Physica A 401 (2014) 118–129

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Directionality of real world networks as predicted by path length in directed and undirected graphs

Yonatan Rosen *, Yoram Louzoun **

Department of Mathematics and Gonda Brain Research Center, Bar-Ilan University, Ramat Gan, 52900, Israel

HIGHLIGHTS

- We propose a novel score to detect whether networks have an inherent order.
- We show that real world networks contain an inherent order.
- The same score helps ordering the nodes in such a network.
- When applied to neural networks, the classification into sensory and motor neurons emerges directly from this order.

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form 24 October 2013
Available online 21 January 2014

Keywords:
Directionality
Centrality
Directed networks
Real-world networks

ABSTRACT

Many real world networks either support ordered processes, or are actually representations of such processes. However, the same networks contain large strong connectivity components and long circles, which hide a possible inherent order, since each vertex can be reached from each vertex in a directed path. Thus, the presence of an inherent directionality in networks may be hidden. We here discuss a possible definition of such a directionality and propose a method to detect it.

Several common algorithms, such as the betweenness centrality or the degree, measure various aspects of centrality in networks. However, they do not address directly the issue of inherent directionality. The goal of the algorithm discussed here is the detection of global directionality in directed networks. Such an algorithm is essential to detangle complex networks into ordered process.

We show that indeed the vast majority of measured real world networks have a clear directionality. Moreover, this directionality can be used to classify vertices in these networks from sources to sinks. Such an algorithm can be highly useful in order to extract a meaning from large interaction networks assembled in many domains.

1. Introduction

Many natural and artificial networks have an inherent direction. In neural networks [1] information is first received in sensor neurons, processed in interim layers and then the network output may be carried out through motor neurons. In power grids [2], electricity flows from generating stations to switching stations; and in a similar fashion in other distribution networks such as oil or gas networks [3], substances are delivered from manufacturers to industrial and private consumers. In signal transduction pathways [4], a chain of cascading events is usually started by membrane bound receptor and then diffused through signaling molecules such as enzymes or hormones and ends in the bio-molecules involved in the outcome

* Corresponding author. Tel.: +972 35317165.
** Corresponding author. Tel.: +972 35317610.
E-mail addresses: RosenY77@gmail.com (Y. Rosen), louzouy@math.biu.ac.il (Y. Louzoun).

0378-4371/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2014.01.005

of the cellular process. When the direction is induced by the timing of events, this timing can actually be used to produce networks, as is the case for directed networks constructed from the time series of stocks returns using partial correlations [5,6], or Granger causality [7].

A direction in network can emerge even if no time is involved through a hierarchy, where nodes high in the hierarchy are ordered before those lower in the hierarchy. Typical examples would be directed linguistic networks, such as concepts in an encyclopedia where some terms are more general and others are more specialized [8], associations between words [9] or other semantic relations among concepts (e.g. "is-a" or "part-of") [10]. Non-knowledge representations physical networks that are time independent could be food webs [11–14], where edges represent causal predator–prey relations, or hierarchical anatomical networks, such as the brain connectome [15], where edges are inclusion relations from large scale areas to more subtle ones. Another source of directionality is based on sources and sinks of information. In such cases, the place where information is created "precedes" the place where it is lost (e.g. email networks, where people email to many other people, but some people are sources of information, while others are only using information. In this case the people using information email the people producing it, but the content is different between producers of information and consumers of it, and this process is not explicitly ordered by time precedence of the emails).

All these contradicting cases of directionality show that the general term of direction is ill defined. However, in graph theory, this concept can be translated to the basic question of whether the specific directions of edges affect generic measures of the network. In the absence of directionality, one would expect that changing randomly the direction of some edges in the network (from $A \rightarrow B$ to $B \rightarrow A$) would have no effect on global properties of the network. If there is a preferred direction to the network, we expect that randomly changing the direction of edges would ruin this order, and have some effect on global measures of the network. In the following sections, we limit ourselves to this minimal definition of directionality.

Note that this limited definition of directionality is global and is measured at the level of the entire network. Specifically, one can ask, whether the large amount of directed networks currently assembled have an inherent directionality.

The parallel question at the vertex level is: If indeed directionality exists in a network, is possible to sort the nodes along a virtual axis induced by this directionality? We here show that a simple measure can answer these two questions.

Directionality naturally induces hierarchy, but the opposite claim is not true. Hierarchy may be unrelated to directionality. For example, the degree distribution in an undirected network induces hierarchy where "important" vertices are more connected. However, such a hierarchy can always be defined, even in the absence of directionality. In other words, in a random directed network, a hierarchy can be defined, but it does not imply that randomly altering the direction of edges affect any global network property.

One could assume that position in a network could be defined based on some property of a vertex, such as its degree, or its betweenness centrality. However, these measures do not specifically measure directionality. Some other measures were proposed to estimate hierarchy, such as closeness centrality [16], defined as the mean of shortest path distance between a vertex and all other vertices reachable from it. In this measure, it is assumed that important vertices reach others through short distances. The PageRank algorithm [17] is a form of eigenvector centrality [18,19] where the score of each vertex depends on the PageRank score of the vertices pointing to it. The PageRank score vector is a stationary probability distribution expressing the chance of reaching a vertex by a random walk. The attraction basin hierarchy was proposed [8] as the ratio between the size of the incoming and outgoing edges at different distances. Finally, in a network drawing algorithm [20], the score vector is a unique minimizer of an energy that is supposed to order vertices in a hierarchy. The common feature of all these measures is that while they can partially uncover a present order, they will induce an order between vertices, even if it does not exist. Thus, these algorithms cannot test whether a given network indeed has an inherent directionality. On the other hand, a measure was proposed to evaluate the extent of directionality in the whole network level [21], defined as the fraction of edges in a network that are not included in cycles. However, it does not order the vertices and it involves high computational complexity regarding cases of substantial edge betweenness.

To clarify, what measure can be used to simply define direction in a network, the special case of DAGs, (directed acyclic graphs), will be examined. A DAG is a network that does not contain directed circles (i.e. directed paths from a vertex to itself). In such a network if u points to v, v cannot point to u, since otherwise these two paths would create a directed circle. In a DAG, there is clearly a coherent direction, since progressively vertices cannot point to previous ones. Such networks could be citation networks [22] or food webs [13]. In citation networks, the direction is derived from the order of the publication in time, while food webs are approximately acyclic since the food chain is typically layered. The directional hierarchy in DAGs, known as topological sorting [23], is also well-defined: It is a linear ordering of the vertices such that for every edge (u, v), u comes before v in the ordering. Indeed multiple successful algorithms have been proposed to deal with DAGs, such as topological sorting and Depth First Search [23,24]. However, as the vast majority of natural and artificial networks contain cycles, such algorithms cannot be applied successfully, and appropriate methods to detect directionality are necessary.

We here propose to expand the definition of directional hierarchy to non-DAG networks. In a DAG, this definition is of local nature, since the existence of an edge (u, v) forces that u comes before v. However, this is untrue in a non-DAG network: for a shortest path cycle of length n that contains an edge (u, v), u is located before v, but simultaneously v is located $n - 1$ edges before u. Therefore directional hierarchy in a non-DAG network can be defined such that, first, if there is a shortest path from u to v, but not the other way, the score of u is necessarily higher than v. Second, if u and v are unconnected (not in a cycle or one way relations), and u comes before more vertices than v in the terms of the first condition, the score of u again must be higher.