Physica A 387 (2008) 4959–4966

Contents lists available at ScienceDirect

Physica A

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

Centrality properties of directed module members in social networks

Péter Pollner[a], Gergely Palla[a], Dániel Ábel[b], András Vicsek[c], Illés J. Farkas[a], Imre Derényi[b], Tamás Vicsek[a,b,*]

[a] Statistical and Biological Physics Research Group of Hungarian Academy of Sciences, 1117 Budapest, Pázmány P. stny. 1/A, Hungary

[b] Department of Biological Physics, Eötvös University, 1117 Budapest, Pázmány P. stny. 1/A, Hungary

[c] Outline Foundation, 1023 Budapest, Frankel Leó u. 21-23. II/7, Hungary

ARTICLE INFO

Article history:
Received 9 April 2008
Available online 18 April 2008

PACS:
89.75.Hc
89.75.Fb
89.65.Ef
89.75.Da

Keywords:
Network modules
Overlapping modules
Directed networks
Clustering
Community finding

ABSTRACT

Several recent studies of complex networks have suggested algorithms for locating network communities, also called modules or clusters, which are mostly defined as groups of nodes with dense internal connections. Along with the rapid development of these clustering techniques, the ability of revealing overlaps between communities has become very important as well. An efficient search technique for locating overlapping modules is provided by the Clique Percolation Method (CPM) and its extension to directed graphs, the CPMd algorithm. Here we investigate the centrality properties of directed module members in social networks obtained from e-mail exchanges and from sociometric questionnaires. Our results indicate that nodes in the overlaps between modules play a central role in the studied systems. Furthermore, the two different types of networks show interesting differences in the relation between the centrality measures and the role of the nodes in the directed modules.

1. Introduction

A widespread approach to the analysis of complex social and economic phenomena is to assemble the participating individuals or objects and their interactions into a network (nodes and links) and to infer functional characteristics of the entire system from this static web of connections [1–5]. Over the past few years, several broadly studied *large-scale* properties of real-world webs have been uncovered, *e.g.* the broad (scale-free) distribution of node degree [6], overrepresented small subgraphs [7,8] and various signatures of hierarchical/modular organisation [9]. Also, many useful measures have been defined to quantify the importance of the *individual nodes* in the networks. If a vertex lies on many shortest paths running between other vertices, it plays a central role in information flows Ref. [10] and is responsible for the vulnerability of the system [11]. In a social network the actors capable of reaching the others with the lowest number of steps (being the closest to others on average) have the greatest influence [12,13].

In recent years there has been a rapidly growing interest in the study of the *intermediate-scale* network structures as well. These units, made up of vertices more densely connected to each other than to the rest of the network, are often referred to as communities, modules, clusters or cohesive groups [14–17]. In the various types of networks' these groups can represent a number of things *e.g.* , communities of people [14,18], functional units in molecular biology [19,20] and

* Corresponding author at: Statistical and Biological Physics Research Group of Hungarian Academy of Sciences, 1117 Budapest, Pázmány P. stny. 1/A, Hungary.

E-mail address: vicsek@angel.elte.hu (T. Vicsek).

URL: http://www.cfinder.org (T. Vicsek).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.04.025

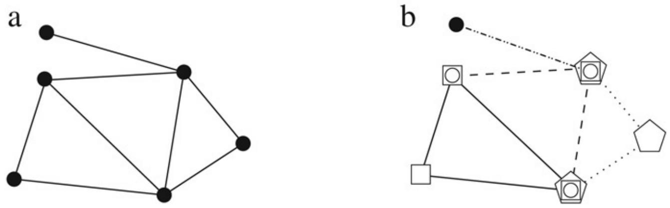

Fig. 1. Visualisation of the CPM method. A community is explored by walking over neighbouring building blocks. The blocks are triangles in this example, that are k-cliques with size $k = 3$. At figure (a) the graph is depicted without marking of blocks. At figure (b) the same graph is plotted with marking the nodes with different symbols for each different block. This graph contains three blocks, and one node, which is not a member of any blocks. The nodes of each blocks are marked by circles, rectangles and pentagons respectively. The unclassifiable node is marked by a filled ellipse. The block built from rectangles and the block of circles are neighbours since they differ only in one node. The block of pentagons and circles are neighbours as well. This graph contains only one community, since all blocks are reachable from each other by walking over neighbours. In this case the walks take at most two steps *e.g.* from block of rectangles to block of circles and than to block of pentagons.

a set of tightly coupled stocks or industrial sectors in economy [21]. Search techniques allowing overlaps between such modules are becoming very important [22–30]. Indeed, communities in real-world graphs are often inherently overlapping; each person in a social web belongs usually to several groups (family, colleagues, friends, etc.), and proteins in a protein interaction network may participate in multiple complexes. Modules, and also some small subgraphs, are appropriate for "coarse-graining" complex networks: each module/subgraph can be represented as a node and two such nodes can be linked, if the corresponding modules/subgraphs are connected (or overlap) [17,31,32].

In this paper we study the centrality properties of module members (nodes) in four social networks representing either e-mail connections between people or social links obtained from questionnaires. Both types of data sets can be represented by *directed networks*, therefore, the modules are located using a directed community finding method, the Directed Clique Percolation Method (CPMd) [33]. Our focus is on the correlations between general single-node properties (*e.g.* the closeness) and the role/importance of the members in the communities, (*e.g.* the number of communities they participate in).

2. Directed network modules

2.1. Directed clique percolation method (CPMd)

The CPMd is a natural extension of the Clique Percolation Method (CPM) [17,34]. In the CPM a community is built up from adjacent blocks of the same size k. These blocks correspond to k-cliques, that are subgraphs with the highest possible density: each pair selected from the k nodes of the subgraph is connected. Two blocks are considered adjacent if they overlap with each other as strongly as possible, *i.e.*, they share $k - 1$ nodes. A community is a set of blocks that can be reached from one to the other through a sequence of adjacent blocks [34]. Note, that any block belongs always to exactly one community, however, there can be nodes belonging to several communities. (e.g. if blocks overlap only in a single node.) The CPM community searching method is visualised in Fig. 1: the communities found by CPM contain only nodes that are densely connected. Nodes with only a few connections or nodes that do not participate in a densely connected subgraph are not classified into any community.

The CPM method is robust against removal or insertion of a single link. Due to the local nature of this approach, such perturbations can alter only the communities containing at least one of the end points of the link. (In contrast, for global methods like modularity optimisation the removal or insertion of a single link can result in the change of the overall community structure.)

Finally, we note that the CPM will find the same communities in a given subgraph irrespective to the fact whether the subgraph is linked to a larger network or not. Therefore, a heterogeneous network can be analysed by first dividing it into homogeneous parts, and applying the method to these subnetworks separately. Homogeneous parts of a network can be determined e.g. from external informations like geographical locations of the nodes etc.

The undirected approach detailed above can be made inherently directed by replacing the (undirected) building blocks with *directed k-cliques*. In this Clique *P*ercolation *M*ethod with *d*irected cliques (CPMd) the building blocks of a community are defined as complete subgraphs of size k in which an ordering can be made such that between any pair of nodes there is a directed link pointing from the node with the higher order towards the lower one [33]. See Fig. 2. In the absence of bidirectional links (also called double links) the above condition is equivalent to

- the absence of directed loops (closed directed paths) in the directed k-cliques.
- the directionality of all links within the k-clique, where any directed link points from a node with a higher order to a lower one.
- the topology, where the number of the nearest out-neighbours (neighbours reached along an out-link) is different for each node in the k-clique.