

Duality between Time Series and Networks

Andriana S. L. O. Campanharo^{1,2,3}, M. Irmak Sirer², R. Dean Malmgren^{2,4}, Fernando M. Ramos⁵,
Luís A. Nunes Amaral^{2,6,7*}

1 Laboratory for Computing and Applied Mathematics, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil, **2** Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, United States of America, **3** Department of Computer Engineering, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, **4** Datascope Analytics, Chicago, Illinois, United States of America, **5** Center for Earth System Science, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil, **6** Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, United States of America, **7** Howard Hughes Medical Institute, Northwestern University, Evanston, Illinois, United States of America

Abstract

Studying the interaction between a system's components and the temporal evolution of the system are two common ways to uncover and characterize its internal workings. Recently, several maps from a time series to a network have been proposed with the intent of using network metrics to characterize time series. Although these maps demonstrate that different time series result in networks with distinct topological properties, it remains unclear how these topological properties relate to the original time series. Here, we propose a map from a time series to a network with an approximate inverse operation, making it possible to use network statistics to characterize time series and time series statistics to characterize networks. As a proof of concept, we generate an ensemble of time series ranging from periodic to random and confirm that application of the proposed map retains much of the information encoded in the original time series (or networks) after application of the map (or its inverse). Our results suggest that network analysis can be used to distinguish different dynamic regimes in time series and, perhaps more importantly, time series analysis can provide a powerful set of tools that augment the traditional network analysis toolkit to quantify networks in new and useful ways.

Citation: Campanharo ASLO, Sirer MI, Malmgren RD, Ramos FM, Amaral LAN (2011) Duality between Time Series and Networks. PLoS ONE 6(8): e23378. doi:10.1371/journal.pone.0023378

Editor: Matjaz Perc, University of Maribor, Slovenia

Received: June 29, 2011; **Accepted:** July 14, 2011; **Published:** August 11, 2011

Copyright: © 2011 Campanharo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil - doctoral fellowship), and National Science Foundation (NSF) award SBE 0624318. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: amaral@northwestern.edu

Introduction

In the context of dynamical systems, time series analysis is frequently used to identify the underlying nature of a phenomenon of interest from a sequence of observations and to forecast future outcomes. Over time, researchers accumulated a large number of time series analysis techniques, ranging from time-frequency methods, such as Fourier and wavelet transforms [1–3], to nonlinear methods, such as phase-space embeddings, Lyapunov exponents, correlation dimensions and entropies [4–6]. These techniques allow researchers to summarize the characteristics of a time series into compact metrics, which can then be used to understand the dynamics or predict how the system will evolve with time.

Obviously, these measures do not preserve all of the properties of a time series, so there is considerable research toward developing novel metrics that capture additional information or quantify time series in new ways [7–10]. One of the most interesting advances is mapping a time series into a network, based on different concepts such as correlations [11,12], visibility [13,14], recurrence analysis [15], transition probabilities [16–18] and phase-space reconstructions [19,20] (a complete list of all the proposed maps can be found in Donner *et al.*, (2010) [21] and references therein). These studies have demonstrated that distinct features of a time series can be mapped onto networks with distinct topological properties. This finding suggests that it may be possible to differentiate properties of time series using network measures.

However, it remains unclear, for example, how these topological properties relate to the original time series.

At the root of this issue is the fact that most of these maps $\mathcal{M} : \mathcal{T} \rightarrow \mathcal{G}$ from the time series domain \mathcal{T} to the network domain \mathcal{G} do not have a natural inverse operation $\mathcal{M}^{-1} : \mathcal{G} \rightarrow \mathcal{T}$. Recently, some attempts to construct an invertible map have been proposed [18,22,23]. However, they are either sensitive to arbitrarily chosen parameters [22,23] or they use information obtained from a given map \mathcal{M} to build an inverse operation \mathcal{M}^{-1} [18]. Consequently, they are not applicable to real world networks, where \mathcal{M} is not known in advance.

A fully invertible map makes it possible to create a “dual” representation of a time series and its network counterpart and directly relate common network statistics back to the original time series and vice-versa. This dual representation would not only allow time series analysis to benefit from the recent surge in network related research [24,25], but network theory would be able to draw on more than three centuries of theoretical and applied developments in time series analysis. In this paper, we take a significant step toward realizing this goal by introducing a map from time series to networks that has a natural and robust inverse.

Methods

Let \mathcal{M} be a map from a continuous time series $X \in \mathcal{T}$ to a network $g \in \mathcal{G}$, where $X = \{x(t) | t \in \mathbb{N}, x(t) \in \mathbb{R}\}$ and $g = \{\mathcal{N}, \mathcal{A}\}$ consists of a set of nodes \mathcal{N} and arcs \mathcal{A} . Ideally, such a map would

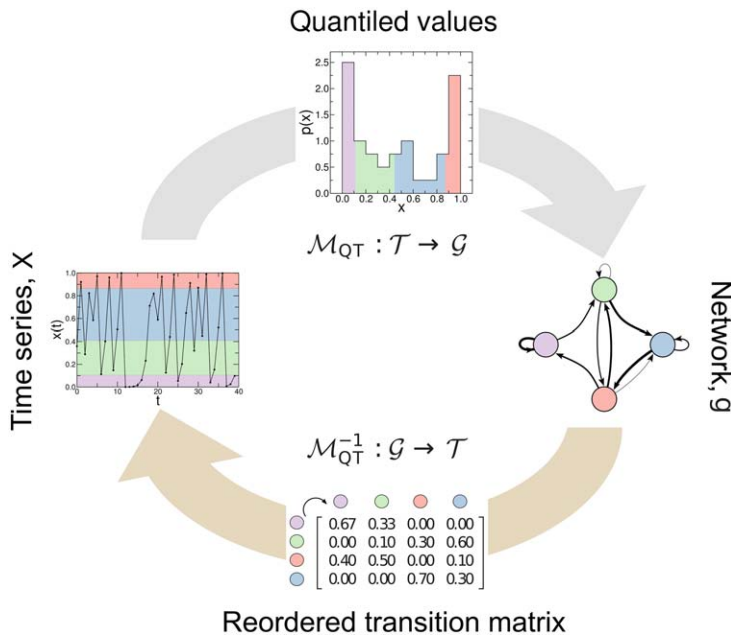


Figure 1. Illustration of the proposed map. Forward map: A time series X is split into $Q=4$ quantiles (colored shading) and each quantile q_i is assigned to a node $n_i \in \mathcal{N}$ in the corresponding network g . Two nodes n_i and n_j are then connected in the network with a weighted arc $(n_i, n_j, w_{ij}) \in \mathcal{A}$ where the weight w_{ij} of the arc is given by the probability that a point in quantile q_i is followed by a point in quantile q_j . Repeated transitions between quantiles results in arcs in the network with larger weights (represented by thicker lines). Inverse map: The weighted adjacency matrix W of network g is first normalized such that it is a Markov transition matrix with $\sum_j w_{ij} = 1$. The association between nodes and quantiles is obtained by reordering W to have large w_{ij} near to the diagonal such that the resulting time series is as continuously smooth as possible [29]. The time series is constructed by repeatedly moving from node n_i to node n_j with probability w_{ij} and choosing a random number from the corresponding quantile q_j until we have obtained a time series of length T . doi:10.1371/journal.pone.0023378.g001

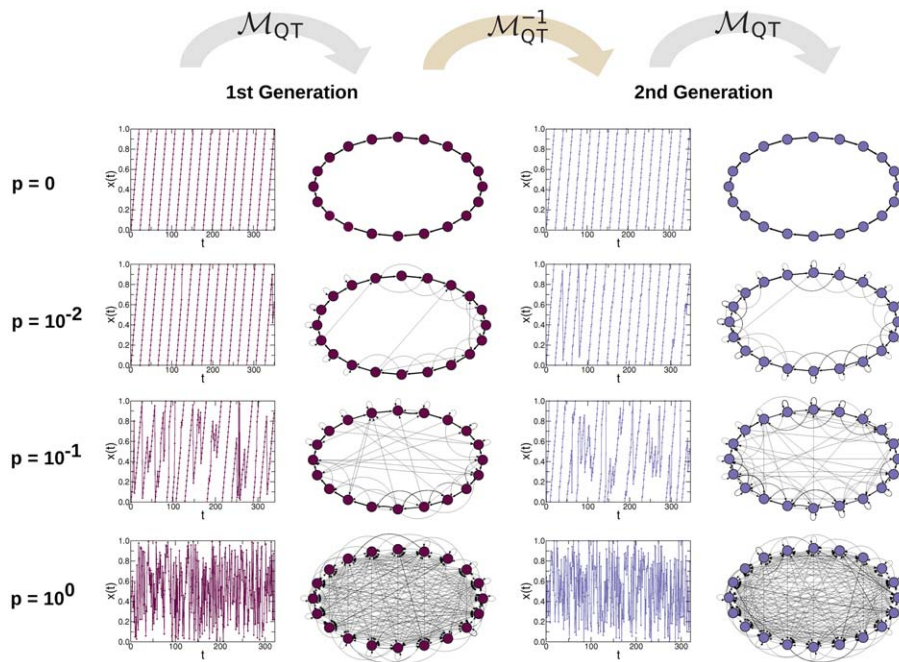


Figure 2. Qualitative assessment of the faithfulness of the proposed map and its inverse. We generate first generation time series from the toy time series model (Eq. 2) ranging from periodic ($p=0$) to random ($p=1$) with $T=320$ and $\delta=0.05$. We then construct the first generation networks using $Q=20$ quantiles by applying M_{QT} from the corresponding time series. Time series with different values of p result in networks with different topologies. As the toy time series becomes more random, the corresponding networks also become increasingly random. We construct the second generation time series and the second generation networks by sequentially applying M_{QT}^{-1} and M_{QT} , respectively. These panels suggest that the first and second generation time series and networks have similar properties, supporting the hypothesis that it may be possible to use time series analysis to characterize the topology of networks and networks analysis to characterize the structure of time series. doi:10.1371/journal.pone.0023378.g002

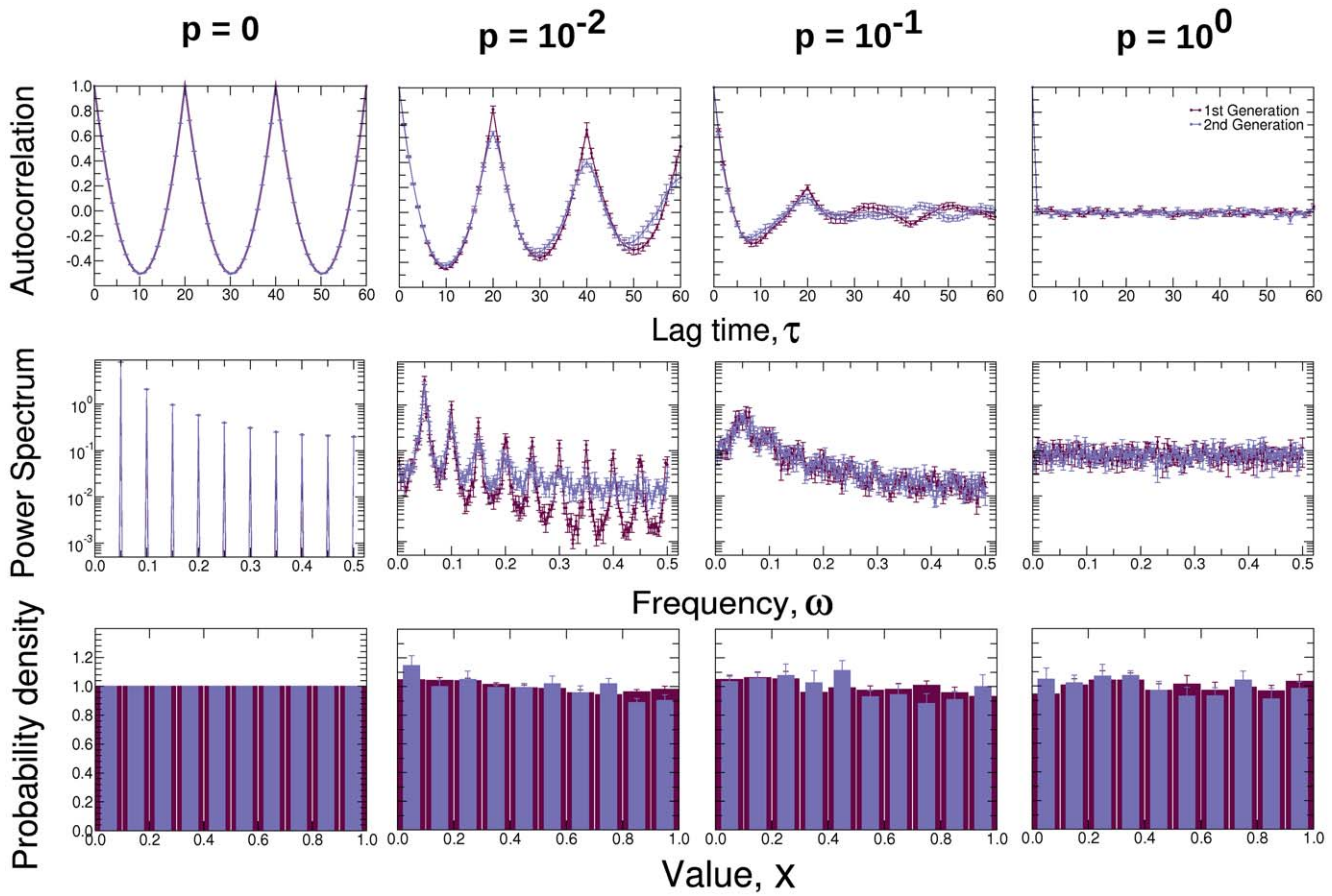


Figure 3. Comparison of statistical properties of first generation and second generation time series. We compare the means of these properties over 10 different realizations of first and second generation time series. Error bars denote standard deviation across realizations. For both the first and second generation time series, the autocorrelation function and the power spectrum reveal a distinct signal when the time series are periodic ($p=0$), which disappears when the time series become random ($p=1$). As expected from the toy model that has no biases toward particular values, both the first and second generation time series have values that are uniformly distributed between 0 and 1 for all values of p . doi:10.1371/journal.pone.0023378.g003

preserve *all* information of the original time series, possibly by a bijective map \mathcal{M}_b where each time series X maps to exactly one network $g = \mathcal{M}_b[X]$ that is invertibly mapped to the exact same time series $X = \mathcal{M}_b^{-1}[\mathcal{M}_b[X]]$. In practice, this is impossible; continuous time series have uncountably many values whereas networks are limited to a countable set of nodes \mathcal{N} and connections \mathcal{A} between them. Thus, any map from a continuous time series X to a network g must discretize the time series in some manner. Here, we use a simple discretization of X that is not sensitive to the distribution of its values. Specifically, given a time series X , we identify its Q quantiles and assign each quantile q_i to a node $n_i \in \mathcal{N}$ in the corresponding network. Two nodes n_i and n_j are then connected in the network with a weighted arc $(n_i, n_j, w_{ij}) \in \mathcal{A}$ where the weight w_{ij} of each arc is the transition probability in a Markov model estimated from the aggregate time series (Fig. 1).

The proposed map, here denoted by \mathcal{M}_{QT} , has two important properties. First, it is surjective. Given a time series X with $t = 1, 2, \dots, T$ points and the number of quantiles Q , the map will produce one and only one network $g = \mathcal{M}_{QT}[X]$. Note that distinct time series X and X' can be mapped onto the same network $g = \mathcal{M}_{QT}[X] = \mathcal{M}_{QT}[X']$ although the network space is large enough that this does not typically happen in practice.

Second, if $T > Q$, the resulting network is weighted, directed and connected. Third, \mathcal{M}_{QT} is insensitive to the distribution of values of X . The “forward” map only requires the specification of the parameter Q . This is in contrast to the maps proposed earlier, where the structures of the resulting networks are very sensitive to the choice of several parameters like time delay, embedding dimension and threshold distance; demanding expert guesses commonly used in techniques like phase-space reconstruction and recurrence analysis [26–28].

The map proposed here has the significant advantage that it has a “natural” inverse operation – a realization of a random walk on the network with transition probability w_{ij} given by the weighted adjacency matrix W such that $\sum_j w_{ij} = 1$ (Fig. 1). Starting from a random node, we construct a time series by performing a random walk in which the probability of moving from node n_i to node n_j is w_{ij} . If we identify each node in the network with a particular quantile in the resulting time series X , we can construct the time series by dividing its domain into Q quantiles and for each step of the random walk choosing a value within the corresponding quantile at random with uniform probability. In the absence of *a priori* knowledge of a direct correspondence between quantiles and nodes we assume smoothness in the resulting time series. In this way, nodes can be associated to quantiles by reordering the

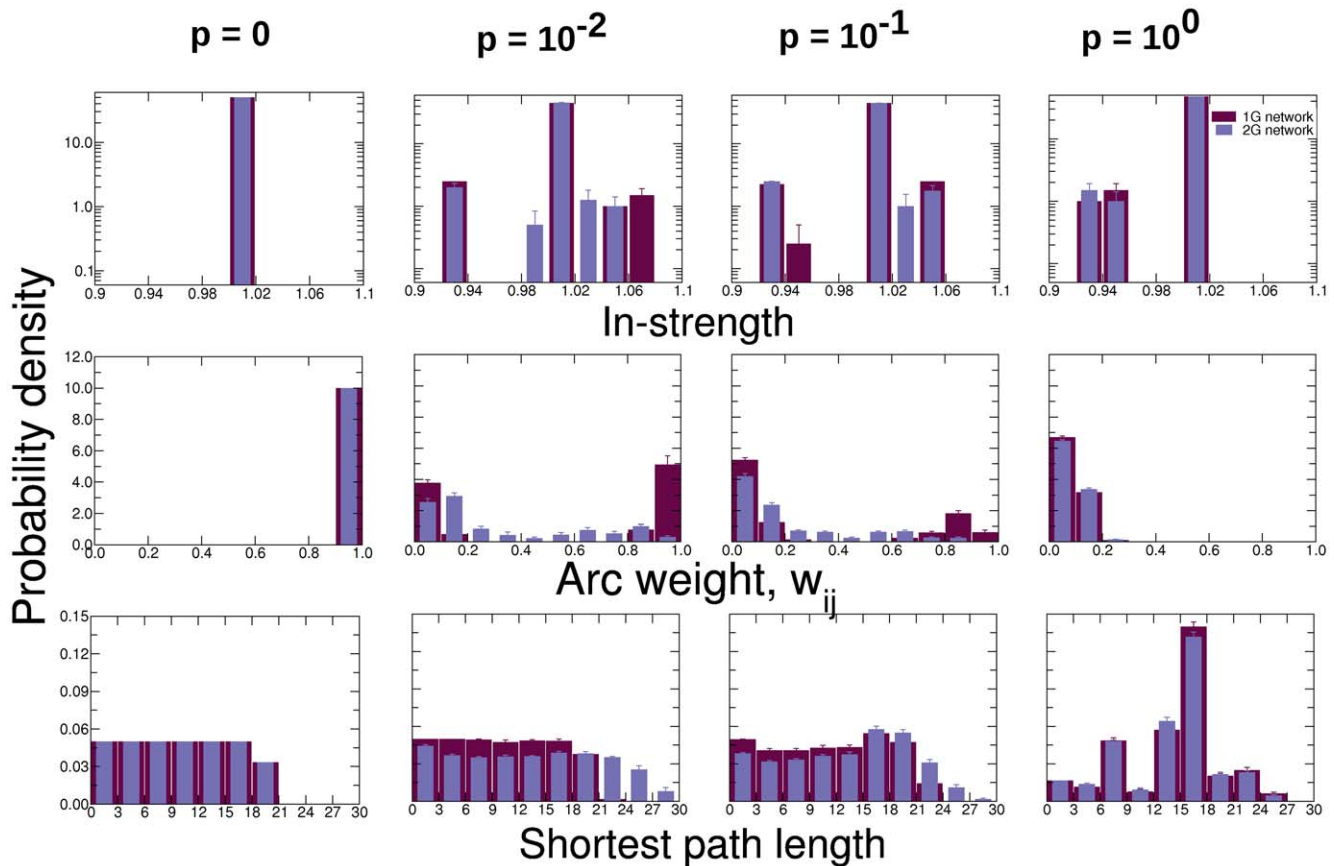


Figure 4. Comparison of topological properties of first generation and second generation networks. We compare the means of these properties over 10 different realizations of first and second generation networks. Error bars denote standard deviation across realizations. In-strength is unitary for every node when the first and second generation networks are regular ($p=0$) and, as the networks become increasingly random (left to right), the in-strength distribution broadens due to the redistribution of the weights. Note that the out-strength of a node is unitary in all cases, since the weights are Markovian probabilities. Arc weights are unitary for every arc when the first and second generation networks are regular ($p=0$). As p increases, the arc weight distribution of the first and second generation networks shows presence of small weights ($w \sim 0$) as well as large weights ($w \sim 1$). The shortest path length is calculated as the minimum sum of inverted weights on a path from one node to another. Shortest path lengths [25] are uniformly distributed when the first and second generation networks are regular ($p=0$). As p increases, random shortcuts generally decrease the distance between nodes, although for some cases, larger path lengths also arise due to redistribution of weights on the shortest path to other nodes. As the networks become more random (left to right), the shortest path distribution becomes increasingly peaked.
doi:10.1371/journal.pone.0023378.g004

weighted adjacency matrix W to have large w_{ij} near to the diagonal [29] such that the resulting time series is as “smooth” as possible – a property that is common to many empirical time series. To find the ordering of W close to the optimal ordering, we use simulated annealing [30] with a cost function that weights each element by its distance to the diagonal [31]:

$$C = \frac{1}{N} \sum_{i,j=1}^N W_{ij} |i-j|, \quad (1)$$

where N is the order of the transition probability matrix.

For every iteration in the simulated annealing search, we use $\mathcal{O}(N^2)$ moves in which segments of contiguous nodes attempts to change positions in the ordering. We accept or reject each attempted move following a standard Metropolis algorithm. For each attempt, we randomly pick: (a) a segment of contiguous nodes and (b) a new position for the first node – the remaining nodes will be placed keeping the order relative to the first node. The first node and its new position are picked from a uniform distribution; the width of the segment is picked from a Gaussian distribution

whose variance depends linearly on both the temperature T and the size of the network N – for low temperatures only changes of single nodes are proposed. We compute the value of the cost function for the new order C' and we accept the change with probability $p = \exp[(C - C')/T]$ [29].

Like \mathcal{M}_{QT} , the proposed inverse map, here denoted by \mathcal{M}_{QT}^{-1} , has several important properties. It is also surjective; given a network g the map will produce a time series $X = \mathcal{M}_{QT}^{-1}[g, \varepsilon]$ over a realization ε , but distinct networks g and g' can be mapped onto the same time series $X = \mathcal{M}_{QT}^{-1}[g, \varepsilon] = \mathcal{M}_{QT}^{-1}[g', \varepsilon']$. However, it is not strictly one-to-one since it has a stochastic element. That is, $\mathcal{M}_{QT}^{-1}[g, \varepsilon] \neq \mathcal{M}_{QT}^{-1}[g, \varepsilon']$. Note that even though the proposed map is not one-to-one, the time series obtained by applying the inverse map with different realizations will have very similar properties. In contrast, previous inverse maps [22,23] depend on the arbitrary choice of node labels and the resulting time series are highly sensitive to this choice.

Results

To verify the extent to which the properties of the original time series or network are recovered when \mathcal{M}_{QT} and \mathcal{M}_{QT}^{-1} are applied

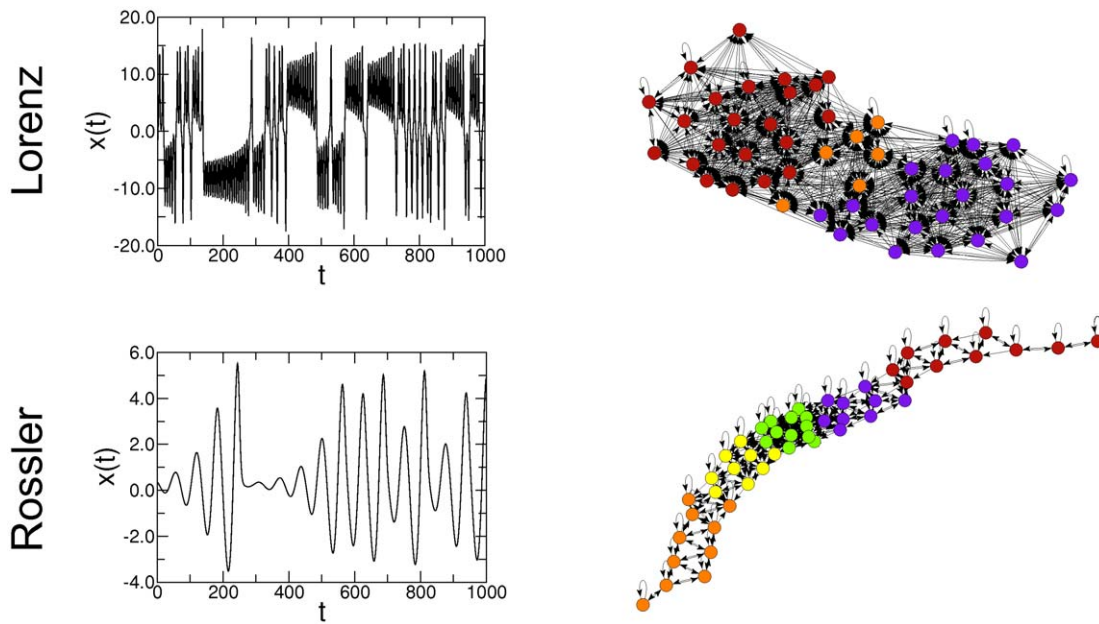


Figure 5. Illustration of the forward map M_{QT} to chaotic time series from Lorenz and Rossler systems. We use $T = 10,000$ time points of the x variable of the chaotic Lorenz and Rossler equations and construct networks using $Q = 50$ quantiles by applying the forward map. Each node is colored according to the module to which it belongs. The resulting networks display clear differences in topologies. The network of Lorenz's system is bulky with two large modules. It has a modularity value of $M = 0.4547$, that is much larger than the mean (standard error) modularity value $\langle M \rangle_D = 0.0298(0.0018)$ obtained from networks built from the randomizations of the original time series. Furthermore, the two lobes of the Lorenz attractor are mapped into the two largest connected modules in the network. On the other hand, the network of Rossler's system presents an elongated, chain-like pattern due the strong periodicity present in its corresponding time series. The network of Rossler's system is also modular, with five small modules and it has a modularity value of 0.6437 . This value is much larger than the mean (standard error) modularity value $0.0280(0.0017)$ obtained from networks built from the randomizations of the original time series.
doi:10.1371/journal.pone.0023378.g005

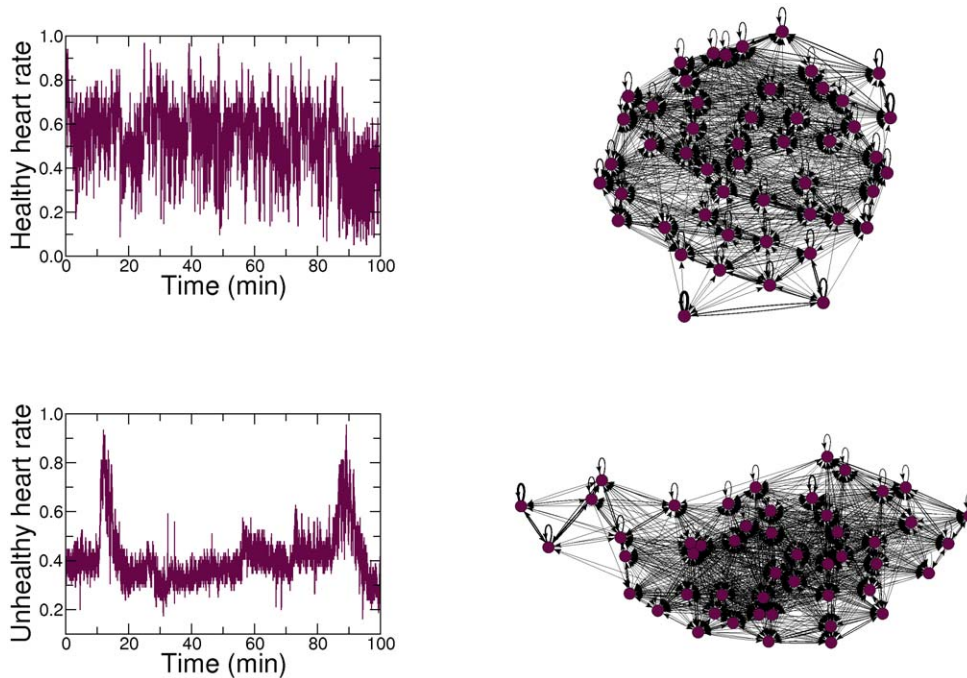


Figure 6. Illustration of the proposed forward map to the problem of detecting differences in the data structures of patients in different health conditions. We use 100-minute normalized heart rate time series from a healthy subject (upper panel) and a subject with severe congestive heart failure (lower panel) sampled every ≈ 0.01 seconds ($T = 10,000$ points) [36]. We construct the networks using $Q = 50$ quantiles by applying M_{QT} from the corresponding time series. The resulting networks display clear differences in topology, which are especially apparent on the relatively separated cluster in the network associated with the unhealthy subject. These differences in topology are confirmed by generating networks with different number of nodes (Fig. 7) and using time series from different healthy and unhealthy subjects (Fig. 8).
doi:10.1371/journal.pone.0023378.g006

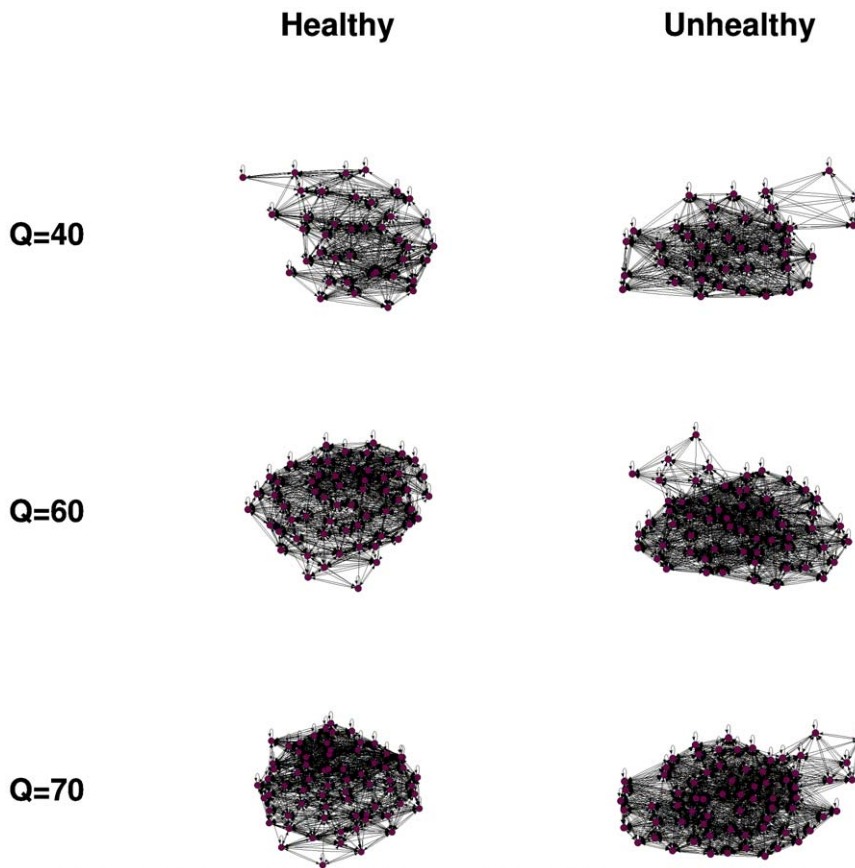


Figure 7. Application of the proposed forward map to the heart rate time series using different number of quantiles. We apply \mathcal{M}_{QT} using $Q=40,60$ and 70 nodes to time series from healthy (left panels) and unhealthy subjects (right panels). Note the visual similarity of these networks with the networks presented in Figure 6, attesting the robustness of the results of the proposed forward map, regardless of the value of Q .
doi:10.1371/journal.pone.0023378.g007

sequentially, we introduce an ensemble of time series that range from periodic to random:

$$x(t) = \begin{cases} \text{mod}(x(t-1) + \delta + \eta, 1), & \text{with probability } p \\ \text{mod}(x(t-1) + \delta, 1), & \text{otherwise} \end{cases} \quad (2)$$

where δ is a constant, p parameterizes the probability that noise modifies the otherwise periodic time series, and η is a random variable drawn from a uniform distribution in $[0,1]$. We choose $p=0, 10^{-2}, 10^{-1}$ and 10^0 and $\delta=0.05$ and generate numerous time series with $T=320$ points. We then apply the forward map with $Q=20$ quantiles to the generated time series and obtain the resulting networks. We refer to these time series and networks as the “first generation” time series and networks, respectively. Figure 2 shows that time series with different properties are mapped onto networks with visually distinct topologies. Specifically, as the time series become more random, the corresponding networks become increasingly more random, much like the small-world network model of Watts & Strogatz [32].

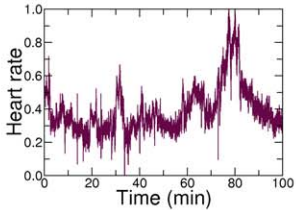
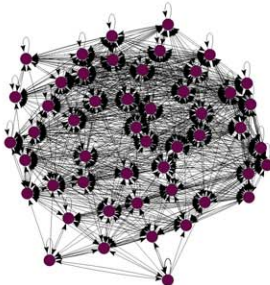
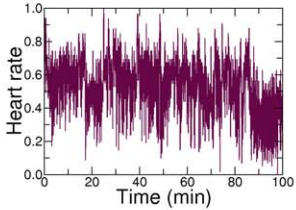
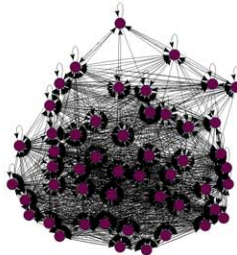
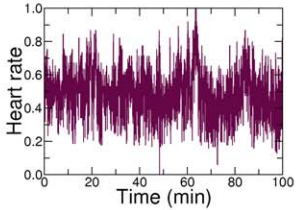
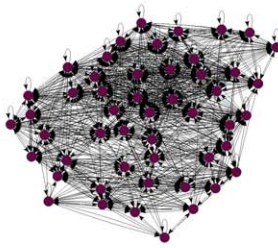
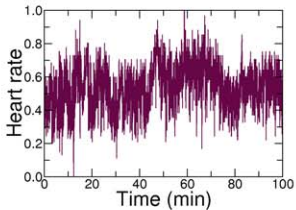
We next apply the map \mathcal{M}_{QT}^{-1} to each of the first generation networks and obtain the “second generation” time series, again with $T=320$ points. For simplicity, we assign each quantile to the

corresponding quantile from the first generation time series. The visual similarity between the first generation time series X and the second generation time series $\mathcal{M}_{QT}^{-1}[\mathcal{M}_{QT}[X], \varepsilon]$ is apparent, regardless of the value of p (Fig. 2). We quantitatively demonstrate the faithfulness of the proposed map in the time series domain \mathcal{T} by comparing the autocorrelation function, the power spectrum and the distribution of the first and second generation time series (Fig. 3).

Finally, we apply \mathcal{M}_{QT} to the second generation time series using $Q=20$ quantiles to obtain the corresponding “second generation” networks. It is visually apparent that first generation networks g and second generation networks $\mathcal{M}_{QT}[\mathcal{M}_{QT}^{-1}[g], \varepsilon]$ have similar topologies for all values of p (Fig. 2). We quantitatively demonstrate the faithfulness of the map in the network domain \mathcal{G} by comparing the in-strength, arc weight and shortest path length distributions of the first and second generation networks (Fig. 4). Our results show that the topological features of the first generation networks are recovered in the second generation networks for all values of p . The results of Figures 3 and 4 indicate that our method is able to preserve both structured and unstructured information in both the time series and network domains, even after successive mappings.

To further highlight the potential of the forward map described above, we apply it to two time series belonging to different dynamical systems. The first time series is the x variable of the

Healthy



Unhealthy

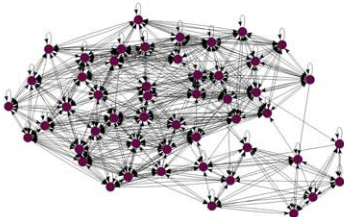
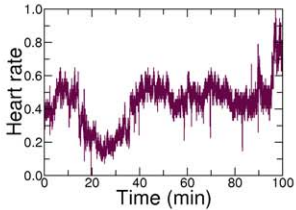
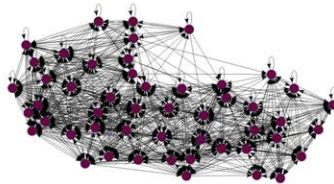
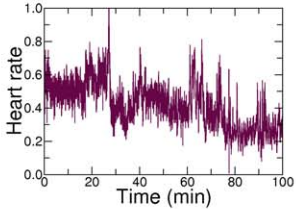


Figure 8. Application of the proposed forward map to the heart rate time series associated to different subjects. We apply \mathcal{M}_{QT} using $Q=50$ nodes to time series from three healthy (left panels) and unhealthy subjects (right panels). Regardless of the number of different subjects, the resulting networks are visually similar with those presented in Figure 6. This is another demonstration of the robustness of \mathcal{M}_{QT} . doi:10.1371/journal.pone.0023378.g008

chaotic Lorenz equations:

$$\begin{aligned} dx/dt &= \sigma(-x+y) & dx/dt &= -y-z \\ dy/dt &= rx-y-xz & dy/dt &= x+ay \\ dz/dt &= bz+xy & dz/dt &= b+z(x-c) \end{aligned} \tag{3} \tag{4}$$

with parameter values $\sigma=10$, $r=28$, and $b=8/3$. Numerical solutions of these equations leads to an attractor embedded in a three-dimensional space with coordinates (x,y,z) [33]. The trajectory rotates about one of two unstable fixed points and eventually escapes to orbit the other fixed point. This behavior is recognizable in the x variable (left panel in Fig. 5) since its values oscillate between the positive and the negative x -region.

The second time series is the x variable of the chaotic Rossler equations:

with parameter values $a=0.432$, $b=2.0$, and $c=4.0$. Its phase-space generates a chaotic attractor with a single lobe, in contrast to the Lorenz attractor which has two. The trajectory within the attractor follows an outward spiral close to the (x,y) plane around an unstable fixed point. Once the trajectory spirals out enough, a second fixed point influences it, causing a rise and twist in the z -dimension [34]. This behavior generates a quasiperiodic oscillatory pattern in the x variable, with max/min peaks/troughs with different amplitudes (left panel in Fig. 5).

In both cases, we apply the forward map with $T=10,000$ and $Q=50$ quantiles. The resulting networks (right panel in Fig. 5)

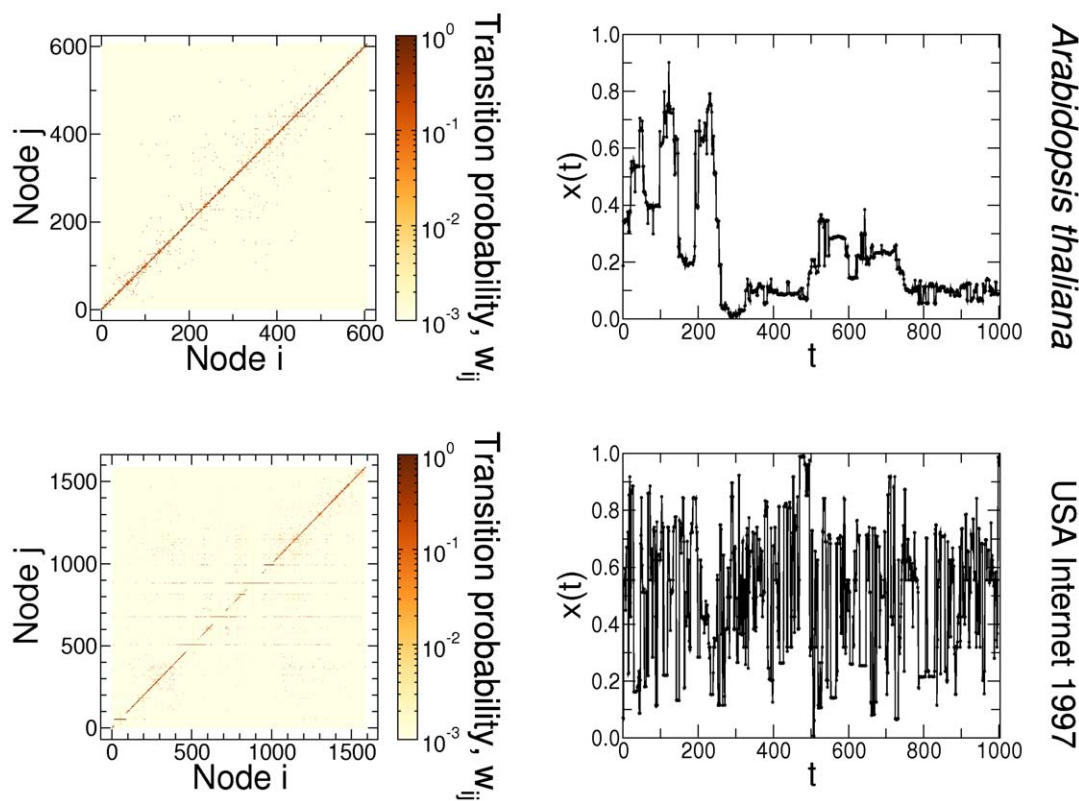


Figure 9. Illustration of the proposed inverse map to different types of real-world networks – metabolic network and the Internet. We use *Arabidopsis thaliana* network with $Q=607$ nodes and USA Internet 1997 with $Q=1,589$ nodes [37,38]. The corresponding adjacency matrices of these two networks are reordered (left panels) and time series with $T=100,000$ points each are generated by applying \mathcal{M}_{QT}^{-1} (for clarity only $T=1,000$ points are shown in the right panels). The resulting time series display clear differences in dynamics. In the first application, the topological features of the metabolic network are translated into a time-series with a high degree of persistence (or long-range correlations), due to the presence of modules in the original structure. In the second one, every time the random walker reaches one of the several hubs, it has a high probability of being sent to a different branch of the network. This behavior produces the noisy signal characteristic of low persistence (short correlations) time series. These differences in dynamics are confirmed by performing random walks over different realizations (Fig. 10), and computing their statistical properties (Fig. 11). doi:10.1371/journal.pone.0023378.g009

Arabidopsis thaliana

USA Internet 1997

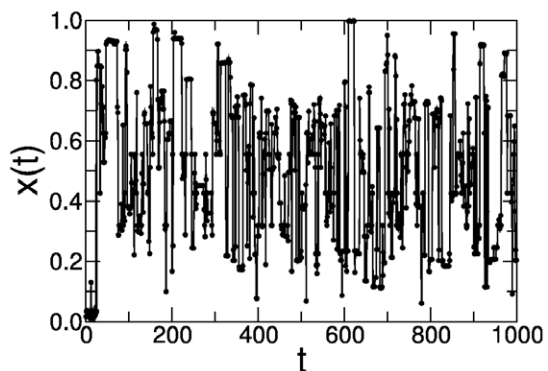
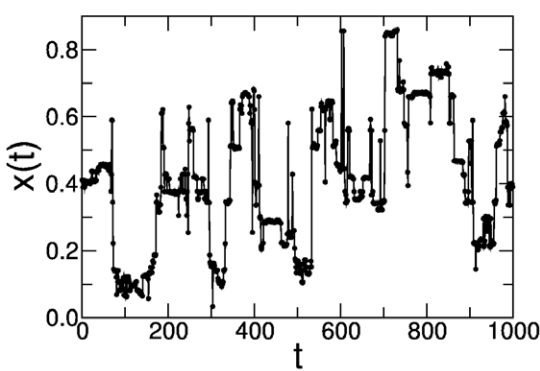
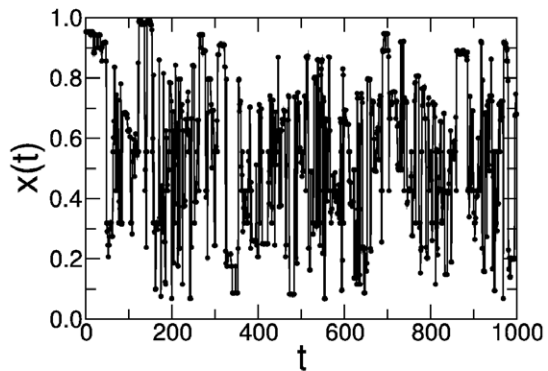
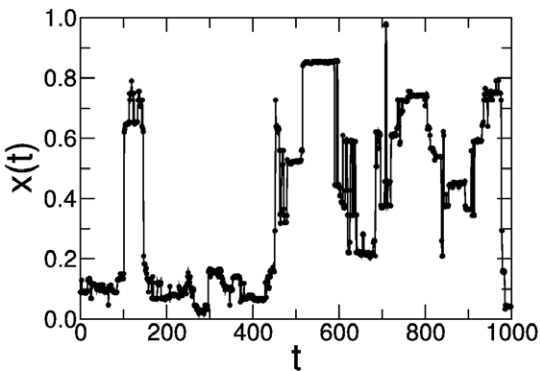
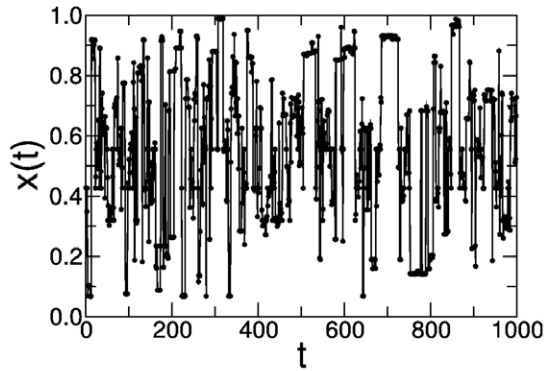
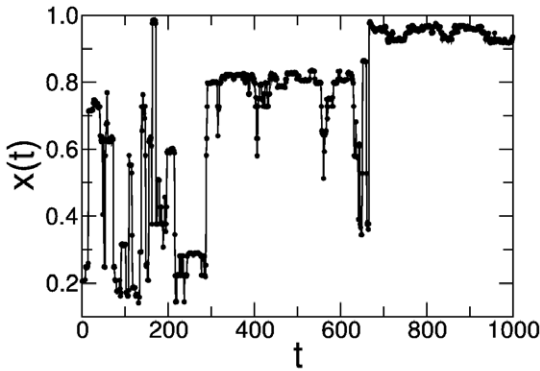
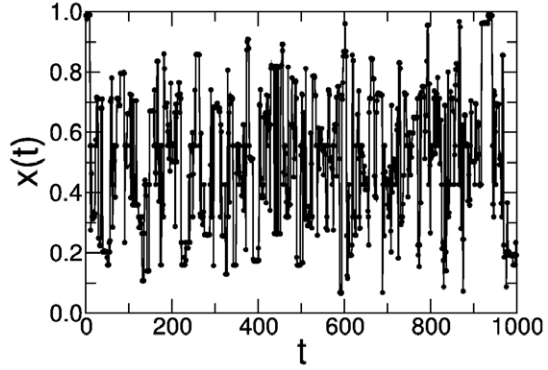
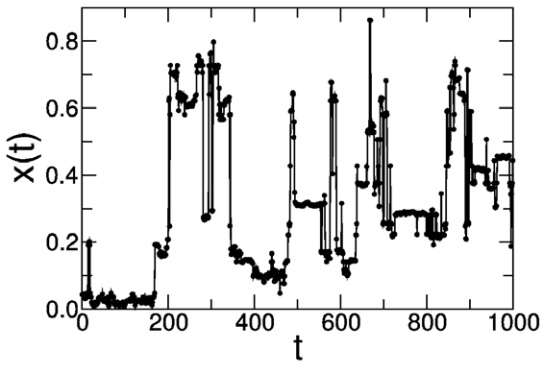


Figure 10. Different realizations of the inverse map M_{QT}^{-1} in the real networks. We perform four realizations of M_{QT}^{-1} to the *Arabidopsis thaliana* metabolic network ($Q=607$ nodes and $T=100,000$ points), and USA Internet 1997 ($Q=1,589$ nodes and $T=100,000$ points). Note the clear similarity of these time series with the time series presented in Figure 9, demonstrating the robustness of the proposed inverse map. doi:10.1371/journal.pone.0023378.g010

display clear differences in topology. The network of Lorenz's system presents a bulky structure, with the two lobes of the Lorenz attractor being mapped into the two largest connected modules in the network. On the other hand, the network of Rossler's system presents an elongated chain-like pattern which stems from the strong quasiperiodicity present in the corresponding time series. The five small modules in this network originate from the different amplitude levels generated by the Rossler attractor.

In order to further illustrate the potential for real-world applications of the forward map, we apply it to the long standing problem of detecting the subtle differences between interbeat interval time series of healthy and unhealthy subjects [35]. Specifically, we obtained two human heart rate time series from PhysioNet [36]; one from a healthy subject and one from a subject with severe congestive heart failure (Fig. 6). The healthy time series is notable for its apparent nonstationarity and "patchiness". On the other hand, congestive heart failure may be associated with the emergence of excessive regularity, as is apparent from the unhealthy time series. We apply the forward map using 100-minute heart rate time series, $T=10,000$ and $Q=50$ quantiles (Fig. 6). The resulting networks display clear differences in

topology, which are especially apparent on the relatively separated cluster in the network associated with the unhealthy subject.

We demonstrate the robustness of the results found in Figure 6 by applying M_{QT} to the healthy and unhealthy heart rate time series over different values of Q . Figure 7 suggests that the forward map is able to produce networks with similar topologies, regardless of the value of Q . As another demonstration of robustness, we apply the forward map to the different healthy and unhealthy heart rate time series. Figure 8 suggests that the forward map is able to produce networks with similar dynamics for both healthy and unhealthy subjects.

We also illustrate the potential for real-world applications of the inverse map described above by applying it to two networks belonging to different network classes (for details, see [37,38]). The first network is the metabolic network of *Arabidopsis thaliana*, with a relatively high modularity, characterized by long open "chain" or closed loops of non-hubs, and a core of a few hubs that are directly reachable from one another. The second, the Internet in 1997, which has a star-like structure with several hubs and low modularity. First, we associate nodes to quantiles by reordering the corresponding adjacency matrices [29]. Next, we obtain time

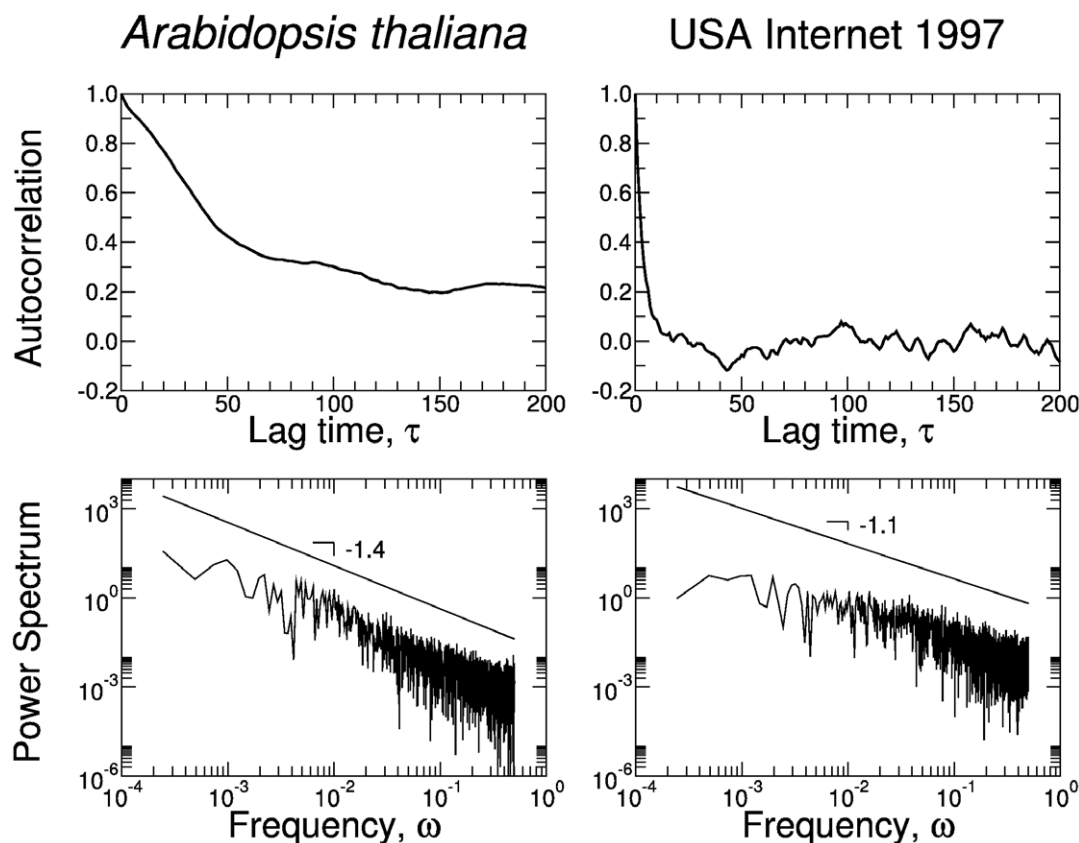


Figure 11. Statistical properties of the time series presented in Figure 9, generated from the *Arabidopsis thaliana* network and the USA Internet 1997. Note that the long-range correlations present in the metabolic network are well captured by the autocorrelation function and the corresponding power density spectrum, which displays a clear power-law scaling. On the other hand, the results in the USA Internet 1997 bear the footprint of the short-correlated signal generated by the Internet network. Note a power-law scaling with a less steep slope. doi:10.1371/journal.pone.0023378.g011

series with $T = 100,000$ points each using networks with $Q = 607$ and $Q = 1,589$ nodes, respectively (Fig. 9). The resulting time series display clear differences in dynamics, which we confirm by performing random walks over different realizations (Fig. 10), and computing their statistical properties (Fig. 11). Our results demonstrate that networks with different topologies result in time series with different dynamics.

Discussion

The proposed map can be extended to include higher-order correlations. Just as a traditional Taylor expansion approximates the value of a time series X near a particular point t_0 by evaluating the derivatives of $x(t)$ near t_0 , \mathcal{M}_{QT} resembles a “wholistic” Taylor expansion – it estimates values near a particular point t_0 by the Markovian probability that $x(t_0 + 1)$ follows $x(t_0)$ with the same accuracy for any point t_0 of the time series. Just as the precision of a Taylor expansion improves as higher-order terms in the expansion are retained, the precision of the map can be improved by incorporating higher-order Markov chains. For example, \mathcal{M}_{QT} can be readily adapted to capture second-order correlations by constructing networks from the second order Markov probability density $p(x(t_0 + 1)|x(t_0), x(t_0 - 1))$, resulting in networks with directed and weighted hyperedges connecting the nodes associated with the quantiles of $x(t_0)$ and $x(t_0 - 1)$ to the node associated with the quantile of $x(t_0)$.

It is worth mentioning that the proposed map procedure touches on a few classic analysis techniques. In some sense, it bears some resemblance to symbolic dynamics, where a continuous system is discretized into a sequence of symbols representing the state of the system [39]. In our map nodes play the role of symbols and a symbolic series is then produced by looking at a particular

path through the network. The proposed map procedure also provides a unique approach to compressing time series data. Since most financial, health and climate time series consist of millions of measurements, our map procedure naturally provides an excellent storage mechanism to compress the T points of these large time series into a list of at most Q^2 values of the Markov transition matrix W . Additional storage savings occurs when W is sufficiently sparse that it is more efficient to store a weighted edge list.

Our results build a bridge connecting time series analysis and network-related research. In this sense, networks can be analyzed by exploring an extensive set of statistical properties of the associated time series. For example, motifs in a network are mapped as periodicities in a time series, which are characterized by looking at the corresponding power spectrum of the time series. At the same time, different dynamical regimes in time series can be analyzed by exploring an extensive set of topological statistics at the associated network domain.

Acknowledgments

We thank R. Guimerà, M. Sales-Pardo, J. Duch, S. Saavedra, and the members of the Amaral lab for insightful comments and suggestions. All figures were generated with PyGrace (<http://pygrace.sourceforge.net>) with color schemes from Colorbrewer (<http://colorbrewer.org>) and Pajek (<http://vlado.fmf.unilj.si/pub/networks/pajek/>).

Author Contributions

Conceived and designed the experiments: ASLOC MIS RDM FMR LANA. Performed the experiments: ASLOC MIS RDM. Analyzed the data: ASLOC MIS RDM FMR LANA. Contributed reagents/materials/analysis tools: ASLOC MIS RDM. Wrote the paper: ASLOC MIS RDM FMR LANA.

References

- Korner TW (1988) Fourier Analysis. Cambridge: Cambridge University Press.
- Box GEP, Jenkins GM, Reinsel GC (2008) Time Series Analysis, Forecasting and Control. John Wiley & Sons, Inc.
- Percival DB, Walden AT (2000) Wavelet Methods for Time Series Analysis. Cambridge University Press.
- Strogatz SH (1994) Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering. Perseus Books Group.
- Kantz H, Schreiber T (2003) Nonlinear Time Series Analysis. Cambridge University Press.
- Campanharo ASLO, Ramos FM, Macau EEN, Rosa RR, Bolzan MJA, et al. (2008) Searching chaos and coherent structures in the atmospheric turbulence above the Amazon forest. *Phil Trans R Soc A* 366.
- Zhang J, Luo X, Small M (2006) Detecting chaos in pseudoperiodic time series without embedding. *Physical Review E* 73.
- Lai C, Chung P, Tseng VS (2010) A novel two-level clustering method for time series data analysis. *Expert Systems with Applications* 37.
- Verplancke T, Looy SV, Steurbaut K, Benoit D, Turck FD, et al. (2010) A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks. *BMC Medical Informatics and Decision Making* 10.
- Ao S (2010) Applied Time Series Analysis and Innovative Computing. Springer.
- Zhang J, Small M (2006) Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters* 96.
- Yang Y, Yang HJ (2008) Complex network-based time series analysis. *Physica A* 387.
- Lacasa L, Luque B, Ballesteros F, Luque J, Nuno JC (2008) From time series to complex networks: The visibility graph. *Proc Natl Acad Sci U S A* 105.
- Luque B, Lacasa L, Ballesteros F, Luque J (2009) Horizontal visibility graphs: Exact results for random time series. *Physical Review E* 80.
- Marwan N, Donges JF, Zou Y, Donner RV, Kurths J (2009) Complex network approach for recurrence analysis of time series. *Physics Letters A* 46.
- Nicolis G, Cantú AG, Nicolis C (2005) Dynamical aspects of interaction networks. *International Journal of Bifurcation and Chaos* 15.
- Li P, Wang BH (2007) Extracting hidden fluctuation patterns of hang seng stock index from network topologies. *Physica A* 378.
- Shirazi AH, Jafari GR, Davoudi J, Peinke J, Tabar MRR, et al. (2009) Mapping stochastic processes onto complex networks. *Journal of Statistical Mechanics: Theory and Experiment*.
- Xu X, Zhang J, Small M (2008) Superfamily phenomena and motifs of networks induced from time series. *Proc Natl Acad Sci U S A* 105.
- Gao Z, Jin N (2009) Complex network from time series based on phase space reconstruction. *Chaos*.
- Donner RV, Small M, Donges JF, Marwan N, Zou Y, et al. (2010) Recurrence-based time series analysis by means of complex network methods. Technical Report arXiv:010.6032. Comments: To be published in *International Journal of Bifurcation and Chaos* (2011).
- Strozzi F, Zaldívar JM, Poljansek K, Bono F, Gutiérrez E (2009) From complex networks to time series analysis and viceversa: Application to metabolic networks. *JRC Scientific and Technical Reports*, EUR 23947 JRC52892.
- Haraguchi Y, Shimada Y, Ikeguchi T, Aihara K (2009) Transformation from complex networks to time series using classical multidimensional scaling. In: *ICANN '09: Proceedings of the 19th International Conference on Artificial Neural Networks Heidelberg, Berlin: Springer-Verlag*.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Review* 45: 167–256.
- Costa LDF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: A survey of measurements. *Adv Phys* 56: 167–242.
- Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. *Physical Review A* 33.
- Takens F (1981) Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence* 898: 366–381.
- Eckmann JP, Kamphorst SO, Ruelle D (1987) Recurrence plots of dynamical systems. *Europhysics Letters* 4.
- Sales-Pardo M, Guimerà R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. *Proc Natl Acad Sci USA* 104: 15224–15229.
- Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220: 671–680.
- Wasserman S, Faust K (1994) *Social Network Analysis*. Cambridge, UK: Cambridge University Press.
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–442.
- Lorenz EN (1963) Deterministic nonperiodic flow. *Journal of Atmospheric Sciences* 20.
- Rossler OE (1976) An equation for continuous chaos. *Physics Letters* 57A.
- Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, et al. (2000) Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation* 101.

36. Physionet - the research resource for complex physiologic signals. Available: <http://www.physionet.org>. Accessed 2011 May, 23.
37. Guimerà R, Sales-Pardo M, Amaral L (2007) Classes of complex networks defined by role-to-role connectivity profiles. *Nature Phys* 3: 63–69.
38. Netgeo - the internet geographic database. Available: <http://www.caida.org/tools/utilities/netgeo>. Accessed 2011 February, 10.
39. Lind D, Marcus B (1995) *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press.