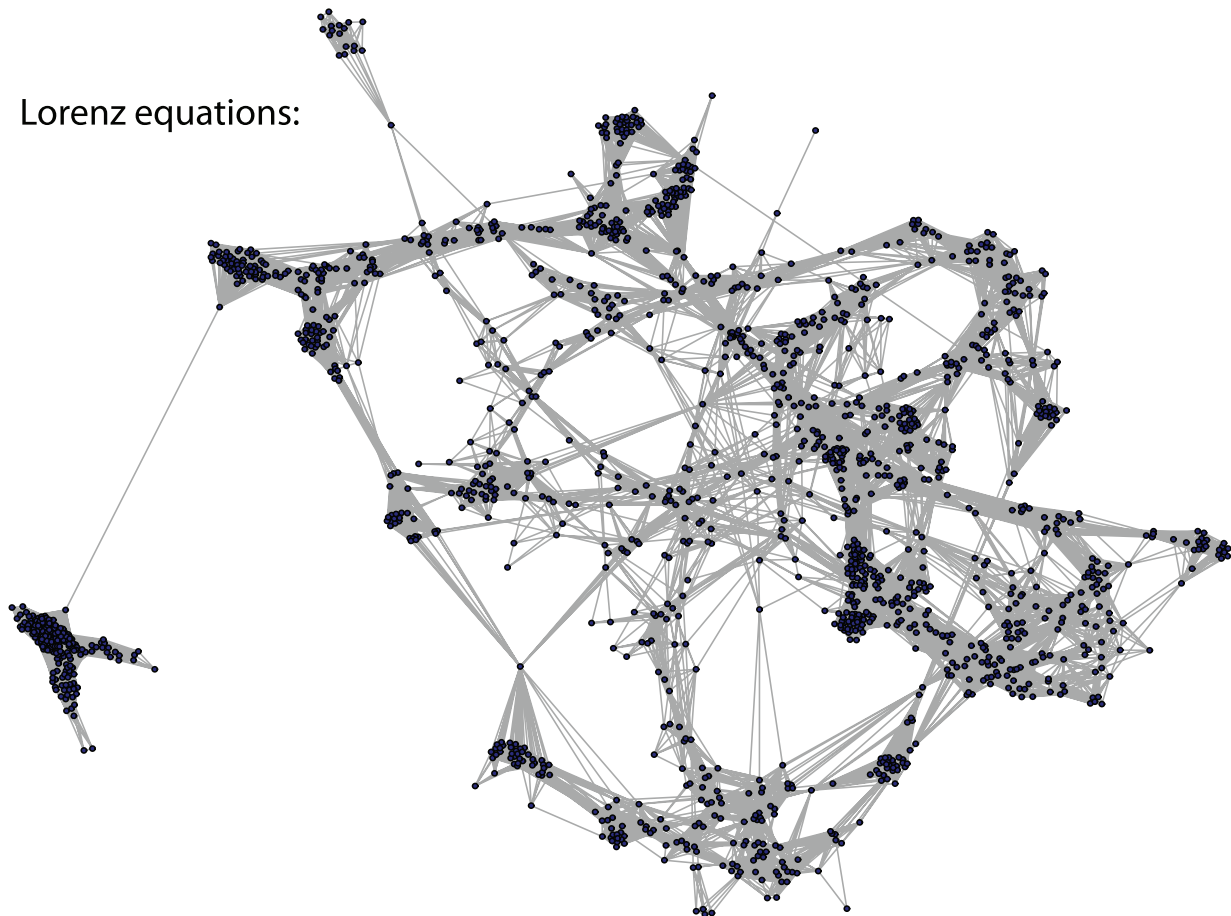
Rössler system:
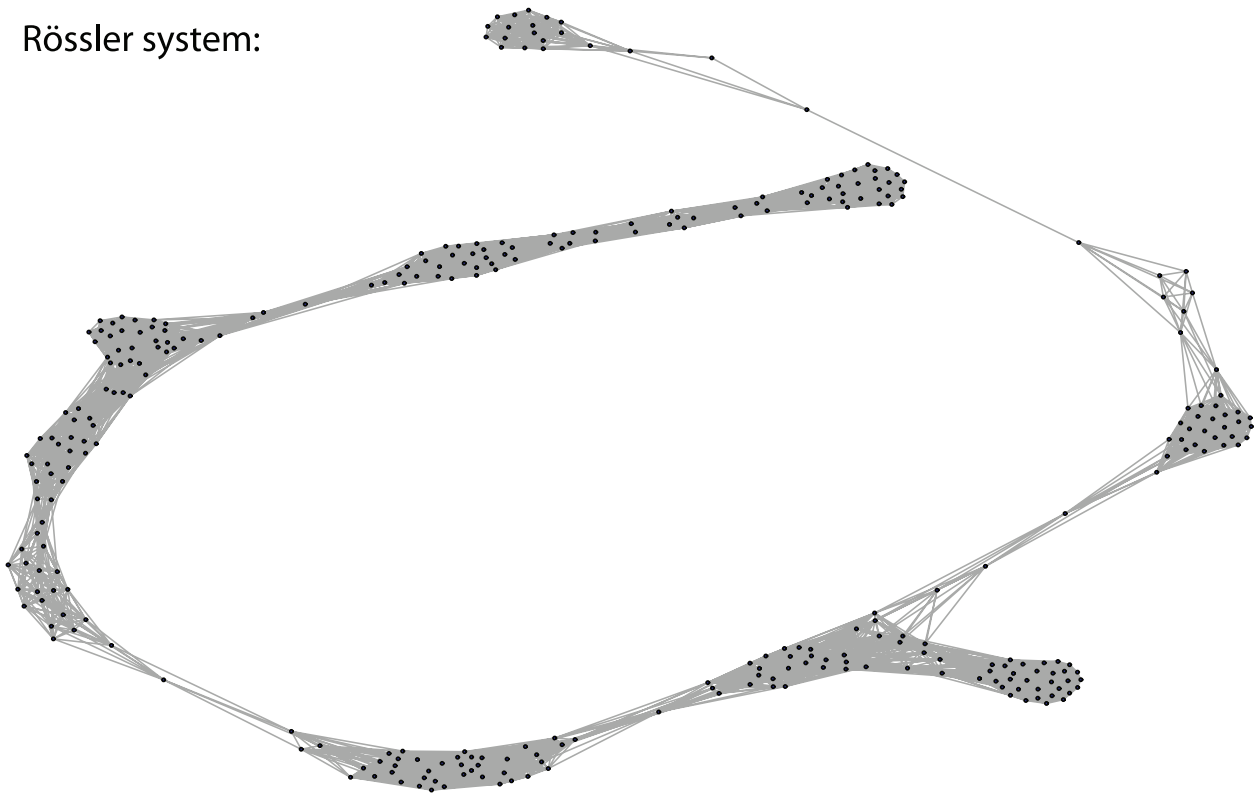
Lorenz equations:

Figure 5. Examples of episode networks. Top: The network was constructed from a Rössler system and consists of 321 nodes. Bottom: The network was constructed from Lorenz equations and consists of 1663 nodes.
doi:10.1371/journal.pone.0027733.g005

these populations. For constructing the episode networks we use as efficient length $n_e^{el} = 4$. The results from this analysis for the average path length and the clustering coefficient are shown in table 1. The variables m_i refer to the mean values whereas σ_i^2 provide the corresponding variance.

It is interesting to see that the average path length is quite similar for the first two data types. Only data type three can be clearly distinguished. The situation is different for the clustering coefficient. Here there is a clear separation between all three data types. We confirm this impression statistically by a two-sample t-test [54] comparing the mean values of the clustering coefficients. For all three tested cases, we obtain a statistically significant result assuming a significance level of $\alpha = 0.05$. That means, the null hypothesis assuming equal means among the groups is rejected. The selected significance level of $\alpha = 0.05$, which implies an expected false positive rate of 5%, appears reasonable in the given context. It is interesting to note that the variance of all variables is quite small indicating that each of the 8 networks comprising a population is representative for the other networks of the population because the measured properties are very similar. This is a desirable property, because it allows an experimental design with small sample sizes involving only a few patients. Especially, for invasive procedures this is an advantage because such procedures are usually accompanied by severe discomfort for the patients and costs for Public Health.

In the top row in Fig. 8 we show the distribution of mutual information values for three episode networks, one for each of the three data types. From these figures one gets the impression that all networks follow a power law distribution in the tails for their corresponding mutual information values. This is quantitatively confirmed by the statistical test suggested in [55] resulting in the p-
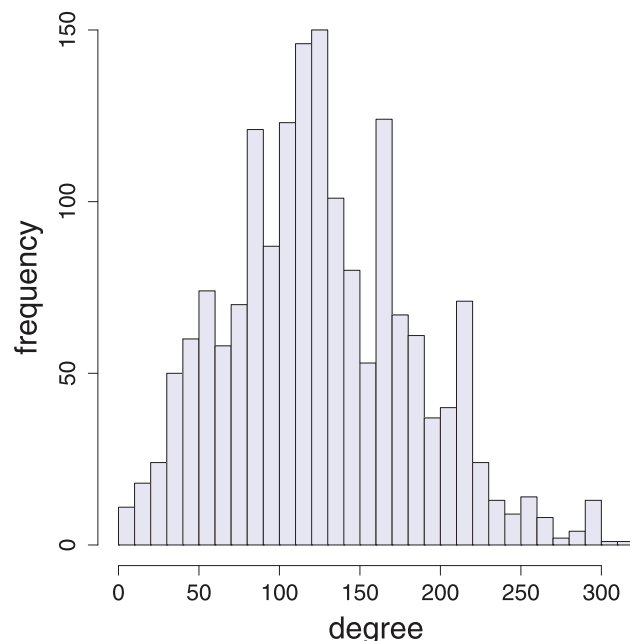
values 0.063, 0.280 and 0.154 for the cut-off values 1.019, 0.96 and 1.02. That means the tails of these mutual information distributions follow a power law, $\sim I^{-\alpha}$, and are not only long tails as for the Rössler system and the Duffing map.

In the second row in Fig. 8 we show the histograms of the degrees for these networks. It is interesting to see that despite the similarity of the distribution of mutual information values for all networks, their degree distributions are remarkably dissimilar. The network from the control patients (left most figure) seems to be between the other two networks with respect to the degree distribution. Quantitatively, this is confirmed by the mean value of the distributions which are 51.8 (left figure - data type 1), 24.9 (middle figure - data type 2) and 238.2 (right figure - data type 3). This observation is plausible because it means that during a seizure the EEG time series becomes more irregular and, hence, the similarity between different episodes is reduced. This leads to a reduction in the connectivity in the episode networks, which can be directly observed in the degree distribution in the middle Fig. 8. In contrast, during seizure free intervals the episodes become more similar leading to an increase in the connectivity of the episode network. However, it is less obvious that episodes of patients for time series of seizure free intervals (data type 3) are more similar than for control patients (data type 1). One implication from this observation is that even for seizure free intervals the EEG activity of such patients is considerably different compared to control patients. This might be a property useful for diagnostic purposes.

Discussion

In summary, in this paper we introduced a method to construct networks from pseudoperiodic time series, which we called *episode networks*. Our method is parametric allowing for the adjustment of the length of an episode, which defines the nodes in the network, and, hence, allows for the modification of the size of a network. We demonstrated, numerically, that it is always beneficial to use an episode length longer than one cycle and we defined the *effective* length of an episode as the solution of an optimization problem for the measure h. The measure h was defined reflecting the average global clustering coefficients of episode networks and their variability with respect to a population of episode networks. Using the *optimal* value n_e^{el} as an episode length to construct an episode network leads to the smallest network that contains, approximately, the same information as larger episode networks, because networks constructed for an episode length smaller than n_e^{el} lead to similar values of h, but larger networks. Another novel feature of our construction method is that it employes estimates of mutual information values to assess the similarity between different intervals of the time series, to construct the connectivity among the nodes. This allows to capture nonlinearities that are doubtlessly present in pseudoperiodic time series. From the application of our network construction method to data from EEG experiments we found that the episode networks corresponding to different groups (patients or control) capture discriminative information from the underlying time series allowing a clear distinction from each other. Furthermore, the revealed differences in the degree distributions might be useful for diagnostic purposes. However, additional studies are necessary with data from independently conducted EEG experiments to establish the robustness of the obtained results with respect to
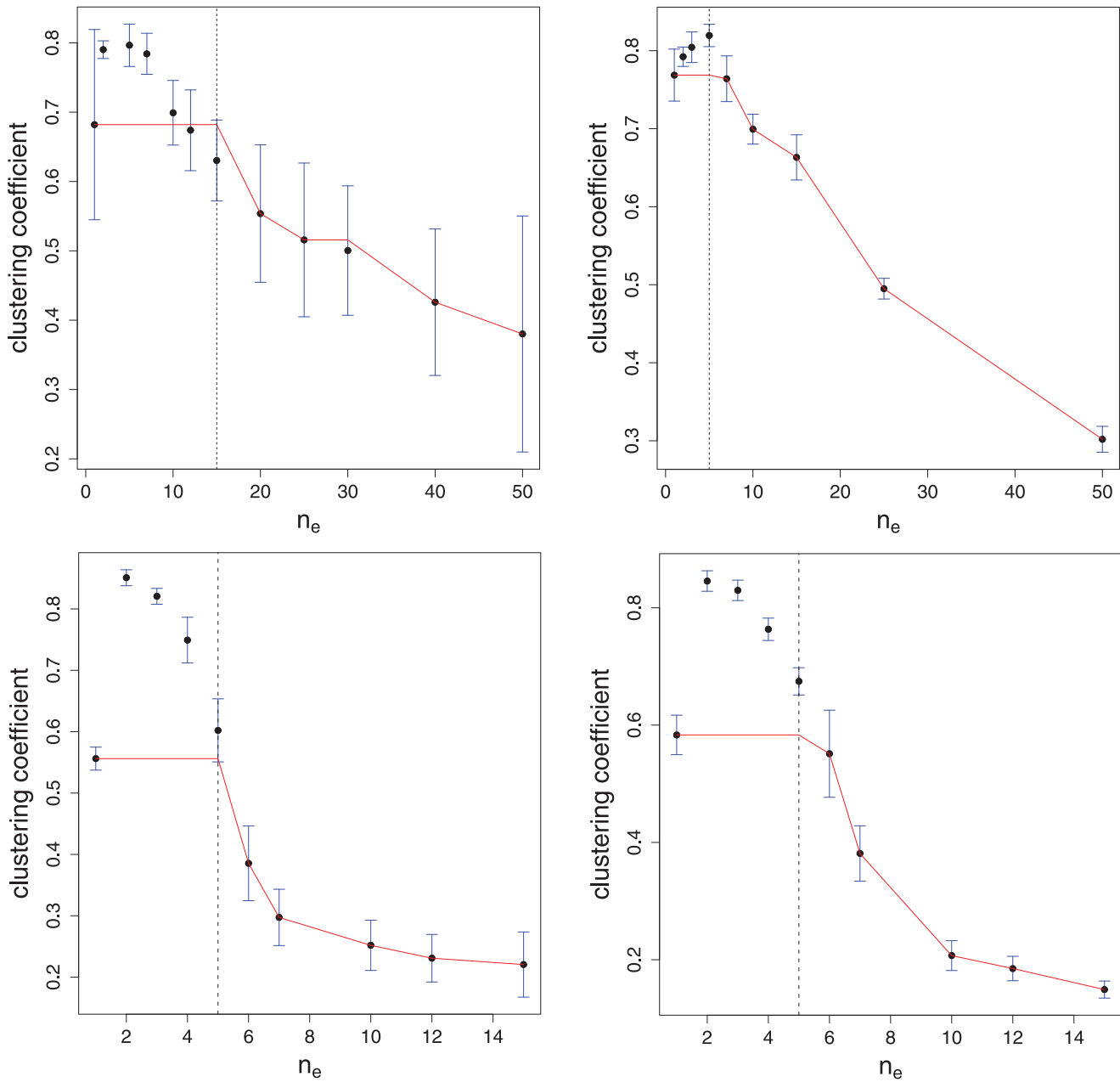

Figure 6. The degree distribution of the episode network constructed from the Lorenz equations, shown in Fig. 5.
doi:10.1371/journal.pone.0027733.g006

Figure 7. Efficient length of an episode. First row: Rössler system. Left: $n_e^{el} = 15$. Right: $n_e^{el} = 5$. Second row: Lorenz equations. Left: $n_e^{el} = 5$. Right: $n_e^{el} = 5$.
doi:10.1371/journal.pone.0027733.g007

Table 1. Numerical results obtained for the EEG dataset [53].

measure	m_1	σ_1^2	m_2	σ_2^2	m_3	σ_3^2
average path length	1.930	0.018	1.934	0.006	1.613	0.014
clustering coefficient	0.393	0.017	0.287	0.005	0.542	0.007

The index of the measures refers to one of the three data types used. m_i corresponds to the mean value of the measure and σ_i^2 to its variance. The indices correspond to: 1: control group; 2: patients - seizure free intervals; 3: patients - seizure intervals.
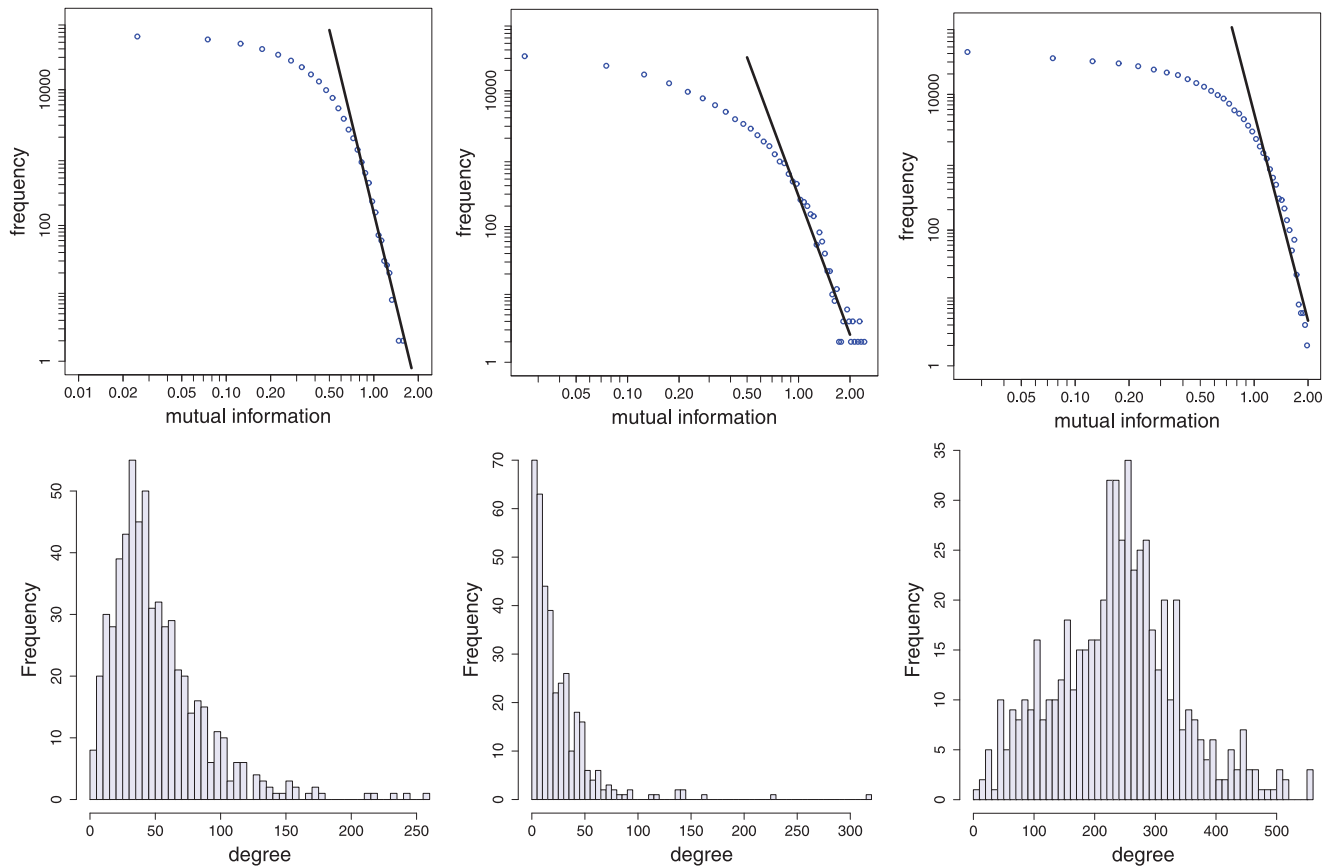doi:10.1371/journal.pone.0027733.t001

Figure 8. Results for the EEG dataset [53]. Left column: data type 1 (control group). Middle column: data type 2 (patients - seizure free intervals). Right column: data type 3 (patients - seizure intervals). The first row shows the distribution of mutual information values. The second row shows the histogram of degrees.
doi:10.1371/journal.pone.0027733.g008

varying experimental settings and protocols. This could also provide valuable insights into the experimental design of EEG experiments and differences among them, leading to a stratification in the way that our method could be applied to data from a certain subgroup of experimental designs.

From the above discussion one might feel tempted to ask if the traditional methodology for time series analysis [56,57] should be substituted by the structural analysis of networks constructed from an underlying time series. However, we do not think that it is necessary to substitute one approach by the other, instead, the analysis of networks constructed from time series data should be considered as a valuable addition to the standard methodology for time series analysis. Interestingly, from a statistical perspective there is, in fact, nothing special about the usage of networks as representation of the data. In the context of gene expression data this has been demonstrated by [58,59] showing that a correlation matrix can be interpreted as a weighted network that contains meaningful information about the interactions among genes. If seen from this angle, networks form an integral part of many methods in multivariate analysis [60].

There are several other methods that have been introduced in recent years to construct a network from time series data [30,31,33] following the spirit of [34]. Due to the fact that time series data are available in many different fields, e.g., biology, chemistry, physics, medicine or the social sciences, methods to convert these data into networks in order to enable a subsequent analysis are certainly of interest for a large variety of different application domains. Specifically, in molecular biology, the concentration of mRNAs is measured by DNA microarrays allowing genome-wide expression levels of all genes to be obtained. For this reason, periodic processes like the cell cycle or the circadian rhythm could be studied by means of *episode networks* [61–63]. From a theoretical perspective, it would be interesting to investigate in-depth the structural properties of *visibility*, *cycle* and *episode networks* to gain a thorough understanding of the coupling between their features and the properties of the time series, with respect to their generation processes.

Acknowledgments

Author Contributions

Conceived and designed the experiments: FES. Performed the experiments: FES. Analyzed the data: FES. Contributed reagents/materials/analysis tools: FES. Wrote the paper: FES.

References

1. Cayley A (1857) On the theory of analytic forms called trees. Philos Mag 13: 19–30.
2. Euler L (1736) Solutio problematis ad geometriam situs pertinentis. Comment Acad Sci U Petrop 8: 128–140.
3. Erdös P, Rényi A (1959) On random graphs. I Publicationes Mathematicae 6: 290–297.
4. Bollobas B (1998) Modern Graph Theory. New York: Springer.
5. Diestel R (2000) Graph Theory. Heidelberg: Springer.
6. Harary F (1967) Graph theory and theoretical physics. London: Academic Press.
7. Bornholdt S, Schuster H, eds. Handbook of Graphs and Networks: From the Genome to the Internet. Weinheim: Wiley-VCH.
8. Dehmer M, Emmert-Streib F, eds. Analysis of Complex Networks: From Biology to Linguistics. Weinheim: Wiley-VCH.
9. Kolaczyk E (2009) Statistical Analysis of Network Data: Methods and Models. New York: Springer.
10. Wasserman S, Faust K (1994) Social Network Analysis. Cambridge; New York: Cambridge University Press.
11. Dehmer M (2010) Structural Analysis of Complex Networks. Boston: Birkhäuser Publishing.
12. Emmert-Streib F, Glazko G (2011) Network biology: A direct approach to study biological function. Wiley Interdiscip Rev Syst Biol Med 3: 379–391.
13. Butte A, Tamayo P, Slonim D, Golub T, Kohane I (2000) Discovering functional relationships between rna expression and chemotherapeutic susceptibility using relevance networks. Proc Natl Acad Sci U S A 97: 12182–6.
14. Altay G, Emmert-Streib F (2010) Inferring the conservative causal core of gene regulatory networks. BMC Systems Biology 4: 132.
15. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, et al. (2007) Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. PLoS Biol 5.
16. Li H, Gui J (2006) Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. Biostatistics 7: 302–317.
17. Palsson B (2006) Systems Biology. Cambridge; New York: Cambridge University Press.
18. Stolovitzky G, Califano A, eds. Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference Wiley-Blackwell.
19. Werhli A, Grzegorczyk M, Husmeier D (2006) Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. Bioinformatics 22: 2523–31.
20. Xing B, van der Laan M (2005) A causal inference approach for constructing transcriptional regulatory networks. Bioinformatics 21: 4007–4013.
21. Friston K (2009) Causal modelling and brain connectivity in functional magnetic resonance imaging. PLoS Biol 7: e1000033.
22. Kaiser M, Hilgetag CC, Kötter R (2010) Hierarchy and dynamics of neural networks. Frontiers in Neuroinformatics 4.
23. Sporns O (2011) Networks of the brain. Cambridge: MIT Press.
24. Pearl J (2000) Causality: Models, Reasoning, and Inference. New York: Cambridge University Press.
25. Shipley B (2000) Cause and Correlation in Biology. New York: Cambridge University Press.
26. Emmert-Streib F, Dehmer M (2011) Networks for Systems Biology: Conceptual Connection of Data and Function. IET Systems Biology 5: 185.
27. Donner RV, Zou Y, Donges JF, Marwan N, Kurths J (2010) Recurrence networks - a novel paradigm for nonlinear time series analysis. New Journal of Physics 12: 033025.
28. Emmert-Streib F, Dehmer M (2010) Identifying Critical Financial Networks of the DJIA: Towards a Network-based Index. Complexity 16: 24–33.
29. Emmert-Streib F, Dehmer M (2010) Influence of the Time Scale on the Construction of Financial Networks. PLoS ONE 5: e12884.
30. Lacasa L, Luque B, Ballesteros F, Luque J, Nuno JC (2008) From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences 105: 4972–4975.
31. Luque B, Lacasa L, Ballesteros F, Luque J (2009) Horizontal visibility graphs: Exact results for random time series. Phys Rev E 80: 046103.
32. Mantegna R (1999) Hierarchical structure in financial markets. European Physical Journal B 11: 193–197.
33. Shirazi AH, Reza Jafari G, Davoudi J, Peinke J, Reza Rahimi Tabar M, et al. Mapping stochastic processes onto complex networks. Journal of Statistical Mechanics: Theory and Experiment 2009: P07046.
34. Zhang J, Small M (2006) Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics. Phys Rev Lett 96: 238701.
35. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69: 066138.
36. Li W (1990) Mutual information functions versus correlation functions. Journal of Statistical Physics 60: 823–837.
37. Margolin A, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7: S7.
38. Meyer P, Kontos K, Bontempi G (2007) Information-theoretic inference of large transcriptional regulatory networks. EURASIP journal on bioinformatics and systems biology 2007: 79879.
39. Cover T, Thomas J (1991) Information Theory John Wiley & Sons, Inc.
40. Shannon C, Weaver W (1949) The Mathematical Theory of Communication. Champaign: University of Illinois Press.
41. Gallager R (1968) Information Theory and Reliable Communication. New York: Wiley.
42. Cormen T, Leiserson C, Rivest R, Leiserson C, Rivest R (2001) Introduction to Algorithms. Cambridge: MIT Press.
43. Zhang J, Luo X, Small M (2006) Detecting chaos in pseudoperiodic time series without embedding. Phys Rev E 73: 016216.
44. Altay G, Emmert-Streib F (2011) Structural Influence of gene networks on their inference: Analysis of C3NET. Biology Direct 6: 31.
45. Watkinson J, Liang K, Wang X, Zheng T, Anastassiou D (2009) Inference of regulatory gene interactions from expression data using three-way mutual information. Ann N Y Acad Sci 1158: 302–13.
46. Glass L, Mackey M (1988) From Clocks to Chaos. Princeton: Princeton University Press.
47. Newman M (2010) Networks: An Introduction. Oxford: Oxford University Press.
48. DasGupta A (2008) Asymptotic theory of statistics and probability. New York: Springer.
49. Strogatz S (1994) Nonlinear dynamics and Chaos: with applications to physics, biology, chemistry, and engineering. Reading: Addison-Wesley.
50. Yamaguchi Y, Mishima N (1984) Structure of strange attractor and homoclinic bifurcation of two-dimensional cubic map. Physics Letters A 104: 179–183.
51. Fell J, Röschke J, Mann K, Schäffner C (1996) Discrimination of sleep stages: a comparison between spectral and nonlinear eeg measures. Electroencephalography and Clinical Neurophysiology 98: 401–410.
52. Stam C (2005) Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clinical Neurophysiology 116: 2266–2301.
53. Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, et al. (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Phys Rev E 64: 061907.
54. Sheskin DJ (2004) Handbook of Parametric and Nonparametric Statistical Procedures, RC Press, Boca Raton, FL, 3rd edition.
55. Clauset A, Shalizi CR, Newman MEJ (2007) Power-law distributions in empirical data. SIAM Reviews.
56. Kantz H, Schreiber T (2004) Nonlinear Time Series Analysis. New York: Cambridge University Press.
57. Shumway RH, Stoffer DS (2000) Time Series Analysis and its Applications. New York: Springer.
58. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between coexpression modules. BMC Systems Biology 1: 54.
59. Yip A, Horvath S (2007) Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 8: 22.
60. Maridia K, Kent J, Bibby J (2003) Multivariate Analysis. San Diego: Academic Press.
61. Keegan KP, Pradhan S, Wang JP, Allada R (2007) Meta-analysis of drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. PLoS Comput Biol 3: e208.
62. Spellman P, Sherlock G, Zhang M, Iyer V, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. Mol Biol of the Cell 9: 3273–3297.
63. Yang R, Su Z (2010) Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics 26: i168–i174.
64. R Development Core Team (2008) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0.
65. Csardi G, Nepusz T (2008) The igraph software package for complex network research. InterJournal, Complex Systems 1695.