LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-677610

# A Guide to Selecting a Network Similarity Method

S. Soundarajan, T. Eliassi-Rad, B. Gallagher

September 29, 2015

9th Annual Machine Learning Symposium
New York, NY, United States
March 13, 2015 through March 13, 2015

**Disclaimer**

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

**A Guide to Selecting a Network Similarity Method**

Sucheta Soundarajan,[1] Tina Eliassi-Rad,[1*] Brian Gallagher[2]
[1] Rutgers University, Piscataway, NJ
[2] Lawrence Livermore National Laboratory, Livermore, CA
[*] eliassi@cs.rutgers.edu

**Abstract**

The study of network data covers diverse domains from social sciences to biology to information technology. While these different networks share important features, the extent of these similarities is not clear. A network similarity method is useful for applications such as detecting when the structure of an online financial network has changed, indicating possible fraud; or for determining when an algorithm developed on one network may be applied to a different network (i.e., a form of transfer learning) [2]. One network similarity method may compare two networks based on simple network features such as edge density, while another may examine more complex (and computationally burdensome) patterns such as communities. So, assuming no knowledge of node-correspondences, how similar are two networks? In [1], we study a variety of network-similarity methods and address the following three questions. (i) How correlated are different network-similarity methods to each other? (ii) How can one automatically find groups of methods that behave comparably? (iii) How can one select a single consensus method from a group of network-similarity methods?

We consider twenty different methods applied to the task of network-similarity ranking, in which one is given a network *G* plus a set of other networks and must rank the other networks in order of their similarity to network *G*. Figure 1 depicts an overview of our approach. We categorized a network-similarity method based on two criteria. First, at what level of the network does it operate? Second, what type of comparison does it use? For the first criterion, we define three levels: micro, mezzo, and macro. As their names suggest, at the micro-level a method extracts features at the node- or neighborhood-level; at the mezzo-level it extracts features from communities; and at the macro-level it extracts features from the global/network level. For the second criterion, we have three types: vector-based, classifier-based, and matching-based. Our work yields several valuable results. First, we show that the various similarity methods, though seemingly different, produce highly correlated rankings in cross-sectional networks. This is not the case in longitudinal networks where the time granularity is chosen inappropriately, leading to *structurally immature* graphs. Second, we observe that some complex methods can be approximated by a much simpler method. For example, a method that compares random walks from two networks is well correlated with a method that simply measures density. Third, we describe how to select a single consensus from a set of rankings, and show that across all considered networks, two methods--namely, random walk with restarts and *NetSimile* [2]--are consistently closest to the consensus ranking.

**References**

[1] S. Soundarajan, T. Eliassi-Rad, and B. Gallagher: A guide to selecting a network similarity method. *SIAM SDM* 2014: 1037-1045.

[2] M. Berlingerio, D. Koutra, T. Eliassi-Rad, and C. Faloutsos: Network similarity via multiple social theories. *ACM / IEEE ASONAM* 2013: 1439-1440.

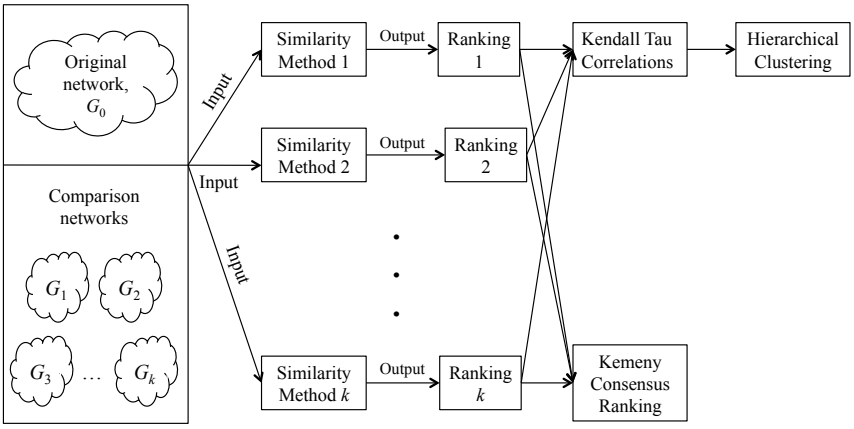


Figure 1: Our flowchart of how to compare various network similarity methods.

Prepared by LLNL under Contract DE-AC52-07NA27344