7th Intercontinental Geoinformation Days

igd.mersin.edu.tr

The modeling and analysis of empirical systems with complex networks

Leyla Naghipour*[1], Khalil Valizadeh Kamran[1], Mohammad Taghi Aalami[2], Vahid Nourani[3]

[1] University of Tabriz, Department of Remote Sensing and GIS, Faculty of Planning and Environmental Sciences, 29 Bahman Blvd., Tabriz, East-Azerbaijan, Iran*
[2] University of Tabriz, Department of Water Resources Engineering, Faculty of Civil Engineering, 29 Bahman Blvd., Tabriz, East-Azerbaijan, Iran
[3] University of Tabriz, Center of Excellence in Hydroinformatics and Faculty of Civil Engineering, 29 Bahman Blvd., Tabriz, East-Azerbaijan, Iran

Keywords
Complex networks
Function
Topology
Collective behavior
Statistical approach

Abstract

Network construction is an acceptable approach for better understanding the behavior of complex system which can be used to reveal the pattern of collective dynamics for realizing physical interactions in the dynamical system. In this case, characterizing functional connectivity of complex networks for studying a broad class of natural and artificial systems from the measures of correlation and causality is of utmost importance to correctly unravel physical phenomena of the system. Many network reconstruction approaches are based on heuristically thresholding the correlation matrices resulting from pairwise correlation analysis according to experimental methods. Other approaches compare the observed correlations against null models in the statistical analyses, obtaining results which are statistically more robust. Different methods were used, including cross-correlation (CC), spectral coherence (SpeCoh), mutual information (MI), transfer entropy (TE), Spearman's rank correlation (SC) and convergent cross-mapping (CCM). The methods were applied to linear and nonlinear collective dynamics by autoregressive moving average (ARMA) and Logistic map (LOG) models, respectively. The dynamics of interconnected units was simulated from different complex topologies widely observed in empirical systems with well-known network models. The methods of MI and CCM were chosen after examining on the artificial cases consisting of desirable features of the real-world systems.

1. Introduction

Complex networks are widely used in many fields throughout the biological, social, information, engineering, and physical sciences to improve our understanding of collective dynamics and function of complex natural and artificial systems evolving in time (Albert and Barabási, 2002; Newman, 2003; Boccaletti et al. 2006).

The networks uncover the system's underlying interaction patterns where a detailed description of dynamics and structure may be impossible due to complex or chaotic behavior. The interconnection between the elements of a real-world complex system should be determined to consider the nature of ongoing interaction. Hence, the intrinsic connectivity between the components needs to be characterized before we can understand the system, and usually this is done by mapping physical (e.g., anatomical) connections onto a network. The effect of that connectivity is frequently investigated by employing the concept of complex networks.

The broad applicability of networks and their success in providing insights into the structure and function of both natural and human-made systems have thus produced considerable excitement across myriad scientific disciplines. For example, transportation networks with airline routes, road and rail networks (Sen et al. 2003; Gastner and Newman 2004; Li et al. 2015); disease containment strategies based on cellular networks data (Lima et al. 2015); characterizing interactions in online social networks (Omodei et al. 2015); understanding the spatio-temporal evolution of an epidemics and infer migration patterns (De Domenico et al. 2013 and Matamalas et al. 2016); organization and functioning of the human's brain (Reis et al. 2014; De

* **Corresponding Author**

*(Naghipour.l@tabrizu.ac.ir) ORCID ID 0000-0001-9000-3851
(valizadeh@tabrizu.ac.ir) ORCID ID 0000-0003-4648-842X
(mtaalami@tabrizu.ac.ir) ORCID ID 0000-0002-5845-9776
(nourani@tabrizu.ac.ir) ORCID ID 0000-0002-6931-7060

Cite this study

Naghipour L, Valizadeh Kamran, K., Aalami, M. T., & Nourani, V. (2023). The modeling and analysis of empirical systems with complex networks. Intercontinental Geoinformation Days (IGD), 7, 138-141, Peshawar, Pakistan

Domenico et al. 2016); good description of protein-genetic interactions (Jeong et al. 2001; Carmi et al. 2006; De Domenico et al. 2015); the modeling of complex climate system (Naghipour et al. 2021 and 2022); analysis of groundwater level (Naghipour et al. 2023).

The quantitative study of networks is fundamental for the characterization of complex systems. Importantly, several features arise in a diverse variety of networks. For example, many networks constructed from empirical data-sets exhibit heavy-tailed degree distributions, the small-world property, and/or modular structures; such structural features can have important implications for information diffusion, robustness against component failure, and many other considerations.

In this study, we will propose a new statistical approach trying to overcome the problems and improve the present understanding of the Earth's climate system and its predictability. This work can also be considered as a comprehensive study for the application of convergent cross-mapping (CCM) method and the evaluation of common methods. Our approach is based on a combination of dynamical systems techniques and statistical analysis. Finally, we show how surrogate models can partially reproduce the nonlinear dynamics.

2. Methodology

In this section, we briefly review the statistical methods most widely adopted for reconstructing the network structure from the observation of collective dynamics. While all the methods differ in the type of correlation, similarity or causality they estimate between two stochastic processes, say X(t) and Y(t), the subsequent steps are the same: i) calculate the same measure for any pair of signals corresponding to the activity in two different nodes; ii) perform a statistical comparison against a null model and, accordingly, iii) keep only the links which reject the null hypothesis of statistically uncorrelated dynamics (see Figure 1).

In the following, a brief description of our procedure is presented by considering two widely used linear correlation statistics, namely Cross-Correlation (CC) and Spectral Coherence (SpeCoh), two widely used information-theoretic correlation measures, namely Mutual Information (MI) and Transfer Entropy (TE), one non-parametric similarity measure, namely Spearman's rank Correlation (SC).

In the next section, we will briefly discuss one method based on the reconstruction of the underlying phase space, namely Convergent Cross-Mapping (CCM), that is widely used to infer causal relationships between two observed dynamics.

According to Figure 1, (a) We generate synthetic network models and correlated dynamics. (b) Methods for inferring correlation, similarity or causality relationships between nodes are used to reconstruct the network connectivity, by performing hypothesis testing for each pairwise interaction. The null hypothesis is that the observed time series are not statistically correlated. (c) The inferred network connectivity is compared against the ground truth and statistical descriptors such as accuracy, negative predictive value, specificity and balanced accuracy are estimated to validate the goodness of the reconstruction (see the study by Naghipour et al. (2021) and (2022) for detailed information).

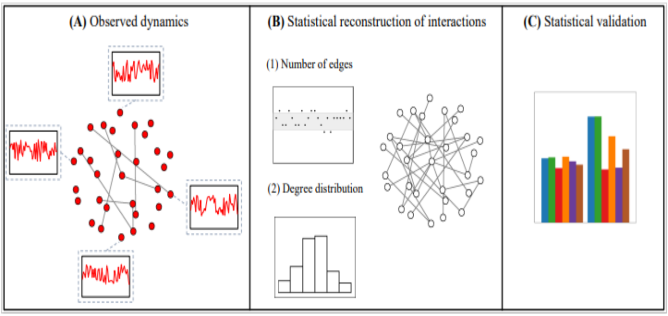

Figure 1. Schematic representation of the procedure used for reconstructing the network connectivity from the analysis of observed collective dynamics.

3. Results

In this work, we simulate 100 independent realizations of different network models -- namely with modular structure (Stochastic Block Model, SBM), scale-free (Barabási-Albert, BA), and small-world (Watts-Strogatz, WS) topology -- for increasing system's size (N=32,64, 128, 256) and time course length (M=64, 128, 256, 512, 1024). This setup allows one to understand the impact on connectivity reconstruction of network size, time series length, as well as the interplay between topology and dynamics. Reconstruction is performed according to the statistical approach previously described and for different correlation measures. The significance level chosen for each analysis in the following is 95%, compatible with a choice of 20 surrogates obtained by using randomly reshuffled series, the null model being the lack of correlation and causality between any pair of dynamics.

In Figure 2, a realization of the Barabási-Albert model, used as ground truth, and networks reconstructed from the observation of linear dynamics with shock propagation. The size of the system is N=32 and the length of temporal measurements is M=64. (B--G) Networks reconstructed with different methods. For each method, three cases are considered: forward dynamics only (1), backward dynamics only (2), and the multiplex network approach obtained from combing forward and backward dynamics (3).

We show in Figure 2 a single realization of a Barabási -Albert network, used as ground truth, simulate linear dynamics with shock propagation and show the resulting reconstructed networks. Once again, all reconstruction techniques described so far are used for forward-only collective dynamics (FOR) and dynamics with time reversal (INT). Methods like cross-correlation and Spearman's rank correlation lead to networks denser than the ground truth, explaining the excess of spurious connectivity observed in previous analyses. Conversely, convergent cross mapping and mutual information lead to networks as sparse as the ground truth, especially the latter. However, mutual information has the undesirable feature to infer edges where they are missing (i.e., high false negative rate), a problem only partially affecting convergent cross-mapping.

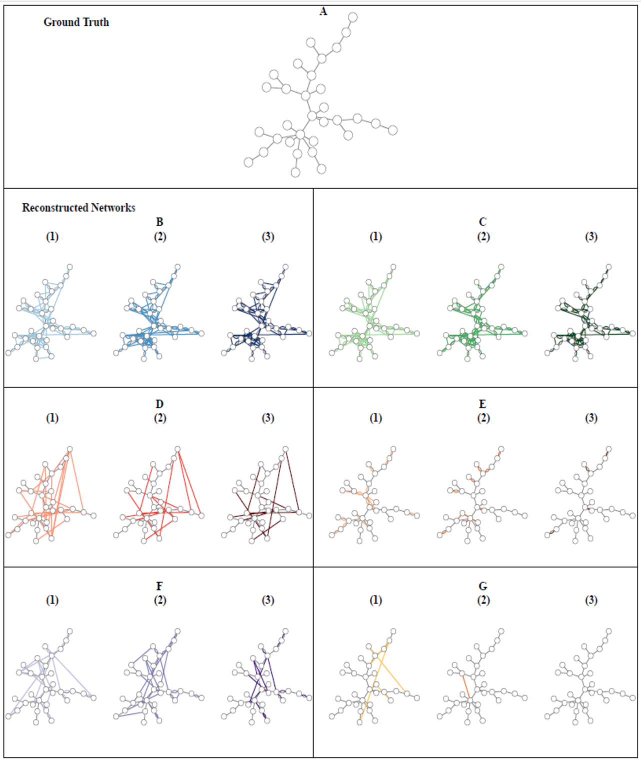

Figure 2. (A) A realization of the Barabási-Albert model with linear dynamics (B--G) Networks reconstructed with different methods, namely (B) cross-correlation, (C) Spearman's rank correlation, (D) spectral coherence, (E) convergent cross mapping, (F) transfer entropy and (G) mutual information.

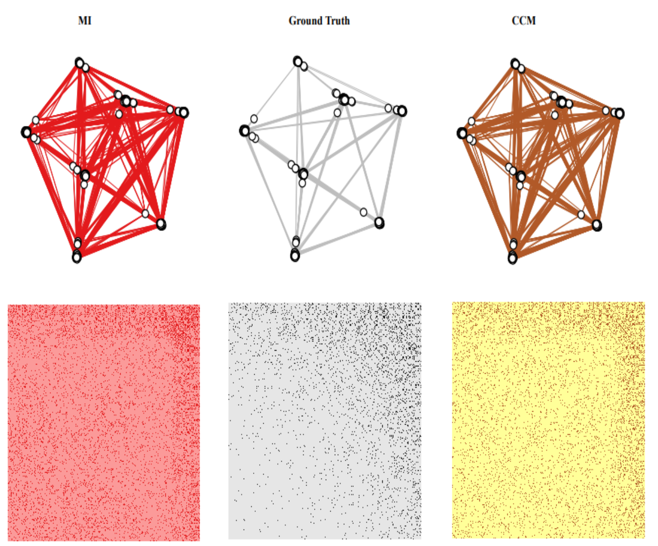

Figure 3. A representation of reconstructed networks and ground truth together with accompanying adjacency matrices.

In Figure 3, topology of the networks follows Lancichinetti-Fortunato-Radicchi model with the simulations from the nonlinearly coupled units with the Logistic map, and one of the network realizations is used as ground truth (gray edges). Inferred connectivity's are represented the reconstructed networks by the selected methods from the analysis presented in Figure 3, namely mutual information (MI) and convergent cross mapping (CCM) with the encoded colors as the previous figures. The dark colors indicate links on the corresponding adjacency matrices of the reconstructed networks. As another representative example, we show in Figure 3, a single realization of Lancichinetti-Fortunato-Radicchi network model, used as ground truth, simulate nonlinear collective dynamics and show the resulting reconstructed networks from the selected methods. Once again, selected reconstruction techniques as satisfied methods, are used, including mutual information and convergent cross mapping. Mutual information leads to networks denser than the ground truth, explaining the excess of spurious connectivity observed in previous analyses with more community detection (not shown in the figure). Convergent cross mapping also leads to networks denser than the ground truth, with less community detection which is desirable in comparison to the reconstructed network by mutual information. However, mutual information and convergent cross mapping have the unpleasant feature to infer edges where they are missing (i.e., high false negative rate), the problem of detecting true communities less affecting convergent cross mapping.

4. Conclusion

We demonstrate the principles of our approach with simple model examples, including BA, WS and SBM, to assess the strengths and weaknesses of the methods by varying system sizes and length of the time-series with performing a robust statistical analysis. The results show that the performance of CCM method is better to model time-series of the deterministic dynamical systems and MI models the stochastic series with high accuracy. In all statistical measures, the ARMA series has less error in compared with the LOG series. In fact, it has been shown that the analysis of collective dynamics might lead to very accurate (or inaccurate) reconstructions depending on the inference approach. However, convergent cross mapping appears to be the most robust, in terms of statistical reconstruction, to the unknown interplay between the structure of an interconnected system and the dynamics of its sites.

Acknowledgement

This study is supported by the research grant of the University of Tabriz (Number 718).

References

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. Reviews of modern physics, 74(1), 47.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. Physics reports, 424(4-5), 175-308.

Carmi, S., Levanon, E. Y., Havlin, S., & Eisenberg, E. (2006). Connectivity and expression in protein networks: proteins in a complex are uniformly expressed. Physical Review E, 73(3), 031909.

De Domenico, M., Lima, A., & Musolesi, M. (2013). Interdependence and predictability of human mobility and social interactions. Pervasive and Mobile Computing, 9(6), 798-807.

De Domenico, M., Nicosia, V., Arenas, A., & Latora, V. (2015). Structural reducibility of multilayer networks. Nature communications, 6(1), 6864.

De Domenico, M., Sasai, S., & Arenas, A. (2016). Mapping multiplex hubs in human functional brain networks. Frontiers in neuroscience, 10, 326.

Gastner, M. T., & Newman, M. E. (2004). Diffusion-based method for producing density-equalizing maps. Proceedings of the National Academy of Sciences, 101(20), 7499-7504.

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. Nature, 411(6833), 41-42.

Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E., & Havlin, S. (2015). Percolation transition in dynamical traffic network with evolving critical bottlenecks. Proceedings of the National Academy of Sciences, 112(3), 669-672.

Lima, A., De Domenico, M., Pejovic, V., & Musolesi, M. (2015). Disease containment strategies based on mobility and information dissemination. Scientific reports, 5(1), 10650.

Matamalas, J. T., De Domenico, M., & Arenas, A. (2016). Assessing reliable human mobility patterns from higher order memory in mobile communications. Journal of The Royal Society Interface, 13(121), 20160203.

Naghipour, L., Aalami, M. T., & Nourani, V. (2021). Reconstruction of network connectivity by the interplay between complex structure and dynamics to discover climate networks. Theoretical and Applied Climatology, 143, 969-987.

Naghipour, L., Aalami, M. T., & Nourani, V. (2022). Unravelling the backbone of climate networks from the analysis of collective dynamics and time reversal. International Journal of Climatology, 42(11), 5535-5553.

Naghipour, L., Aalami, M. T., & Nourani, V. (2023). Collective dynamics analysis based on the multiplex network method to unravel the backbone of fluctuations in groundwater level data. Computers & Geosciences, 172, 105310.

Newman, M. E. (2003). The structure and function of complex networks. SIAM review, 45(2), 167-256.

Omodei, E., De Domenico, M., & Arenas, A. (2015). Characterizing interactions in online social networks during exceptional events. Frontiers in Physics, 3, 59.

Reis, S. D., Hu, Y., Babino, A., Andrade Jr, J. S., Canals, S., Sigman, M., & Makse, H. A. (2014). Avoiding catastrophic failure in correlated networks of networks. Nature Physics, 10(10), 762-767.

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. Physical Review E, 67(3), 036106.