Neurocomputing 206 (2016) 19–27

Contents lists available at ScienceDirect

Neurocomputing

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

Discovery of functional module alignment

Jiang Xie [a,*], Chaojuan Xiang [a], Junfu Xu [a], Wu Zhang [a], Jiao Wang [b,*]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[b] School of Life Sciences, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form
19 April 2016
Accepted 29 April 2016
Available online 1 June 2016

Keywords:
Functional module alignment
Visualization
Protein interaction networks alignment

ABSTRACT

The alignment of protein interaction networks (PINs) could help reveal similar subnetworks that may play important roles in biology. However, it is challenging to visualize the network alignment, especially for those large-scale PINs. In this work, we present a tool, namely FMA-finder, which can visualize large-scale network alignments between different species. Moreover, instead of focusing on the protein pairs to detect functional homologies, functional module alignment (FMA) is proposed in this study. FMAs are biologically meaningful because they are pairs of subnetworks sharing similar functions. The FMA-finder tool provides both analysis and visualization of FMAs. Experiments on the alignment between *Homo sapiens* and *Saccharomyces cerevisiae* protein interaction networks demonstrate that our FMA-finder can visualize and analyze large-scale network alignments.

1. Introduction

Protein interaction networks (PINs) are fundamental to understand the biology process [1]. Currently, many public databases have been constructed to store the experimentally validated and theoretically predicted PPIs of many model species, including FPPI [2], the Database of Interacting Proteins (DIP) [3], Molecular INTeraction database (MINT) [4], protein InterAction database (IntAct) [5], Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [6], and so on. Aligning networks from different species is an efficient way to identify previously undiscovered homologies between proteins of different species and reveal functionally similar subnetworks. And the ultimate goal of network alignment is to transfer the knowledge of protein function from one species to another [7,8]. In the past few years, a number of different aligners have been published, including local alignment and global alignment. Global alignment attempts to find the best mapping between all nodes of two networks, while local alignment is categorized as a one-to-many mapping and finds small subnetworks corresponding to the pathways or protein complexes conserved in PINs of different species [9]. Despite some global pairwise alignment algorithms, such as IsoRank [10], PISwap [11], GRAAL [12], MI-GRAAL [13], GEDEVO [14], NETAL [9] and HGA [15] have been proposed, there are few tools to visualize these network alignments, whereas visualization of network alignments is more sensible compared to statistical evidences.

Since PINs usually contain thousands of nodes and links, the amount of data to be visualized often exceeds the display capability of a computer screen. An interactive visualization is a powerful choice to interpret data. In recent years, some visualization tools were widely applied to analyze networks [16], such as Cytoscape, Pajek, Proviz and BioLayout Express3D. However, fewer tools, e.g. BNMatch [17] and NetCompare [18], focus on analyzing network alignment, and matched proteins are shown with corresponding positions and same colors. It is difficult for BNMatch to handle large scale PINs, and NetCompare only focuses on the topological similarity. Besides, there are some open problems regarding PIN visualization, such as the high number of nodes and connections, and the heterogeneity of nodes (proteins) and edges (interactions). Furthermore, the annotation of proteins and their interactions with biological information makes it more difficult for the visualization [19]. Efforts have been made to solve these problems, and many data reduction techniques have been developed, such as sampling, filtering, clustering, principal components analysis and multidimensional scaling [20].

Clustering, as one of the most critical steps for analyzing large scale networks, has been adopted to extract related structure data for visualization. Recently, many clustering approaches have been proposed to predict functional modules based on PINs, such as hierarchical multi-label classification (HMC) [21], HC-PIN [22], PROCOMOSS [23], CSO [24] and from-function-to-interaction paradigm [25]. Therefore, clustering for both visualization and detecting function modules in PINs are adopted in this work. A functional module [26] is defined as a group of genes or their products that are related by one or more genetic or molecular interactions, e.g. co-regulation, co-expression or membership of a

* Corresponding authors.
E-mail addresses: jiangx@shu.edu.cn (J. Xie), jo717@shu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.neucom.2016.04.055
0925-2312/

protein complex [27]. In this work, functional module alignments (FMAs) which are pairs of aligned modules sharing common functions are proposed.

With respect to the facts mentioned above, we present a tool, namely FMA-finder that takes advantages of clustering to dig into network alignment. Firstly, it provides an interface to explore network alignment visually. Furthermore, some approaches are developed to manipulate the PINs, and some pairs of FMA can be discovered with FMA-finder. The more proteins matched in one pair of FMA, the more convincible the matched protein pairs in FMA should be. Despite of some well-established network alignment quality measures, such as node correctness (NC), edge correctness (EC), induced conserved structure (ICS), symmetric substructure score (S3), the size of the largest connected common subgraph (LCCS) and GO consistence (GOC) [28], FMA is biologically meaningful providing an alternative way to evaluate different aligners.

2. Related work

Fig. 1 shows the schematic illustration of discovering modules with network alignment. The details are addressed as below.

2.1. Finding potential mapping of proteins between two PINs

Given two graphs $G_1=(V_1, E_1)$ and $G_2=(V_2, E_2)$ with nodes denoting proteins while edges their interactions, without loss of generality, we assumed that $|V_1| < |V_2|$. The problem of pairwise alignment is to find a one-to-one mapping function f: $V_1 \rightarrow V_2$, which maps each node in V_1 to those from V_2 that it best matches. Nodes connected by dotted lines shown in Fig. 2 are the matched nodes, and this alignment provides the best mapping of the two networks.

Here, we applied HGA developed by Jiang's group [15] as an example. This is a global network alignment algorithm with high SS, EC and HGp scores and can output alignment rapidly based on GPU programs [29].

2.2. Clustering protein network into hierarchical structure

Clustering, as one of the most critical steps for visualization of large scale network, is widely adopted to extract network modules from PINs. Here, a multi-level hierarchical clustering method named Louvain method [30] is adopted to transform PINs into multidimensional modules, which is different from other traditional clustering methods that divide networks into only one level, such as k-means, MCL, spectral and RNSC [31]. It has been found that this clustering algorithm outperforms all others, including algorithms proposed by Radicchi et al. [32] and Aslam et al. [33], thus providing the best option for visualization purposes with less edge crossings [34]. As shown in Fig. 3, this unsupervised algorithm first divides networks into clusters based on the local optimization of Newman–Girvan modularity in the neighborhood of each vertex, and further divides clusters into sub-clusters. In this way, multi-level clusters revealing the hierarchical structures can be obtained and are useful for multi-dimensional visualization. And the modular and hierarchical modular network structures have some advantages, including better robustness, adaptivity and evolvability of network function [35].

At the same time, by integrating clustering with network alignment, we can accomplish the following goals: (i) to find aligned modules which help to understand functional modules across species and (ii) to confirm the more accurately aligned protein pairs based on aligned modules.

2.3. Visualization of networks on the web

Document-driven documents (D3), a new web-based library, has recently become a very popular toolkit to construct interactive visualizations on the web [20]. Comparing with Cytoscape.js [36], it provides more meaningful layouts and supports direct manipulation of document elements (namely, webpage elements) by binding data to document elements using HTML, SVG and CSS. One

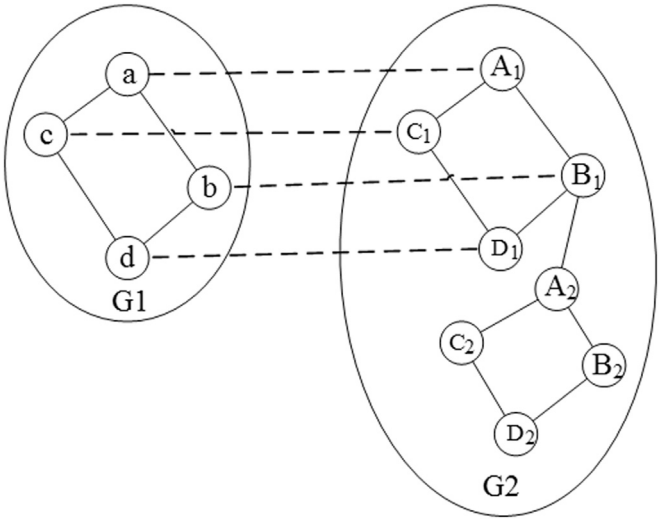

Fig. 2. Example of PIN alignment [29].

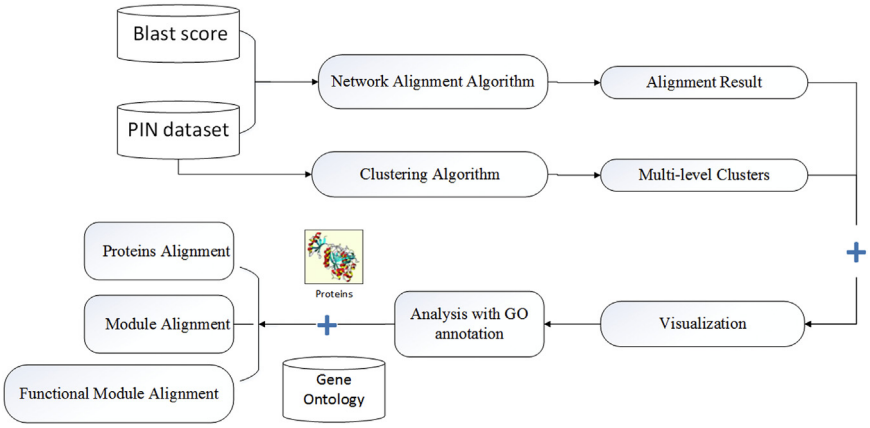

Fig. 1. The schematic illustration of discovering functional module alignments via aligning and clustering of PINs.