BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 2 2012, pages 290–291
doi:10.1093/bioinformatics/btr637

Systems biology

Advance Access publication December 6, 2011

Computing graphlet signatures of network nodes and motifs in Cytoscape with GraphletCounter

Christopher Whelan* and Kemal Sönmez

Biomedical Engineering, Oregon Health and Science University, 20000 NW Walker Road, Beaverton, OR 97006, USA

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Biological network analysis can be enhanced by examining the connections between nodes and the rest of the network. For this purpose we have developed GraphletCounter, an open-source software tool for computing graphlet degree signatures that can operate on its own or as a plug-in to the network analysis environment Cytoscape. A unique characteristic of GraphletCounter is its ability to compute the graphlet signatures of network motifs, which can be specified by files generated by the motif-finding tool mfinder. GraphletCounter displays graphlet signatures for visual inspection within Cytoscape, and can output graphlet data for integration with larger workflows.

Availability and implementation: GraphletCounter is implemented in Java. It can be downloaded from the Cytoscape plugin repository, and is also available at http://sonmezsysbio.org/software/graphletcounter.

Contact: whelanch@ohsu.edu

Received on January 8, 2011; revised on November 1, 2011; accepted on November 14, 2011

1 INTRODUCTION

Advances in systems biology, and in particular the identification and analysis of network motifs (Milo *et al.*, 2002), have provided many insights into the structure and function of complex biological networks. Motif analysis has allowed researchers to identify common, recurring groupings of nodes and connections, and study of the motifs identified has explained the function of motifs as components of a network; for example, a feed-forward loop (FFL) acts as a sign-sensitive delay element (Mangan *et al.*, 2003). However, it would also be helpful to be able to see how the motif components are connected to the rest of the network. If an FFL has a different pattern of connections to the rest of the network than would be expected by chance, study of these connections might provide further insight into the functional role of the motif and how it is integrated into the architecture of the network. Further, different types of biological networks may have different motif connection patterns, yielding information about their higher level architectures.

Pržulj (2007) recently introduced *graphlets* as a tool for examining local structures within large networks. Graphlets are small, induced subgraphs of a network graph. By counting the number of times that a node in a network participates in each possible type of graphlet composed of up to some fixed number of nodes, one can create a vector for that node describing its local connections to the rest of the network, called a *graphlet signature*. In addition to computing the number of graphlets that a node belongs to, one can count the number of times it appears in a topologically distinct position in each graphlet, known as an *automorphism orbit*. These counts can be combined across nodes into a measure that describes the entire network so that it can be compared to other networks or to null models (Pržulj *et al.*, 2004; Pržulj, 2007). Graphlet signatures can be used to predict characteristics of the biological entity represented by the node via clustering or statistical inference algorithms; for example, Milenković and Pržulj (2008) used them to predict the function of uncharacterized nodes in protein–protein interaction (PPI) networks. They have also been used to identify cancer genes (Milenković *et al.*, 2010). GraphCrunch (Milenković *et al.*, 2008) is a software package that uses graphlets to compare networks to random networks from various null models, and is very useful for network comparison.

We believe that graphlets have great potential as a tool for analyzing biological networks, and could be used to provide insight into the ways in which motifs are integrated into the higher level structures of networks. In addition, we think graphlet data could be useful in the context of exploratory network visualization environments. To this end, we have developed GraphletCounter, a tool for collecting graphlet signatures of nodes and motifs in biological networks. In addition to running as a stand-alone utility, GraphletCounter runs as a plugin for the Cytoscape (Shannon *et al.*, 2003) network visualization system, allowing the exploration of graphlet signature data in a visual environment. The software is open source and freely available.

2 METHODS AND IMPLEMENTATION

GraphletCounter computes the graphlet degrees of selected nodes in a network for all graphlets made up of five nodes or fewer. For each node, it outputs the number of times that node participates in one of the 72 distinct automorphism orbits defined by Pržulj (2007). In addition to raw counts, GraphletCounter also computes weighted counts that eliminate the redundancy between orbits, using the method described by Milenković and Pržulj (2008). Currently, we assume that the network is undirected, and treat edges as if they were undirected when operating on a directed network. The defined set of graphlets do not consider self-edges or multiple edges between two nodes, and therefore we ignore such edges. Extension of the graphlet counting methodology to more complicated networks would be an interesting direction for future enhancements and study.

2.1 Computing the graphlet signatures of motifs

In addition to the ability to count graphlet signatures for individual nodes in a network, GraphletCounter can also compute graphlet signatures for network motifs, or more generally for sets of nodes within the network. Let the set of

*To whom correspondence should be addressed.

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

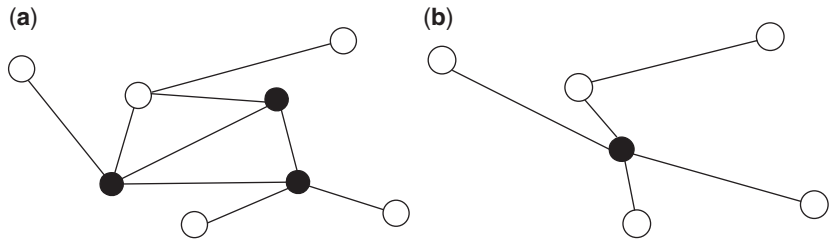

Fig. 1. Collapsing an example motif into a single node so that its graphlet signature can be computed. (**a**) A portion of a network containing a FFL motif (filled nodes). (**b**) The nodes composing the motif have been collapsed into a single node, allowing computation of a graphlet signature.

nodes be represented by $S = \{S_1, \ldots, S_n\}$. The algorithm to enable computation of the graphlet signature of S is then:

(1) Add a new node to the network, M, to represent S.

(2) For each edge between a node S_i in the subgraph and a node G not in the subgraph, create a new edge between M and G if one does not already exist.

(3) Remove each node S_i from the network, as well as all of its edges.

After this procedure, the new node M has the same neighbors in the rest of the network as the original subgraph had. Therefore, if we treat S as a network component, its connectivity to the rest of the network can be analyzed using graphlet signatures. A visual example of this process is shown in Figure 1. GraphletCounter currently accepts input files produced by the popular motif-finding utility mfinder (Kashtan *et al.*, 2004), and analyzes each of the motifs in the input file independently, restoring the network to its original state before processing the next motif. The code is designed to be extensible to input from other motif-finding tools, or tools designed to identify network clusters or communities.

2.2 User interfaces

GraphletCounter can be run as a plug-in to Cytoscape, and has been tested in version 2.8. The user can elect to compute graphlet counts for every node in the network, or only for nodes selected in the current Cytoscape session. Alternatively, the user can upload a list of motifs in mfinder format, and the plugin will compute graphlet counts for those motifs as described in Section 2.1. In the Cytoscape results panel as shown in Figure 2, the user can view the mean graphlet counts for the current selection, as well as the raw counts and a histogram of counts for each node. The counts can also be exported in a text format for further processing. All functions are also available as a command line program for batch processing. In this mode, the user provides a simple text representation of the network, as well as lists of nodes or motifs selected for processing.

Conflict of Interest: none declared.

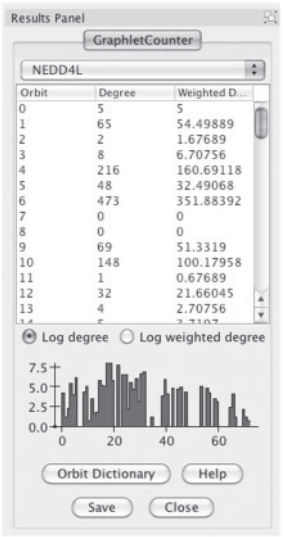

Fig. 2. A screenshot of the GraphletCounter results panel in Cytoscape.

REFERENCES

Kashtan,N. *et al*. (2004) Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, **20**, 1746–1758.

Mangan,S. *et al*. (2003) The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.*, **334**, 197–204.

Milenković,T. and Pržulj,N. (2008) Uncovering biological network function via graphlet degree signatures. *Cancer Infor.*, **6**, 257–273.

Milenković,T. *et al*. (2008) GraphCrunch: a tool for large network analyses. *BMC Bioinformatics*, **9**, 70.

Milenković, T. *et al*. (2010) Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related functional genomics data. *J. R Soc. Interface*, **7**, 423–437.

Milo,R. *et al*. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.

Pržulj,N. *et al*. (2004) Modeling interactome: scale-free or geometric? *Bioinformatics*, **20**, 3508–3515.

Pržulj,N. (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **23**, e177–e183.

Shannon,P. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016