

Chapter 2

Problems and Techniques

Abstract When biological networks are considered, the extraction of interesting knowledge often involves subgraphs isomorphism check that is known to be NP-complete. For this reason, many approaches try to simplify the problem under consideration by considering structures simpler than graphs, such as trees or paths. Furthermore, the number of existing approximate techniques is notably greater than the number of exact methods. In this chapter, we provide an overview of three important problems defined on biological networks: network alignment, network clustering, and motifs extraction from biological networks. For each of these problems, we also describe some of the most important techniques proposed to approach them.

Keywords Biological network analysis • Graph alignment • Protein–protein interaction network clustering • Community search • Graph motif extraction • Global and local alignment

2.1 Network Alignment

Let N_1 and N_2 be two input networks. The alignment problem consists of finding a set of conserved edges across N_1 and N_2 , leading to a (non-necessarily connected) conserved subgraph between them. In this case, the problem is also known as *pairwise alignment*. *Multiple alignments* is an extension of pairwise alignment such that a set of networks N_1, \dots, N_n is considered in input, and it is usually computationally more difficult to perform. In the following we refer to pairwise network alignment, and all the notions we will introduce can extend to multiple alignment.

The problem of biological network alignment can be distinguished in *global alignment* and *local alignment*. Global alignment aims at finding a unique (possibly, the best one) overall alignment between N_1 and N_2 , in such a way that a one-to-one correspondence is found between nodes in N_1 and nodes in N_2 . The result is made of a set of pairs of non-overlapping subgraphs of N_1 and N_2 . Local alignment aims instead at finding multiple, unrelated regions of isomorphism among the input networks, each region implying a mapping independently of the others. Therefore,

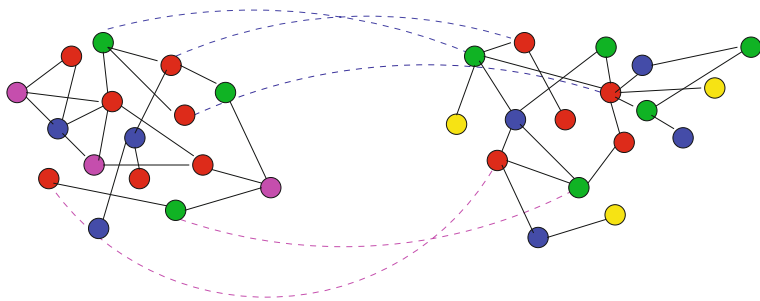


Fig. 2.1 Global alignment

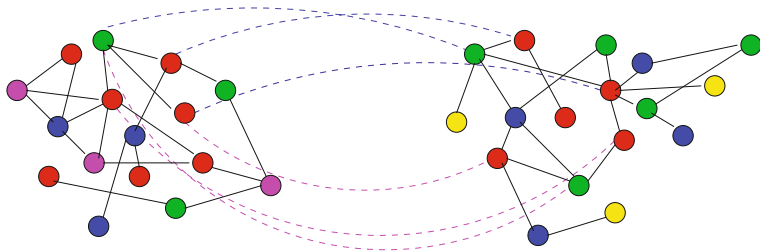


Fig. 2.2 Local alignment

the found correspondences may involve overlapping subgraphs. Figures 2.1 and 2.2 illustrate global alignment and local alignment, respectively.

Network alignment can be approached also if N_1 and N_2 are of different types, leading to a kind of *heterogeneous* alignment. Usually, in this case, the two input networks are merged and statistical approaches are then applied to extract the most significant subgraphs from the integrated network.

2.1.1 Techniques

Network alignment involves the problem of subgraph isomorphism checking, that is known to be NP-complete [27]. Therefore, the proposed techniques are often based on approximate and heuristic algorithms.

2.1.1.1 Global Alignment

Singh et al. [65] present *IsoRank*, an algorithm for pairwise global alignment of PPI networks working in two stages: it first associates a score with each possible match between nodes of the two networks and, then, constructs the mapping for the global network alignment by extracting mutually consistent matches according to a bipartite

graph weighted matching performed on the two entire networks. *IsoRank* has been extended in [64] to perform multiple alignments by approximate multipartite graph weighted matching. In [39] the *IsoRankN* (IsoRank-Nibble) tool is proposed, that is, a global multiple-network alignment tool based on spectral clustering on the induced graph of pairwise alignment scores. In [34] a graph-based maximum structural matching formulation for pairwise global network alignment is introduced, combining a Lagrangian relaxation approach with a branch-and-bound method. MI-GRAAL [36] can integrate any number and type of similarity measures between network nodes (e.g., sequence similarity, functional similarity, etc.) and find a combination of similarity measures yielding the largest contiguous (i.e., connected) alignments. In [62] a scalable algorithm for multiple alignment is presented based on clustering methods and graph matching techniques to detect conserved interactions while simultaneously attempting to maximize the sequence similarity of nodes involved in the alignment. Finally, in [44] an evolutionary-based global alignment algorithm is proposed, while in [45] a greedy method is used, based on an alignment scoring matrix derived from both biological and topological information of the input networks to find the best global network alignment.

2.1.1.2 Local Alignment

Kelley et al. [31] proposed *PathBLAST*, that is, a procedure for pairwise alignment combining interaction topology and protein sequence similarity. They search for high scoring pathway alignments involving two paths, one for each network, in which proteins of the first path are paired with putative homologs occurring in the same order in the second path. *PathBLAST* is extended in [59] for multiple alignments, based on the generation of a network alignment graph where each node consists of a group of sequence-similar proteins, one for each species, and each link between a pair of nodes represents a conserved protein interaction between the corresponding protein groups. *PathBLAST* has also been used in [8] to resolve ambiguous functional orthology relationships in PPI networks. In [35], a technique for pairwise alignment is proposed based on duplication/divergence models and on efficient heuristics to solve a graph optimization problem. *Bi-GRAPPIN* [23] is based on maximum weight matching of bipartite graphs, resulting from comparing the adjacent nodes of pairs of proteins occurring in the input networks. The idea is that proteins belonging to different networks should be matched looking not only at their own sequence similarity, but also at the similarity of proteins they significantly interact with. In [24] an algorithm for multiple alignments, named *Graemlin*, is presented. *Graemlin* aligns an arbitrary number of networks to individuate conserved functional modules, greedily assigning the aligned proteins to non-overlapping homology classes and progressively aligning multiple input networks. The algorithm also allows searching for different conserved topologies defined by the user. In [14] the algorithm *C3Part-M*, based on a non-heuristic approach exploiting a correspondence multigraph formalism to extract connected components conserved in multiple networks, is presented and compared

with *NetworkBlast-M* [30], another technique recently proposed based on a novel representation of multiple networks that is only linear in their size.

Finally, *AlignNemo* is proposed in [13] that builds a weighted alignment graph from the input networks, extracts all connected subgraphs of a given size from the alignment graph, and uses them as seeds for the alignment solution, by expanding each seed in an iterative fashion.

2.1.1.3 Other Approaches

ABiNet [19, 20] is an algorithm performing *asymmetric alignment*. In particular, given two input networks, the one associated to the best-characterized organism (called *Master*) is exploited as a fingerprint to guide the alignment process to the second input network (called *Slave*), so that generated results preferably retain the structural characteristics of the Master network. Technically, this is obtained by generating from the Master a finite automaton, called *alignment model*, which is then fed with (linearization of) the Slave for the purpose of extracting, via the Viterbi algorithm, matching subgraphs. *ABiNet* performs both querying and global alignment.

Finally, the approach [67] has been proposed in order to align heterogeneous networks, for example, PPI and disease networks.

2.1.2 Querying

Network querying consists on analyzing an input network, called *target network*, searching for the occurrences of a *query network* of interest. Such a problem “is aimed at transferring biological knowledge within and across species” [60], since the found subnetworks may correspond to cellular components involved in the same biological processes or performing similar functions than the components in the query.

2.1.2.1 Techniques

Network querying approaches may be divided into two main categories: those ones searching for efficient solutions under particular conditions, e.g., the query is not a general graph but it is a path or a tree, and other approaches where the query is a specific small graph in input, often representing a functional module of another well-characterized organism. *MetaPathwayHunter* [50] is an algorithm to query metabolic networks by multi-source trees that are directed acyclic graphs whose corresponding undirected graphs are trees where nodes may present both incoming and outgoing edges. In [16, 63] *QPath* and *QNet* are presented, respectively. *QPath* queries a PPI network by a query pathway consisting of a linear chain of interacting proteins belonging to another organism. The algorithm works similarly to sequence align-

ment, by aligning the query pathway to putative pathways in the target network, so that proteins in analogous positions have similar sequences. *QNet* is an extension of *QPath* in which the queries are trees or graphs with limited treewidth. In [68] the two problems of path matching and graph matching are considered. An exact algorithm called *SAGA* is presented to search for subgraphs of arbitrary structure in a large graph, grouping related vertices in the target network for each vertex in the query. *NetMatch* [21] is a Cytoscape plugin allowing for approximated queries, that is, graphs where some nodes are specified and others are wildcards (which can match an unspecified number of elements). *NetMatch* captures the topological similarity between the query and target graphs, without taking into account any information about node similarities. In [22] a technique is proposed based on maximum weight matching of bipartite graphs. *Torque* [10] is an algorithm based on dynamic programming and integer linear programming to search for a matching set of proteins that are sequence-similar to the query proteins, by relaxing the topology constraints of the query. Finally, note that, sometimes, methods for local alignment can be also successfully exploited to perform network querying, e.g., [20, 31, 35].

2.2 Network Clustering

The analysis of protein–protein interaction networks may result in the detection of protein complexes helping in understanding the mechanisms regulating cell life, in describing the evolutionary orthology signal (e.g., [29]), in predicting the biological functions of uncharacterized proteins, and, more importantly, for therapeutic purposes. The problem of detecting protein complexes using PPI networks can be computationally addressed by using clustering techniques aiming at grouping together proteins which share a large number of interactions. Possible uncharacterized proteins in a cluster may be assigned to the biological function recognized for that module, and groups of proteins performing the same tasks can be singled out this way.

PPI networks have various characteristics which have to be taken into account when developing clustering algorithms for detecting functional complexes.

2.2.1 Techniques

MCODE (*Molecular COMplex DETection*) [7] relies on a node weighting procedure by local neighborhood density and outward traversal from a locally dense seed protein, in order to isolate the dense regions according to given input parameters. The algorithm allows fine-tuning of clusters of interest without considering the rest of the network and allows examination of cluster interconnectivity, which is relevant for protein networks. The algorithm may use a “fluff” option which increases the

size of the modules to find and allows for overlapping among the output complexes, since the nodes added to a cluster are not marked as already used.

CFINDER [1] is based on the clique percolation concept (see [15, 46]). The idea behind this method is that a cluster can be interpreted as the union of small fully connected subgraphs that share nodes, where a parameter is used to specify the minimum number of shared nodes. CFinder extracts all the maximal complete subgraphs, i.e., the maximal cliques in the input PPI network. Then a clique–clique overlap matrix is built such that each entry contains the number of common nodes between the two corresponding cliques, and each diagonal entry is the clique size k . The k -cliques communities can be found by deleting every entry off the diagonal having a value less than $k - 1$, and every diagonal entry less than k . The remaining separate components will be the k -cliques communities. CFinder allows overlap between communities.

The greedy local expansion methods RANCoC [55], MF- PINCoC [52], and PINCoC [53] expand a single protein randomly selected by adding/removing connected proteins that best contribute to improve a given quality function based on the concept of co-clustering [40]. In order to escape poor local maxima, with a given probability, the protein causing the minimal decrease of the quality function is removed in MF- PINCoC [52] and PINCoC [53]. Instead RANCoC removes, with a fixed probability, a protein at random, even if the value of the quality function diminishes. This strategy is more efficient in terms of computation than that applied in the methods [52, 53], and it is more efficacious in avoiding entrapments in local optimal solutions. All the three algorithms work until either a preset of maximum number of iterations is reached, or the solution cannot further be improved. Both MF- PINCoC and RANCoC [55] allow overlapping clusters.

A typical instance of cost-based local search is RNSC (*Restricted Neighborhood Search Clustering*) [33], which explores the solution space of all the possible clusterings in order to minimize a cost function reflecting the number of inter-cluster and intra-cluster edges. The algorithm begins with a random clustering, and attempts to find a clustering with the best cost by repeatedly moving one node from a cluster to another one. A list of tabular moves is used to forbid cycling back to previously examined solutions. In order to output clusters likely to correspond to true protein complexes, thresholds for minimum cluster size, minimum density, and functional homogeneity must be set. Only clusters satisfying these criteria are given as the final result. This obviously implies that many proteins are not assigned to any cluster.

Several community discovery algorithms have been proposed based on the optimization of a modularity-based function (see e.g., [25]). Modularity measures the fraction of edges falling within communities, subtracted by what would be expected if the edges were randomly placed. In particular, QCUT [58] is an efficient heuristic algorithm applied to detect protein complexes. QCUT optimizes modularity by combining spectral graph partitioning and local search. By optimizing modularity, communities that are smaller than a certain scale or have relatively high inter-community density may be merged into a single cluster. In order to overcome this drawback, the authors introduce an algorithm that recursively applies QCUT to divide a community into subcommunities. In order to avoid over-partitioning, a statistical test is applied to determine whether a community indeed contains intrinsic subcommunity.

One of the first methods based on flow simulation for detecting protein complexes in a PPI network is the *Markov Clustering algorithm MCL* [17, 66]. MCL is based on the concept of a random walk on a graph to retrieve cluster structure and uses algebraic operations on a distance matrix associated with the graph. In a random walk the direction to be followed at each node is given by chance. MCL simulates many random walks (or flows) within a graph by strengthening flow where it is strong, and weakening it where it is weak. By repeating this process, a number of regions with strong internal flow (the clusters), separated by boundary with no flow, will appear. The flow is simulated by algebraic operations on a stochastic Markov matrix, such as flow expansion and an inflation operator which raises each entry of the matrix to a given power, and then rescales the matrix so that the column sum equals 1. By repeating a number of times squaring, inflating, and scaling, the matrix tends to an equilibrium state that shows the cluster structure. The inflation parameter influences the number of clusters.

In [51, 54] genetic algorithms have been applied to PPI networks, referred as *GA-PPI*, by performing an extensive experimental evaluation aiming at exploring the capability of genetic algorithms to find clusters in PPI networks, when different topological-based fitness functions are employed. The adopted representation of individuals is the graph-based adjacency representation, originally proposed in [49], where an individual of the population consists of n genes, each corresponding to a node of the graph modeling the PPI network. A value j assigned to the i th gene is interpreted as a link between the proteins i and j , and implies that i and j belong to the same cluster. In particular, in [51] the fitness functions of *conductance*, *expansion*, *cut ratio*, *normalized cut*, reported from [38], are employed, while in [54], the cost functions of the *RNSC* algorithm [33] have been used.

2.3 Network Motif Extraction

The concept of *motif* has been exploited in different applications of computational biology [3, 47]. Depending on the context, *what is a motif* may assume sensibly different meanings. In general, motifs are always associated to repetitive objects. For example, a repeated substring can be considered a motif when its frequency is greater than a fixed threshold, or instead when it is much different than expected [5].

Also in the context of biological networks, a motif can be defined according to its *frequency* or to its *statistical significance* [12]. In the first case, a motif is a subgraph that appears more than a threshold number of times in an input network; in the second case, a motif is a subgraph that appears more often than expected by chance. In particular, to measure the statistical significance of the motifs, many works compare the number of appearances of the motifs in the biological network with the number of appearances in a number of randomized networks [18], by exploiting suitable statistical indices such as *p value* and *z score* [43].

Despite the similarity with sequences, that is evident from the two definitions above, network motifs present important differences w.r.t. string motifs, the main

important of which concerns the computational complexity of the problem of motif extraction, that is, polynomial for strings and exponential (in the size of the input) for networks.

Given a biological network N , a *motif* can be defined according to its *frequency* or to its *statistical significance* [12]. In the first case, a motif is a subgraph appearing more than a threshold number of times in N ; in the second case, it is a subgraph occurring more often than expected by chance. In particular, to measure the statistical significance of a motif, many works compare its number of occurrences with those detected in a number of randomized networks [18], by exploiting suitable statistical indices such as *p value* and *z score* [43].

2.3.1 Techniques

Shen-Orr et al. [61] defined *network motifs* as “patterns of interconnections that recur in many different parts of a network at frequencies much higher than those found in randomized networks.” They discovered three highly significant motifs composed by three–four nodes among which the most famous is the “feed-forward loop,” whose importance has been shown also in further studies [41, 42]. The technique presented in [61] laid the foundations for different extensions, such as [9, 11, 69]. In [69], composite motifs consisting of two kinds of interactions are extracted by exploiting edges of different colors in the network modeling. In particular, two types (colors) of edges are considered, representing protein–protein and transcription–regulation interactions, and algorithms are developed for detecting network motifs in networks with multiple types of edges. In [9], topological motifs derived from families of mutually similar, but not necessarily identical, patterns are discussed and extracted. The authors developed a search algorithm to extract topological motifs called *graph alignment*, in analogy to sequence alignment, that is based on a scoring function. All the approaches mentioned above relate the concept of *motif* only to the network topology. As observed in [37], there are biological networks (e.g., metabolic networks) where a purely topological definition of motifs seems to be inappropriate as similar topologies can give rise to very different functions. Thus, the authors of [37] introduce a new definition of motifs in the context of metabolic networks, such that the components of the network play the central part and the topology can be added only as a further constraint. Similarly to [37], in [48] the concept of motif is related to both graph structure and node similarity. In particular, the author presents a three-step exact approach based on the application of the notion of maximality, used extensively in strings and arrays [2, 4, 6, 26, 28, 56, 57], to graphs. In [32] the two notions of *structural* and *biological network motifs* are distinguished, focusing on the latter one that the authors explain as biologically significant small connected subgraphs regardless of the structure. They introduce five algorithms for the discovery of biological network motifs reducing the number of subgraphs to search by removing a number of edges from the original network and, at the same time, increasing the discovery rate for biological network motifs.

The search of significant motifs in biological networks is pioneered by Shen-Orr et al. [61], where *network motifs* have been defined as “patterns of interconnections that recur in many different parts of a network at frequencies much higher than those found in randomized networks.” The authors of [61] studied the transcriptional regulation network of *Escherichia coli*, by searching for small motifs composed by three–four nodes. In particular, three highly significant motifs characterizing such network have been discovered; the most famous is the “feed-forward loop,” whose importance has been shown also in further studies [41, 42].

The technique presented in [61] laid the foundations for different extensions, the main of which are [9, 11, 69].

In [69], composite motifs consisting of two kinds of interactions have been taken into account by exploiting edges of different colors in the network modeling. They modelled an integrated cellular interaction network by two types (colors) of edges, representing protein–protein and transcription–regulation interactions, and developed algorithms for detecting network motifs in networks with multiple types of edges.

In [9], topological motifs derived from families of mutually similar, but not necessarily identical, patterns have been discussed and extracted from the gene regulatory network of *Escherichia coli*. The authors developed a search algorithm to extract topological motifs called *graph alignment*, in analogy to sequence alignment that is based on a scoring function.

In [11], n -nodes “bridge” and “brick” motifs are searched for in complex networks, and the presence of such motifs has been associated with network topology, but not with network size. The authors proposed a method for performing simultaneously the detection of global statistical features and local connection structures, and the location of functionally and statistically significant network motifs.

All the approaches [9, 11, 61, 69] relate the concept of *motif* only to the network topology, without any consideration of possible properties shared by network nodes in terms of their mutual similarity.

As observed in [37], there are biological networks (e.g., metabolic networks) where a purely topological definition of motifs seems to be inappropriate as similar topologies can give rise to very different functions. Thus, the authors of [37] introduce a new definition of motifs in the context of metabolic networks, such that the components of the network play the central part and the topology can be added only as a further constraint.

Similarly to [37], in [48] the concept of motif is related to both graph structure and node similarity. In particular, the author presents a three-step exact approach based on the application of the notion of maximality, used extensively in strings, to graphs.

The two works [37, 48] open the way for the definition of new, exact or approximate, approaches for motif extraction taking into account not only the network topology but also the biological properties of the interacting components.

References

- Adamcsek, B., et al.: CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* **22**(8), 1021–1023 (2006)
- Amelio, A., Apostolico, A., Rombo, S.E.: Image compression by 2D motif basis. In: Data Compression Conference (DCC'11), pp. 153–162 (2011)
- Apostolico, A., et al.: Finding 3d motifs in ribosomal rna structures. *Nucl. Acids Res.* (2008)
- Apostolico, A., Parida, L.: Incremental paradigms of motif discovery. *J. Comput. Biol.* **11**(1), 15–25 (2004)
- Apostolico, A., Bock, M.E., Lonardi, S.: Monotony of surprise and large-scale quest for unusual words. *J. Comput. Biol.* **10**(2/3), 283–311 (2003)
- Apostolico, A., Parida, L., Rombo, S.E.: Motif patterns in 2D. *Theor. Comput. Sci.* **390**(1), 40–55 (2008)
- Bader, G., Hogue, H.: An automated method for finding molecular complexes in large protein-protein interaction networks. *BMC Bioinform.* **4**(2) (2003)
- Bandyopadhyay, S., Sharan, R., Ideker, T.: Systematic identification of functional orthologs based on protein network comparison. *Genome Res.* **16**(3), 428–435 (2006)
- Berg, J., Lassig, M.: Local graph alignment and motif search in biological networks. *Proc. Natl. Acad. Sci. USA* **101**(41), 14689–14694 (2004)
- Bruckner, S., Hüffner, F., Karp, R.M., Shamir, R., Sharan, R.: Torque: topology-free querying of protein interaction networks. *Nucl. Acids Res.* **37**(Web-Server-Issue), 106–108 (2009)
- Cheng, C.Y., Huang, C.Y., Sun, C.T.: Mining bridge and brick motifs from complex biological networks for functionally and statistically significant discovery. *IEEE Trans. Syst. Man Cybern. Part B* **38**(1), 17–24 (2008)
- Ciriello, G., Guerra, C.: A review on models and algorithms for motif discovery in protein-protein interaction network. *Brief. Funct. Genomics Proteomics* (2008)
- Ciriello, G., Mina, M., Guzzi, P.H., Cannataro, M., Guerra, C.: AlignNemo: A local network alignment method to integrate homology and topology. *PLOS One* **7**(6), e38,107 (2012)
- Denielou, Y.P., Boyer, F., Viari, A., Sagot, M.F.: Multiple alignment of biological networks: a flexible approach. In: CPM'09 (2009)
- Derenyi, I., Palla, G., Vicsek, T.: Clique percolation in random networks. *Phys. Rev. Lett.* **94**(16), 160–202 (2005)
- Dost, B., et al.: Qnet: a tool for querying protein interaction networks. In: RECOMB'07, pp. 1–15 (2007)
- Enright, A., Dongen, S., Ouzounis, C.: An efficient algorithm for large-scale detection of protein families. *Nucl. Acids Res.* **30**(7), 1575–84 (2002)
- Erdos, P., Renyi, A.: On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61 (1960)
- Ferraro, N., Palopoli, L., Panni, S., Rombo, S.E.: Master-slave biological network alignment. In: 6th International symposium on Bioinformatics Research and Applications (ISBRA 2010), pp. 215–229 (2010)
- Ferraro, N., et al.: Asymmetric comparison and querying of biological networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 876–889 (2011)
- Ferro, A., et al.: Netmatch: a cytoscape plugin for searching biological networks. *Bioinformatics* (2007)
- Fionda, V., Palopoli, L., Panni, S., Rombo, S.E.: Protein-protein interaction network querying by a “focus and zoom” approach. In: BIRD'08, pp. 331–346 (2008)
- Fionda, V., Panni, S., Palopoli, L., Rombo, S.E.: A technique to search functional similarities in PPI networks. *Int. J. Data Mining Bioinform.* (To appear)
- Flannick, J., Novak, A., Graemlin, S., et al.: General and robust alignment of multiple large interaction networks. *Genome Res.* **16**(9), 1169–1181 (2006)
- Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010)
- Furfaro, A., Groccia, M.C., Rombo, S.E.: Image classification based on 2D feature motifs. In: Flexible Query Answering Systems (FQAS 2013) (2013)

27. Garey, M., Johnson, D.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, New York (1979)
28. Grossi, R., Pisanti, N., Crochemore, M., Sagot, M.F.: Bases of motifs for generating repeated patterns with wild cards. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2**(3), 159–177 (2000)
29. Jancura, P., et al.: A methodology for detecting the orthology signal in a PPI network at a functional complex level. *BMC Bioinform.* (2011)
30. Kalaev, M., Bafna, V., Sharan, R.: Fast and accurate alignment of multiple protein networks. In: *RECOMB'08* (2008)
31. Kelley, B., Yuan, B., Lewitter, F., Sharan, R., Stockwell, B.R., Ideker, T.: Pathblast: a tool for alignment of protein interaction networks. *Nucl. Acid Res.* **32**, W83–W88 (2004)
32. Kim, W., Li, M., Wang, J., Pan, Y.: Biological network motif detection and evaluation. *BMC Syst. Biol.* **5**(Suppl 3), S5 (2011)
33. King, A.D., Pržulj, N., Jurisica, I.: Protein complex prediction via cost-based clustering. *Bioinformatics* **20**(17), 3013–3020 (2004)
34. Klau, G.W.: A new graph-based method for pairwise global network alignment. *BMC Bioinform.* **10**(Suppl. 1), S59 (2009)
35. Koyuturk, M., Kim, Y., Topkara, U., Subramaniam, S., Szpankowski, W., Grama, A.: Pairwise alignment of protein interaction networks. *J. Comput. Biol.* **13**(2), 182–199 (2006)
36. Kuchaiev, O., Przulj, N.: Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics* **27**(10), 1390–1396 (2011)
37. Lacroix, V., Fernandes, C.G., Sagot, M.F.: Motif search in graphs: application to metabolic networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **3**(4), 360–368 (2006)
38. Leskovec, J., Lang, K., Mahoney, M.: Empirical comparison of algorithms for network community detection. In: *Proceedings of the International World Wide Web Conference (WWW)*, pp. 631–640 (2010)
39. Liao, C.S., et al.: Isorankn: spectral methods for global alignment of multiple protein networks. *Bioinformatics* **25**, i253–i258 (2009)
40. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: a survey. *IEEE Trans. Comput. Biol. Bioinform.* **1**(1), 24–45 (2004)
41. Mangan, S., Alon, U.: Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**(21), 11980–11985 (2003)
42. Mangan, S., Itzkovitz, S., Zaslaver, A., Alon, U.: The incoherent feed-forward loop accelerates the response-time of the gal system of escherichia coli. *J. Mol. Biol.* **356**(5), 1073–1081 (2005)
43. Milo, R., et al.: Network motifs: simple building blocks of complex networks. *Science* **298**(5594), 824–827 (2002)
44. Mongiov, M., Sharan, R.: Global alignment of protein-protein interaction networks. In: Mamitsuka, H., DeLisi, C. Kanehisa, M. (eds.) *Data Mining for Systems Biology, Methods in Molecular Biology*, vol. 939, pp. 21–34. Humana Press (2013)
45. Neyshabur, B., Khademl, A., Hashemifar, S., Arab, S.S.: NETAL: a new graph-based method for global alignment of protein?protein interaction networks. *Bioinformatics* **29**(13), 11,654–1662 (2013)
46. Palla, G., et al.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005)
47. Parida, L.: *Pattern Discovery in Bioinformatics. Theory and Algorithms*. Chapman and Hall/CRC (2008)
48. Parida, L.: Discovering topological motifs using a compact notation. *J. Comput. Biol.* **14**(3), 46–69 (2007)
49. Park, Y., Song, M.: A genetic algorithm for clustering problems. In: *Proceedings of 3rd Annual Conference on Genetic Algorithms*, pp. 2–9 (1989)
50. Pinter, R., et al.: Alignment of metabolic pathways. *Bioinformatics* **21**(16), 3401–3408 (2005)
51. Pizzuti, C., Rombo, S.E.: Experimental evaluation of topological-based fitness functions to detect complexes in PPI networks. In: *Genetic and Evolutionary Computation Conference (GECCO)*, pp. 193–200 (2012)

52. Pizzuti, C., Rombo, S.E.: Multi-functional protein clustering in PPI networks. In: Proceedings of the 2nd International Conference on Bioinformatics Research and Development (BIRD), pp. 318–330 (2008)
53. Pizzuti, C., Rombo, S.E.: Pincoc: a co-clustering based approach to analyze protein-protein interaction networks. In: Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning, pp. 821–830 (2007)
54. Pizzuti, C., Rombo, S.E.: Restricted neighborhood search clustering revisited: an evolutionary computation perspective. In: Proceedings of the 8th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB), pp. 59–68 (2013)
55. Pizzuti, C., Rombo, S.E.: A coclustering approach for mining large protein-protein interaction networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9**(3), 717–730 (2012)
56. Rombo, S.E.: Optimal extraction of motif patterns in 2D. *Inf. Process. Lett.* **109**(17), 1015–1020 (2009)
57. Rombo, S.E.: Extracting string motif bases for quorum higher than two. *Theor. Comput. Sci.* **460**, 94–103 (2012)
58. Ruan, J., Zhang, W.: Identifying network communities with a high resolution. *Phys. Rev. E* **77**(1) (2008)
59. Sharan, R., et al.: From the cover: conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* **102**(6), 1974–1979 (2005)
60. Sharan, R., Ideker, T.: Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* **24**(4), 427–433 (2006)
61. Shen-Orr, S.S., Milo, R., Mangan, S., Alon, U.: Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature* **31**, 64–68 (2002)
62. Shih, Y.K., Parthasarathy, S.: Scalable global alignment for multiple biological networks. *BMC Bioinform.* **13**(Suppl 3), S11 (2012)
63. Shlomi, T., et al.: Qpath: a method for querying pathways in a protein-protein interaction network. *BMC Bioinform.* **7** (2006)
64. Singh, R., Xu, J., Berger, B.: Global alignment of multiple protein interaction networks. In: *PSB'08* (2008)
65. Singh, R., Xu, J., Berger, B.: Pairwise global alignment of protein interaction networks by matching neighborhood topology. In: *Research in Computational Molecular Biology (RECOMB 2007)*, pp. 16–31 (2007)
66. Van Dongen, S.: Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* **30**(1), 121–141 (2008)
67. Wu, X., Liu, Q., Jiang, R.: Align human interactome with phenome to identify causative genes and networks underlying disease families. *Bioinformatics* **25**(1), 98–104 (2009)
68. Yang, Q., Sze, S.H.: Saga: a subgraph matching tool for biological graphs. *J. Comput. Biol.* **14**(1), 56–67 (2007)
69. Yeger-Lotem, E., et al.: Network motifs in integrated cellular networks of transcription regulation and protein-protein interaction. *Proc. Natl. Acad. Sci. USA* **101**(16), 5934–5939 (2004)

Discriminative Pattern Discovery on Biological Networks

Fassetti, F.; Rombo, S.E.; Serrao, C.

2017, X, 45 p. 4 illus., Softcover

ISBN: 978-3-319-63476-0