Accepted Manuscript

Connecting network science and information theory

Henrique F. de Arruda, Filipi N. Silva, Cesar H. Comin, Diego R. Amancio, Luciano da F. Costa

PII: S0378-4371(18)31343-8
DOI: https://doi.org/10.1016/j.physa.2018.10.005
Reference: PHYSA 20263

To appear in: *Physica A*

Received date : 13 July 2018
Revised date : 28 September 2018

Please cite this article as: H.F. de Arruda, et al., Connecting network science and information theory, *Physica A* (2018), https://doi.org/10.1016/j.physa.2018.10.005

Connecting Network Science and Information Theory

Henrique F. de Arruda[a,*], Filipi N. Silva[b,c], Cesar H. Comin[b], Diego R. Amancio[a,c], Luciano da F. Costa[b]

[a]*Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, SP, Brazil.*
[b]*São Carlos Institute of Physics, University of São Paulo, São Carlos, SP, Brazil*
[c]*School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA*

Abstract

A framework integrating information theory and network science is proposed. By incorporating and integrating concepts such as complexity, coding, topological projections and network dynamics, the proposed network-based framework paves the way not only to extending traditional information science, but also to modeling, characterizing and analyzing a broad class of real-world problems, from language communication to DNA coding. Basically, an original network is supposed to be transmitted, with or without compaction, through a sequence of symbols or time-series obtained by sampling its topology by some network dynamics, such as random walks. We show that the degree of compression is ultimately related to the ability to predict the frequency of symbols based on the topology of the original network and the adopted dynamics. The potential of the proposed approach is illustrated with respect to the efficiency of transmitting several types of topologies by using a variety of random walks. Several interesting results are obtained, including the behavior of the Barabási-Albert model oscillating between high and low performance depending on the considered dynamics, and the distinct performances obtained for two geographical models.

Keywords: Information theory, Complex networks, Compression, Data compression, Random walks, and Network dynamics.

1. Introduction

Much effort in science and technology has been focused on the study of information theory [1] and network science [2, 3], two seemingly independent realms. In information theory, probabilities are assigned to symbols and used to derive important results, such as minimum bandwidth and minimal sampling rates. On the other hand, in *network science* [2], focus is given to understanding the intricate topology of complex networks, and its interaction with various types of dynamics. Interestingly, these two different perspectives — broadly related to time series compaction and studies of topology/dynamics complexity — can be shown to ultimately be closely interrelated. For instance, information theory has been used to define causal relationships between

*Corresponding author
Email address: h.f.arruda@gmail.com (Henrique F. de Arruda)