Physica A 392 (2013) 5511–5515

Contents lists available at ScienceDirect

Physica A

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

Lobby index as a network centrality measure☆

M.G. Campiteli [a], A.J. Holanda [b,*], L.D.H. Soares [c], P.R.C. Soles [c], O. Kinouchi [c,**]

[a] Instituto de Física de São Carlos, Universidade de São Paulo, Brazil
[b] Departamento de Computação e Matemática, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brazil
[c] Departamento de Física, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, Ribeirão Preto – SP, 14040-901, Brazil

HIGHLIGHTS

- We compared the lobby-index with degree, betweenness and Eigenvector centralities.
- We applied the indexes to linguistic (Moby Thesaurus) and biological (Yeast proteome) networks.
- The lobby-index is independent and does not correlate to betweenness centrality.
- The lobby-index outperforms the Eigenvector centrality in ranking important words from the Thesaurus.
- A plot of the lobby-index versus Eigenvector for the Yeast data detects a cluster of Yeast ribosome proteins.

ARTICLE INFO

Article history:
Received 26 April 2013
Received in revised form 27 June 2013
Available online 5 July 2013

Keywords:
Lobby index
Centrality
Degree
Betweenness
Eigenvector
Hirsch index

ABSTRACT

We study the lobby index (l-index for short) as a local node centrality measure for complex networks. The l-index is compared with degree (a local measure), betweenness and Eigenvector centralities (two global measures) in the case of a biological network (Yeast interaction protein–protein network) and a linguistic network (*Moby Thesaurus* II). In both networks, the l-index has a poor correlation with betweenness but correlates with degree and Eigenvector centralities. Although being local, the l-index carries more information about its neighbors than degree centrality. Also, it requires much less time to compute when compared with Eigenvector centrality. Results show that the l-index produces better results than degree and Eigenvector centrality for ranking purposes.

1. Introduction

The Hirsch index (h-index) has been thoroughly studied for scientometrics purposes. It has been applied to networks of individual researchers collaboration [1–5], research groups [6], journals [7,8] and countries [9] obtained from database of citations. In this context, the h-index is the largest integer h such that a node from a given network has at least h neighbors which have a degree of at least h [1].

Korn et al. [10] have proposed a general index to network node centrality based on the h-index. Korn et al. named it as the lobby index (l-index). Korn et al. argue that the proposed index contains a mix of properties of other well known centrality measures. However, they have studied it mainly in the context of artificial networks like the Barabási–Albert model [11].

* Corresponding author.
** Corresponding author. Tel.: +55 1636023779.
E-mail addresses: aholanda@usp.br (A.J. Holanda), osame@ffclrp.usp.br, okinouchi@gmail.com, osame@usp.br (O. Kinouchi).

0378-4371/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.06.065

Like l-index, degree D is a local centrality measure that is equal to the number of links of a given node. If the network is directed, the number of outlinks is the outdegree and the number of inlinks is the indegree. Unlike l-index, betweenness and Eigenvector are global centrality measures that take into account all nodes in the network. The betweenness B of a given node is proportional to the number of geodesic paths (minimal paths between node pairs in the network) that pass through it. It seems to be an important measure for networks where such minimal paths represent transport channels for information (internet, social networks), energy (power grids), materials (airports network) or diseases (social and sexual networks). Eigenvector centrality of a node is proportional to the sum of the centralities of the nodes to which it is connected, α is the largest eigenvalue of $A = a_{ij}$ and n the number of nodes [12]:

$$Ax = \alpha x, \qquad \alpha x_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, n. \tag{1}$$

In this paper, we compare the l-index with degree, betweenness and Eigenvector centralities applied to associative (non-transport) networks to obtain the correlation between these measures.

2. Methods

We calculate the l-index, degree D, betweenness B and Eigenvector E centralities for the nodes in linguistic and biological networks already considered by the physics community. We also plot the dispersion of D versus l, B versus l and E versus l, to verify the correlation between these measures.

We use the linguistic database *Moby Thesaurus* II [13] composed by 30,260 words, for which some network properties have been studied [14,15]. We choose the convention that an outlink goes from a root word to a synonym to construct the network. As an example, in the entry

`set`, `assign`, `assign to`, `assigned`, ...

the word "set" is the root and the link goes to its synonyms. We obtain the directed links "set"→ "assign", "set"→ "assign to" and "set"→ "assigned".

The raw thesaurus presents over 2.5 million links, but there are many words with only inlinks, that is, they are not root words. We worked with a filtered version containing about 1.7 million links where only root words constitute nodes. We choose the outlinks to calculate the centrality measures, and the minimal number of outlinks is 17 and the maximum is 1106.

The biological network is the yeast protein–protein network downloaded from the BioGRID repository [16] that is a curated repository for 5433 proteins and over 150,000 physical and genetic unambiguous interactions.

The BioGRID network is composed by gene products connected by a link [16]. The links include direct physical binding of two proteins, co-existence in a stable complex or genetic interaction as given by one or several experiments described in the literature. As an example, using the entries

```
YFL039C   YBR243C
YFL039C   YKL052C
```

extracted from BioGRID data set, two links are created: "YFL039C"–"YBR243C" and "YFL039C"–"YKL052C", and the network is undirected.

3. Results

3.1. Local measure: degree

In Fig. 1, we present dispersion plots of the l versus D for the networks studied. The l-index is correlated with D ($h \propto D$) in the low D regime ($D \leq 100$) in both networks. However, for higher D, one observes l proportional to $D^{0.4}$ for both networks. The origin of this anomalous exponent is not clear. Notwithstanding, although correlated, the two measures are not redundant. In the thesaurus case, the words with low frequency of use or that are non-polysemous present low l but high degree.

3.2. Global measures: betweenness and eigenvector

We now compare the l-index with two standard global centrality measures, betweenness and Eigenvector. First, in Fig. 2 we present the dispersion plots of l versus B. The l-index presents no strong correlation with B in both networks.

In Fig. 3, we give the dispersion plot for the l-index versus the Eigenvector centrality E for the thesaurus network. In the high E regime the maximal l values is bounded by $h \propto E^{0.4}$, as in the l versus D plot. We observe several nodes with high E but relatively low l (see Inset). Examining these nodes individually, we find that l seems to outperform E in the ranking task, since words with high l also have high E and are basic and important polysemous words. In contrast, terms with high