

From time series to networks in R with the ts2net package

Leonardo N. Ferreira

Center for Humans & Machines, Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany

Complex networks are one of the most prominent tools for modeling complex systems [1]. The network topology permits not just the analysis of the small parts (nodes) in the system but also their relationship (links) and the complex phenomena that could emerge from these links. Furthermore, this model allows network science and graph mining tools to extract information from complex systems [2]. Complex systems are commonly represented by a set of time series with interdependencies or similarities. This set can be modeled as a network where nodes represent time series, and links connect pairs of associated ones. Associations are usually measured by time series distance functions, carefully chosen by the user to capture the desired similarities. Thus, the network construction process consists of two steps. First, the distance between all pairs of time series is calculated and stored in a distance matrix. Then, this distance matrix is transformed into an adjacency matrix using strategies such as k -nearest neighbors, ϵ neighborhood, or complete weighted graph.

R is one of the most used programming languages in statistics, data science, machine learning, complex networks, and complex systems. Different packages provide distance functions and network analysis, but no single package provides the necessary tools to model time series as networks. This abstract presents `ts2net`, a package to construct networks from time series in R. This package provides tools for measuring linear, non-linear, and event-based associations between time series. Distance calculations run in parallel using multi-core programming to speed up the modeling process. `ts2net` provides methods to run the calculations in supercomputers and computer clusters using multiple jobs. The package is open-access and released under the MIT license. The source code is available on GitHub at github.com/lnferreira/ts2net and the package can be installed from the Comprehensive R Archive Network (CRAN) by running `install.packages("ts2net")` on R.



tances between all pairs of time series. The `ts_dist` function calculates all pairs of distances and returns a distance matrix D . It runs serial or in parallel using different cores. The `ts_dist_part` calculates distances in part of the time series set, which is particularly useful to run in parallel as jobs in supercomputers or computer clusters (HPC). The `ts_dist_part.file` function works similarly to `ts_dist_part`, but it reads the time series from serialized objects (RDS files) in a directory. This means that `ts_dist_part.file` requires much less memory and should be preferred when memory consumption is a concern, e.g., huge data sets or data sets with very long time series.

A distance function is required to construct the distance matrix D . The `ts2net` package provides the most used time series comparison functions: Pearson correlation (`tsdist_cor`), cross-correlation (`tsdist_ccf`), dynamic time warping DTW (`tsdist_dtw`), normalized mutual information (`tsdist_nmi`), variation of information (`tsdist_voi`), maximal information coefficient (`tsdist_mic`), events synchronization (`tsdist_es`), and van Rossum distance (`tsdist_vr`). Some distances provide statistical tests that can be considered during the network construction process to avoid spurious links. Other distance functions can be easily implemented or adapted to be used by the package.

The second step in the network construction process is the transformation of D into an adjacency matrix A . The `ts2net` package provides four ways of network construction: k -NN (`net_knn`), ϵ -NN (`net_enn`), or complete weighted networks (`net_weighted`) [2]. Fig. 1 illustrates the whole net construction process using a set of temperature time series from 27 cities in the US.

In summary, the `ts2net` package makes complex system modeling with networks much simpler. Together with other R packages, `ts2net` permits the analysis and information extraction using network science and graph mining tools.

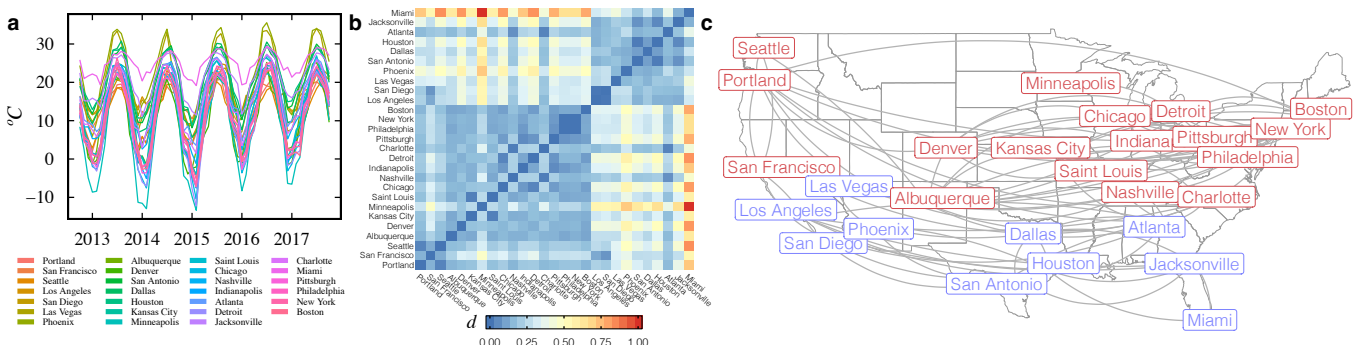


Fig. 1: Transforming time series into a network using `ts2net`. (a) Historical temperature of 27 cities in the US. (b) Distance matrix (normalized DTW) for the data set. (c) The ϵ -NN network using the 30% shortest distances. Node colors (blue and red) represent communities.

[1] Melanie Mitchell. Complex systems: Network thinking. *Artificial Intelligence*, 170(18):1194–1212, 2006.
[2] T.C. Silva and L. Zhao. *Machine Learning in Complex Networks*. Springer International Publishing, 2018.