Supplemental Material
The distance backbone of directed networks

Felipe Xavier Costa[1,2,3,†], Rion Brattig Correia[1,3,†], and Luis M. Rocha[1,3,⋆]

[1] Systems Science and Industrial Engineering Department, Binghamton University (State University of New York), Binghamton NY 13902, USA
[2] Department of Physics, State University of New York at Albany, Albany NY 12222, USA
[3] Instituto Gulbenkian de Ciência, Oeiras 2780-156, Portugal
[†]equal contribution; [⋆]rocha@binghamton.edu

Abstract. In weighted graphs the shortest path between two nodes is often reached through an indirect path, out of all possible connections, leading to structural redundancies which play key roles in the dynamics and evolution of complex networks. We have previously developed a parameter-free, algebraically-principled methodology to uncover such redundancy and reveal the distance backbone of weighted graphs, which has been shown to be important in transmission dynamics, inference of important paths, and quantifying the robustness of networks. However, the method was developed for undirected graphs. Here we expand this methodology to weighted directed graphs and study the redundancy and robustness found in nine networks ranging from social, biomedical, and technical systems. We found that similarly to undirected graphs, directed graphs in general also contain a large amount of redundancy, as measured by the size of their (directed) distance backbone. Our methodology adds an additional tool to the principled sparsification of complex networks and the measure of their robustness.

Keywords: Directed networks, Weighted graphs, Network backbones, Sparsification. Shortest path, Redundancy

S1 Networks Descriptions

S1.1 Co-morbidity risks

Co-morbidity is the existence of more than one health condition in the same patient. This co-morbidity risk network connects medical and psychiatric conditions by their associated increased risk [1]. It was derived from longitudinal healthcare records from approximately 1 million United States military veterans, a population disproportionately impacted by psychiatric morbidity and psychological trauma. In this network, nodes corresponds to a physical or mental condition, while the edge weights measures the log odds ι_{ij} of having a primary health condition, x_i, increases the risk of having a secondary health condition, x_j. We consider, as a measurement of interaction strength, the probability of the increased risk given by $p_{ij} = \dfrac{1}{e^{-\iota_{ij}} + 1}$.

S1.2 Species-species interactions

Inter-species interactions (SSI) is an important resource for zoonosis studies in epidemiology as it describes host-pathogen interactions among species. In these networks species-species interactions indicate the possibility of one species being found in or on another species—a cargo-carrier relationship [6]. Many of such interactions are of the type: pathogen x_i was found in host x_j, however due to the nature of the underlying evidence it is not assumed all interactions to be of this type. Interactions can also be commensal (neither beneficial nor costly) or mutualistic (beneficial to both species), or vector-host. Additionally, an organism that is pathogenic to one host may be non-pathogenic in another. Species and interactions were extracted from metadata files containing a *host* tag and identified in 2,706,620 nucleotide sequence from the National Center for Biotechnology Information (NBCI). Here we use the total number of nucleotide sequence supporting such interactions as a relative similarity measure, $p_{i,j}$, that a specie x_i is found in or on specie x_j.

S1.3 Drug interactions

Drug-drug interactions (DDI) can cause obnoxious and often serious adverse reactions from the co-administration of multiple drugs, putting the elderly or co-morbid patients at increased risk of hospitalization. Here we study a known DDI network based on drug dispensation data in the city of Indianapolis (IN, USA) affecting 13% of the patient population of the major health-care provider in the city. Drug dispensation were collected from patient electronic health records for about 31% of the city population during a 2 year period starting in 2017. In this network nodes are drugs and edges represent the conditional likelihood of patients who administered a known DDI, p_{ij}, given they were prescribed and dispensed drug x_i for administration [2,5].

S1.4 Telephone calls

This network connects a population of more than 700 university students over a period of four weeks who participated in the Copenhagen Networks Study [4]. The original data included networks of physical proximity (estimated via Bluetooth), phone calls, text messages, and information about Facebook friendships. Data collection date was not disclosed for privacy reasons but noted it started on a Sunday during school term, and was released two years after data collection took place. Here, we study the high-resolution network of phone calls between participants where nodes corresponds to students and edges are weighted by the average call duration from a caller x_i to a callee x_j, as a measure of distance d_{ij}.

S1.5 Water pipes

A benchmark water distribution system, the EXNET network [3]. From this system, we construct a graph with nodes being pipe junctions and edge weights are the pipe length d_{ij} connecting the junctions.

References

1. Alexander-Bloch, A.F., Raznahan, A., Shinohara, R.T., Mathias, S.R., Bathulapalli, H., Bhalla, I.P., Goulet, J.L., Satterthwaite, T.D., Bassett, D.S., Glahn, D.C., Brandt, C.A.: The architecture of co-morbidity networks of physical and mental health conditions in military veterans. Proc. R. Soc. A 476, 20190790 (2020)
2. Correia, R.B., de Araújo Kohler, L.P., Mattos, M.M., Rocha, L.M.: City-wide analysis of electronic health records reveals gender and age biases in the administration of known drug-drug interactions. npj Digital Medicine 2 (2019)
3. Farmani, R., Savic, D.A., Walters, G.A.: Exnet benchmark problem for multi-objective optimization of large water systems. Modelling and control for participatory planning and managing water systems (2004)
4. Sapiezynski, P., Stopczynski, A., Lassen, D.D., Lehmann, S.: Interaction data from the copenhagen networks study. Sci. Data 6, 315 (2019)
5. Sánchez-Valle, J., Correia, R.B., Camacho-Artacho, M., Lepore, R., Mattos, M.M., Rocha, L.M., Valencia, A.: Large-scale multi-center comparison of electronic health records uncovers high prevalence and biases in the co-administration of drugs known to interact (2022), in preparation
6. Wardeh, M., Risley, C., McIntyre, M.K., Setzkorn, C., Baylis, M.: Database of host-pathogen and related species interactions, and their global distribution. Sci. Data 2, 150049 (2015)