

RESEARCH

Open Access



ConnectedAlign: a PPI network alignment method for identifying conserved protein complexes across multiple species

Jianliang Gao^{1*}, Bo Song², Xiaohua Hu², Fengxia Yan³ and Jianxin Wang¹

From 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017) Honolulu, Hawaii, USA. 30 May - 2 June 2017

Abstract

Background: In bioinformatics, network alignment algorithms have been applied to protein-protein interaction (PPI) networks to discover evolutionary conserved substructures at the system level. However, most previous methods aim to maximize the similarity of aligned proteins in pairwise networks, while concerning little about the feature of connectivity in these substructures, such as the protein complexes.

Results: In this paper, we identify the problem of finding conserved protein complexes, which requires the aligned proteins in a PPI network to form a connected subnetwork. By taking the feature of connectivity into consideration, we propose ConnectedAlign, an efficient method to find conserved protein complexes from multiple PPI networks. The proposed method improves the coverage significantly without compromising of the consistency in the aligned results. In this way, the knowledge of protein complexes in well-studied species can be extended to that of poor-studied species.

Conclusions: We conducted extensive experiments on real PPI networks of four species, including human, yeast, fruit fly and worm. The experimental results demonstrate dominant benefits of the proposed method in finding protein complexes across multiple species.

Keywords: Network alignment, Big data, Graph data analysis

Background

A protein complex is a biomolecular that contains a number of proteins interacting with each other to perform different cellular functions which is described in many prior works such as the work proposed by Hu et al. in [1]. The identification of protein complexes in a protein-protein interaction (PPI) network [2] can, therefore, lead to a better understanding of the roles of such a network in different cellular systems. It is for this reason that the protein complex identification problem has received a lot of attentions, and a considerable number of techniques and algorithms have been proposed to address such problem.

Graph structure is widely adopted in many applications [3, 4]. By representing a PPI network as a graph [5], whose vertices represent proteins and edges as interactions between proteins, these algorithms are able to identify clusters in single PPI network based on different graph properties [6]. For example, an uncertain graph model based method is proposed to detect protein complex from a PPI network [7]. To identify protein complexes, previous works proposed to consider not just topological but also biological information in the network [1]. However, they all focused on finding protein complexes in a single PPI network, and finding conserved protein complexes from multiple PPI networks still remains challenging.

Network alignment provides a possible way to identify protein complexes from multiple PPI networks [8]. Conserving functional and topological features are two goals for network alignment. Functional module represents a

*Correspondence: gaojianliang@csu.edu.cn

¹School of Information Science and Engineering, Central South University, 410083 Changsha, China

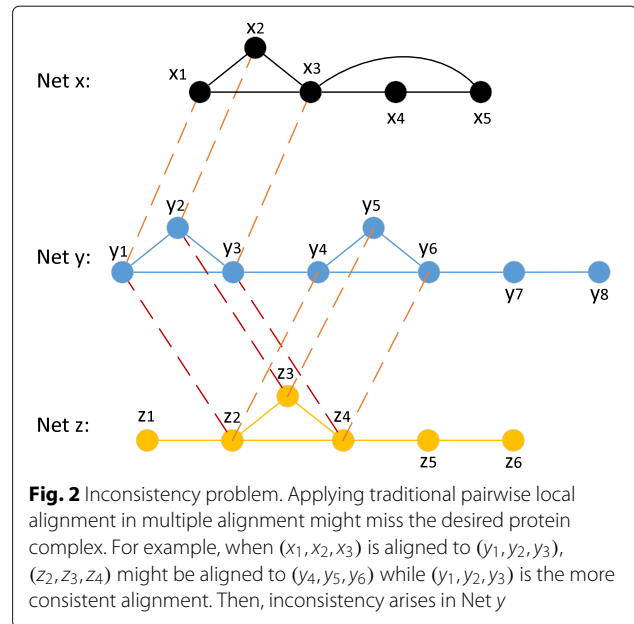
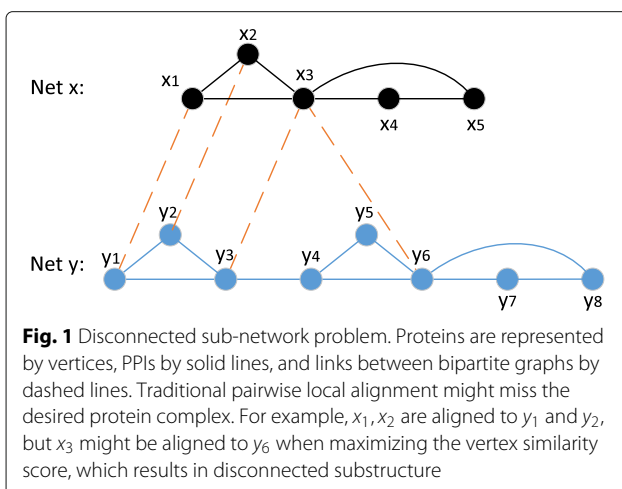
Full list of author information is available at the end of the article



collection of molecular interactions that work together to achieve a particular functional objective in a biological process, while topological module represents locally dense neighborhoods in a PPI network [9]. Network alignment can be categorized into two classes: global alignment and local alignment. Global alignment [10] finds overall best functional orthologs among entire PPI networks, while local alignment identify smaller conserved subnetworks in part of the networks [11]. In the context of local alignment, when a given small network is aligned with large networks, the problem can be projected as network query problem. In this paper, we concern more on the local alignment, which is more related to our problem.

Traditional pairwise network alignment detects functional orthologs of proteins in PPI networks by maximizing the similarity between proteins, while ignoring the subnetwork structure of protein complex. Therefore, the disconnected subnetwork problem might be caused when applying those methods to identify conserved protein complexes. For example, in Fig. 1, there are two PPI networks Net *x* and Net *y*. When aligning complex (x_1, x_2, x_3) in Net *x* to Net *y*, protein x_1 and x_2 are aligned with y_1 and y_2 . But only maximizing pairwise similarity of proteins might lead x_3 to be aligned with y_6 , which results in disconnected subnetwork in the alignment and doesn't meet well with the requirement of protein complex.

Aligning multiple networks promises additional insights into the protein complexes as well as the knowledge-transfer across multiple species. However the alignment of multiple PPI networks has additional challenges. For example, if directly applying the methods of pairwise network alignment to the multiple network alignment, inconsistency problem might be caused. For example, as shown in Fig. 2, the substructure (x_1, x_2, x_3) in Net *x* is aligned with (y_1, y_2, y_3) in Net *y*. When they are expected to be further aligned with the (z_2, z_3, z_4) in Net *z* from consistent perspective, (y_4, y_5, y_6) might be the best alignment



instead if it was a pairwise alignment between Net *y* and Net *z*. However, since the goal of multiple network alignment is to find conserved protein complexes across all PPI networks, (y_1, y_2, y_3) should be a better result.

In this paper, we propose a new approach to find conserved protein complexes by network alignment. The main contributions are as follows:

- We identify the problem of finding conserved protein complex via aligning multiple PPI networks. In this way, the knowledge of protein complexes in well-studied species can be extended to that of many poor-studied species.
- We propose an efficient method to find conserved protein complexes from multiple PPI networks. In this method, we take the feature of subnetwork connections into consideration, which improves the coverage significantly without compromising the consistency of aligned results.
- Extensive experiments are conducted on the PPI networks of four species including human, yeast, fruit fly and worm. These results in terms of coverage and consistency illustrate the dominant benefit of the proposed method in finding protein complexes across species.

Method

Problem definition

Definition 1. Target network: A PPI network $G_t = (V_t, E_t)$ is called target network if the given protein complexes to be aligned belong to G_t , where V_t is the set of proteins and E_t is the set of interactions between them.

The knowledge such as protein complexes of a target network can be extended to other PPI networks via network alignment. We define the other PPI networks as *aligned networks*.

Definition 2. *Aligned networks:* Let $G = \{G_i\}$ ($1 \leq i \leq \xi$) be the set of aligned networks, where ξ is the number of PPI networks to be aligned with target network. $G_i = (V_i, E_i)$ ($1 \leq i \leq \xi$) is the i^{th} PPI network to be aligned, where V_i, E_i are the sets of proteins and their interactions.

Given target network, aligned networks and protein complexes of target network, we define the input of the problem as follows.

Input: (1) The set of aligned networks $G = \{G_i, 1 \leq i \leq \xi\}$, where ξ is the number of aligned networks. (2) The set of well studied protein complexes in target network $G_t: S = \{S_1, S_2, \dots, S_\zeta\}$, where ζ is the number of protein complexes to be aligned.

Then the alignment result as the output is defined as follows.

Output: Without loss of generality, for any protein complex $M_0, M_0 \in S$, the alignment result is a matchset $M = \{M_1, M_2, \dots, M_\xi\}$ consists of a set of ξ subnetworks, where $M_k \subseteq G_k, 1 \leq k \leq \xi, G_k \in G$, which satisfies: (1) any $M_k \subseteq G_k$ is a connected subnetwork of G_k ; (2) maximizing the similarity score of $\{M_0, M_1, M_2, \dots, M_\xi\}$.

With the definitions and notations above, our algorithm of finding protein complexes across multiple PPI networks via network alignment mainly follows two procedures: assigning scores to proteins according to both biological and structural features, and then heuristically selecting proteins that form connected subnetwork in each PPI network which finally achieves optimized total score for multiple PPI networks.

Scoring strategy of network alignment

Overall, we utilize both the biological similarity between proteins and the topological structure to assign scores on subnetworks for subsequent heuristic selections of proteins. Formally, given a protein complex of target network $M_0 \subseteq G_t$, its match result $\{M_1, M_2, \dots, M_\xi\}$ in aligned networks, where $M_k \subseteq G_k$, is assigned with a real-valued score Φ :

$$\Phi = \sum_{k \in \{1, \dots, \xi\}} \sum_{v_j \in V_{M_k}} (\alpha * \delta_{bio}(v_j) + (1 - \alpha) * \delta_{topo}(v_j)) \tag{1}$$

where ξ is the number of PPI networks, V_{M_k} is the set of proteins in M_k , α is a coefficient to trade off biological and topological scores, δ_{bio} and δ_{topo} are the biological and topological scores respectively. In the following, we will describe the details of determining the δ_{bio} and δ_{topo} .

Assume $M_k \subseteq G_k$ is the current subnetwork to be assigned a score, where $G_k, 1 \leq k \leq \xi$, is the current aligned network. At each time, choose another PPI network denoted as $G_h, (h \neq k) \wedge (1 \leq h \leq \xi)$, then G_t, G_k, G_h construct a group of triple networks. Denote $M_h \subseteq G_h$ as subnetwork of G_h to align with M_0 . For every h , we calculate score for the proteins in G_k in the triple networks.

We use Fig. 3 as an example to show the method of assigning scores, where M_0 is the target subnetwork in target network Net x consisting of (x_1, x_2, x_3) , M_k is the subnetwork in aligned network Net y to be assigned scores consisting of (y_1, y_2, y_3) . And the subnetwork of (z_2, z_3, z_4) in aligned network Net z is to be aligned with M_0 .

Definition 3. *Link:* If a pair of proteins (u, v) comes from different PPI networks, and (u, v) are sequence similar, then (u, v) is called a link.

Sequence similarity [12] can be obtained with the BLASTP method [13]. We connect a dashed line to denote a link in this paper.

Definition 4. *Thread:* If triple proteins (u, v, t) comes from three different PPI networks, and there exist links between (u, v) , (u, t) and (v, t) at the same time. Then they form a thread.

The biological score of a protein consists of: (1) the number of links with the subnetwork M_0 , (2) the number of links with the subnetwork M_h , and (3) the number of threads among these three subnetworks which contain the current protein. We denote these three scores as $\delta_{bio}^1, \delta_{bio}^2, \delta_{bio}^3$. Taking y_1 in Fig. 3 as example, there are links (y_1, x_1) , (y_1, z_2) and thread (y_1, x_1, z_2) . Therefore, $\delta_{bio}^1, \delta_{bio}^2, \delta_{bio}^3$ of vertex are all "1". To avoid excessive influence of one

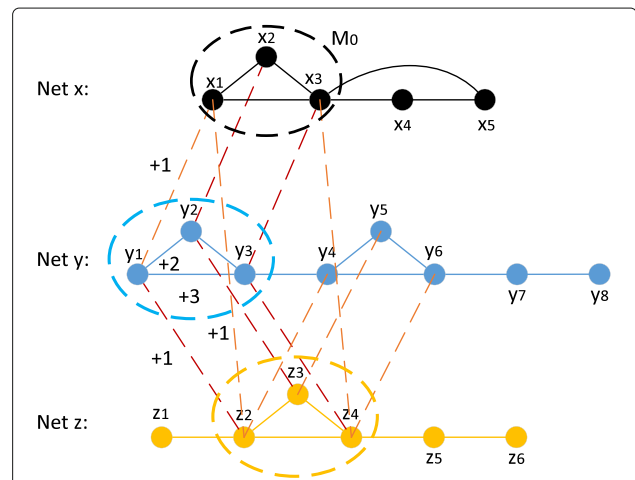


Fig. 3 Illustration of assigning scores. Net x is the target network, and M_0 is the given protein complex. Net y is an aligned network. Taking y_1 as example, its scores $(\delta_{bio}^1, \delta_{bio}^2, \delta_{bio}^3, \delta_{topo}^1, \delta_{topo}^2)$ are 1, 1, 1, 2, 3, respectively

factor, we adopt a transform techniques by multiplying a coefficient. The biological score of a protein u is:

$$\delta_{bio}(u) = (\delta_{bio}^1)^\lambda + (\delta_{bio}^2)^\lambda + (\delta_{bio}^3)^\lambda \quad (2)$$

where $\delta_{bio}^1, \delta_{bio}^2, \delta_{bio}^3$ are the numbers of links with M_0, M_h and the number of threads respectively. e is Euler number and λ ($\lambda > 1$) is the parameter of transform.

Definition 5. *Component: a connected graph $G_c = (V_c, E_c)$ is a component of subnetwork M_k if $G_c \subseteq M_k$.*

The topological score of a vertex consists of (1) the degree of current vertex; (2) the size of the maximal component that includes the current vertex. As the same with biological score, we adopt a transform techniques by multiplying a coefficient. The topological score of a vertex u is:

$$\delta_{topo}(u) = (\delta_{topo}^1)^\omega + (\delta_{topo}^2)^\omega \quad (3)$$

where δ_{topo}^1 is u 's degree in its subnetwork, and δ_{topo}^2 is the size of the maximal component that includes u . ω is a parameter of transform. In our method, $\omega > 1$.

Alignment algorithm

Given the multiple PPI networks and target protein complex from the target PPI network, the alignment process is shown in Algorithm 1, which It mainly includes:

(1) Generate initial candidate pools.

Only those proteins that have links with given protein complex can be selected as candidate proteins since links represent the biological similarity between proteins across PPI networks according to Definition 3. For each aligned network G_i $1 \leq i \leq \xi$, we construct a pool for a given protein complex M_0 , where $M_0 \in G_t$. All vertices in G_i are put into the pool of G_i if they have links with any vertex in M_0 , as shown in Line 5 of Algorithm 1. Then, the initial subnetworks M are selected randomly from the pools.

(2) Simulated annealing process.

Simulated annealing process adopts iteration method for global optimal solution. In each loop, a protein from the candidate pool is chosen randomly to be determined as aligned protein in the corresponding PPI network (Line 14 of Algorithm 1). On the other hand, there are two kinds of proteins that could be moved out from the current alignment solution (Line 13 of Algorithm 1). The first kind is the protein whose score is the lowest in the current solution: $\{v|v \in V_{M_\varepsilon} \wedge \text{argmin}_v \text{score}(v)\}$. The other

Algorithm 1: Alignment algorithm

Input: Set of PPI networks to be aligned $\{G_1, G_2, \dots, G_\xi\}$, a target PPI network G_t and protein complex $M_0 \subseteq G_t, T_{max}, \theta, K, N$

Output: Set of matched protein complexes $M = \{M_1, M_2, \dots, M_\xi\}$

```

1  $T_0 = T_{max}$ ;
  /* Initiate candidate pools: */
2 foreach  $g \in \{G_1, \dots, G_\xi\}$  do
3    $pool_g = NULL$ ;
4   foreach  $u \in V_{M_0}$  do
5      $pool_g = pool_g \cup \{v|v \text{ exists a link } \wedge v \in V_g\}$ 
6   end foreach
7    $M_g = \text{random}(pool_g, |V_{M_0}| * \theta)$ ;
8 end foreach
  /* Simulated annealing process: */
9 for  $i \leftarrow 1$  to  $K$  do
10   $T_i = T_0 - i * (T_{max} - T_{min})/K$ ;
11  foreach  $j \leftarrow 1$  to  $N$  do
12     $\varepsilon = j \text{ MOD } \xi$ ;
13     $v_{out} = \text{RandomGetOne}\{v|v \in V_{M_\varepsilon} \wedge (\text{degree}(v) == 0 \vee (\text{argmin}_v \text{score}(v)))\}$ ;
14     $v_{in} = \text{RandomGetOne}\{v|v \in pool_\varepsilon \wedge v \notin M_\varepsilon\}$ ;
15     $M'_\varepsilon = M_\varepsilon - v_{out} + v_{in}$ ;
16     $\Delta\Phi = \sum_{g \in \{G_1, \dots, G_\xi\}} (\Phi(M'_g) - \Phi(M_g))$ ;
17    if  $(\Delta\Phi > 0)$  or  $(\text{rand}(0, 1) < e^{\frac{\Delta\Phi}{sT_i}})$  then
18       $M_\varepsilon = M'_\varepsilon$ ;
19    end if
20  end foreach
21 end for
22 for  $i \leftarrow 1$  to  $\xi$  do
23    $M_i = \text{largest component in } M_i$ 
24 end for
25 return  $M = \{M_1, M_2, \dots, M_\xi\}$ 

```

kind is the protein whose corresponding vertex in the current subnetwork is not connected with other vertices, i.e., its degree is zero. As shown in Line 16–19 of Algorithm 1, if the new candidate solution achieves higher score, it will take place the previous solution. If not, it still has chance to replace the prior solution with a probability of $(\text{rand}(0, 1) < e^{\frac{\Delta\Phi}{sT_i}})$, where $e^{\frac{\Delta\Phi}{sT_i}}$ returns the selection threshold for the selection of simulated annealing process. Finally, the algorithm returns the best solution as the alignment of protein complexes $M = \{M_1, M_2, \dots, M_\xi\}$.

Results and discussion

In this section, we evaluate the performance of our method through extensive experiments. We compare our method to LocalAli [14] since LocalAli is the most recent

local alignment method for PPI networks. We measure the coverage and consistency of the alignment networks.

Dataset and experimental setup

Real-world PPI networks of four species are used in our experiments, including *Homo sapiens* (human), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (worm) and *Saccharomyces cerevisiae* (yeast) [15]. The detailed numbers of proteins and interactions for each species are listed in the Table 1.

We also obtained the corresponding sequences of all proteins from manually annotated and reviewed database UniProtKB/Swiss-Prot [16] for calculating pairwise protein similarity, i.e., e -value, by conducting BLASTP 2.3.0 (downloaded from the NCBI BLAST [17]) and setting e^{-7} as the e -value cutoff, to select the potential homologous proteins across different species. The corresponding Gene Ontology (GO) annotations of the proteins are collected from the Uniprot-GOA database for the alignment evaluations.

As human and yeast are the two best studied species [18], we build data sets by assigning them alternatively as the target PPI network for the alignment, and choose two from the rest of our collected PPI networks as aligned networks. There are total of six datasets generated, with each dataset as a group of multiple PPI networks to perform alignment. The composition of the six datasets are listed in Table 2.

With most local alignment algorithms that are pairwise, LocalAli [14] is one of the few most recent local alignment approaches. In LocalAli, a framework is proposed to reconstruct the evolution history of conserved modules based on a maximum-parsimony evolutionary model. LocalAli aims to identify functionally conserved modules from multiple biological networks, which is able to be used as a comparison method to our proposed algorithm. We run LocalAli with its default parameters on the six datasets in Table 2 to obtain target protein complexes, by retrieving every matchset in its results and obtain whose proteins form a component in the target network. The components from the target network are used as the input of our algorithm. In the experiment, we set the parameters $\alpha = 0.5, \theta = 1.1, K = 20, N = 100, T_{max} = 100, \lambda = 4.5, \omega = 3$. The results are compared with LocalAli in terms of coverage and consistence.

Table 1 Proteins and interactions of four species

PPI networks	# Proteins	# Interactions
A:Human(<i>H.sapiens</i>)	11,258	47,031
B:Worm(<i>C.elegans</i>)	9,302	15,669
C:Fruit Fly(<i>D.melanogaster</i>)	8,725	27,053
D:Yeast(<i>S.cerevisiae</i>)	5,494	54,163

Table 2 Datasets composition

Datasets	Target species A	Aligned species B	Aligned species C
D1	A:Human	B:Worm	C:Fruit Fly
D2	A:Human	B:Worm	D:Yeast
D3	A:Human	C:Fruit Fly	B:Worm
D4	D:Yeast	B:Worm	A:Human
D5	D:Yeast	B:Worm	C:Fruit Fly
D6	D:Yeast	C:Fruit Fly	A:Human

Coverage

A larger and denser connected component can give more insight of common topology of the network and it could be more biologically significant. The coverage analyzes the numbers of proteins in the aligned subnetworks from each aligned PPI networks with the given motifs in the target network.

As shown in Table 3, We compare our algorithm with LocalAli [14] on the six datasets, where D1~D3 are assigning human PPI network as the target network and D4~D6 get the yeast as the target network. For each dataset, since we utilize the largest component in the according target PPI network from the LocalAli as our target protein complex for alignment, the average number of proteins in every target network are all the same to that

Table 3 Comparison of coverage

Dataset	PPI network	Average size of protein complexes		Ratio
		Proposed	LocalAli [14]	
D1	target network A	7.05	7.05	100.00%
	aligned network B	4.73	3.64	129.95%
	aligned network C	5.54	3.07	180.46%
D2	target network A	8.56	8.56	100.00%
	aligned network B	4.99	2.75	181.45%
	aligned network D	9.89	3.99	247.87%
D3	target network A	6.86	6.86	100.00%
	aligned network C	3.07	3.07	100.00%
	aligned network B	7.86	5.79	135.75%
D4	target network D	7.74	7.74	100.00%
	aligned network B	7.09	3.47	204.32%
	aligned network A	8.82	7.79	113.22%
D5	target network D	5.05	5.05	100.00%
	aligned network B	4.28	3.95	108.35%
	aligned network C	3.55	3.32	106.93%
D6	target network D	5.81	5.81	100.00%
	aligned network B	4.31	4.13	104.36%
	aligned network A	6.88	5.75	119.65%

The ratio is the result obtained by dividing the average size of protein complexes of our proposed method by that of LocalAli

of the LocalAli, i.e., ratio is 100% for the target network. The ratio is the result obtained by dividing the average size of protein complexes of our proposed method by that of LocalAli. As in the aligned networks, our method can generate larger sizes of aligned protein complexes than that of the LocalAli among all datasets. One exception is in the dataset D3, where two method obtained equal coverage in one of the aligned networks, while obtaining much higher coverage in the other aligned networks. Similar situation exist in dataset D6. In dataset D1, D2, and D4, our algorithm achieves significantly higher coverage in all aligned networks, with the largest one has nearly 248% coverage to the LocalAli.

Consistency

The calculation of the consistency utilizes the Gene Ontology (GO) annotations associated to each of the proteins, with three basic types of ontologies describing biological properties: biological process (BP), molecular function (MF) and cellular component (CC) [19]. It is assumed that proteins with more similar GO annotations are more functionally coherent [20]. We calculate and analyze such functional similarity by the fraction of aligned proteins that share same GO annotations. The larger the fraction, the more biological significance the alignment has.

The consistency, specifically measured by the mean entropy (ME) and mean normalized entropy (MNE), serves as a specificity metric to measure the quality of alignment. To calculate ME, we first obtain the entropy $E(M)$ of a matchset M , i.e. the protein complexes aligned to one protein complex in the target species among all participated PPI networks, with following formulation:

$$E(M) = E(v_1, v_2, \dots, v_n) = - \sum_{i=1}^d p_i \times \log(p_i) \quad (4)$$

where p_i is the fraction of all proteins in the matchset M with the annotation GO_i , and d represents the total number of different GO terms in M . Thus the aligned matchset with more consistency will have lower entropy. The ME of the matchset is then calculated by averaging the entropies of all matchsets generated from the alignment to all the protein complexes in the target species, and the lower the ME of the alignment results, the higher consistency a method performs, indicating a better biological quality.

Similar to ME, for the MNE, we first calculate the normalized entropy $NE(M)$ for a matchset as:

$$NE(M) = NE(v_1, v_2, \dots, v_n) = - \frac{1}{\log d} \sum_{i=1}^d p_i \times \log p_i \quad (5)$$

where p_i and d have the same interpretation of those in the $E(M)$. The MNE of the alignment results is then computed by calculating the average of the normalized entropy of

all matchsets with their size. The lower MNE, the better functional consistency an alignment method achieves.

The comparison of consistency between the results from LocalAli and our algorithm is shown in Table 4. The ratio is the result obtained by dividing the ME or MNE of our proposed method by that of LocalAli then subtracting one. We can observe that in D1, D4, D5 and D6, our method generates aligned protein complexes with slightly higher ME and MNE than that of the LocalAli, where the ratio of the consistency less to LocalAli range from 0.76 to 6.48%. Meanwhile, we achieve higher ME and MNE than LocalAli in D2 and D3, with 8.12% better consistency at most.

For PPI network alignment, it is more important to achieve the alignment of functional modules than the alignment of proteins alone. The proposed ConnectedAlign achieves this goal without losing the consistency and coverage. In the future, the genome information could be used for biological network alignment [21].

Conclusion

In this paper, we proposed a novel approach to identify conserved protein complexes across different species. Given target protein complexes in the target network, the proposed method can find conserved protein complexes in multiple aligned PPI networks. Since we take the biological feature and topological feature into consideration, including subnetwork connectivity, our method achieves higher coverage significantly, and keeps stable consistence compared with previous network alignment method. The experimental results demonstrate the significant benefits of our proposed alignment method.

Table 4 Comparison of consistency

Dataset	Metrics	Consistency		Ratio
		Proposed	LocalAli [14]	
D1	ME	18.72	17.58	6.48%
	MNE	4.19	4.06	3.20%
D2	ME	33.76	35.97	-6.14%
	MNE	6.34	6.90	-8.12%
D3	ME	18.50	18.63	-0.70%
	MNE	4.27	4.30	-0.70%
D4	ME	35.70	34.59	3.21%
	MNE	6.62	6.57	0.76%
D5	ME	11.1	10.82	2.59%
	MNE	2.90	2.87	1.05%
D6	ME	17.72	17.44	1.61%
	MNE	4.13	4.09	0.98%

The ratio is the result obtained by dividing the ME or MNE of our proposed method by that of LocalAli then subtracting one

Acknowledgements

The abridged abstract of this work was previously published in the Proceedings of the 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017), Lecture Notes in Computer Science: Bioinformatics Research and Applications [22].

Funding

Publication of this article was supported partially by National Natural Science Foundation of China (NSFC) under grant 61471369, 61672536.

About this supplement

This article has been published as part of *BMC Bioinformatics* Volume 19 Supplement 9, 2018: Selected articles from the 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017): bioinformatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-9>.

Authors' contributions

JG, BS, XH and JW conceived the study and developed the model. BS wrote the code, cooperated with JG. JG and FY participated in algorithm development. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹School of Information Science and Engineering, Central South University, 410083 Changsha, China. ²College of Computing & Informatics, Drexel University, 19104 Philadelphia, USA. ³College of Liberal Arts and Sciences, National University of Defence Technology, 410073 Changsha, China.

Published: 13 August 2018

References

- Hu AL, Chan KCC. Utilizing both topological and attribute information for protein complex identification in ppi networks. *IEEE/ACM Trans Comput Biol Bioinforma.* 2013;10(3):780–92.
- Li M, Niu Z, Chen X, Zhong P, Wu F, Pan Y. A reliable neighbor-based method for identifying essential proteins by integrating gene expressions, orthology, and subcellular localization information. *Tsinghua Sci Technol.* 2016;21(6):668–77.
- Feng Q, Huang N, Jiang X, Wang J. Dealing with several parameterized problems by random methods. *Theor Comput Sci.* 2018;734(22):94–104.
- Gao J, Ping Q, Wang J. Resisting re-identification mining on social graph data. *World Wide Web-internet Web Inf Syst.* 2018. <https://doi.org/10.1007/s11280-017-0524-3>.
- Li M, Yang J, Wu FX, Pan Y, Wang J. Dynetviewer: a cytoscape app for dynamic network construction, analysis and visualization. *Bioinformatics.* 2018;34(9):1597–9.
- Maloddoğnin N, Pržulj N. L-graal: Lagrangian graphlet-based network aligner. *Bioinformatics.* 2015;31(13):2182–9.
- Zhao B, Wang J, Li M, Wu FX, Pan Y. Detecting protein complexes based on uncertain graph model. *IEEE/ACM Trans Comput Biol Bioinforma.* 2014;11(3):486–97.
- Faisal FE, Lei M, Crawford J, Milenković T. The post-genomic era of biological network alignment. *EURASIP J Bioinforma Syst Biol.* 2015;2015(1):1–19.
- Bhowmick SS, Seah BS. Clustering and summarizing protein-protein interaction networks: A survey. *IEEE Trans Knowl Data Eng.* 2016;28(3): 638–58.
- Elmsallati A, Clark C, Kalita J. Global alignment of protein-protein interaction networks: A survey. *IEEE/ACM Trans Comput Biol Bioinforma.* 2016;13(4):689–705.
- Gao J, Song B, Ke W, Hu X. Balanceali: Multiple ppi network alignment with balanced high coverage and consistency. *IEEE Trans Nanobioscience.* 2017;16(5):333–40.
- Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000;302(1): 205–17.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402.
- Hu J, Reinert K. Localali: an evolutionary-based local alignment approach to identify functionally conserved modules in multiple networks. *Bioinformatics.* 2015;31(3):363–72.
- Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al. The intact molecular interaction database in 2012. *Nucleic Acids Res.* 2012;40(D1): 841–6.
- Consortium U. The universal protein resource (uniprot) in 2010. *Nucleic Acids Res.* 2010;38(suppl 1):142–8.
- <https://blast.ncbi.nlm.nih.gov/>. Accessed 12 Jan 2018.
- Remmele CW, Luther CH, Balkenhol J, Dandekar T, Müller T, Dittrich MT. Integrated inference and evaluation of host-fungi interaction networks. *Front Microbiol.* 2015;6(764):1–18.
- Huntley RP, Sawford T, Mutowo-Muullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C. The goa database: gene ontology annotation updates for 2015. *Nucleic Acids Res.* 2015;43(D1):1057–63.
- Yu N, Li Z, Yu Z. A survey on encoding schemes for genomic data representation and feature learning from signal processing to machine learning. *Big Data Min Analytics.* 2018;1(3):191–210.
- Bérard S, Chateau A, Pompidor N, Guertin P, Bergeron A, Swenson KM. Aligning the unalignable: bacteriophage whole genome alignments. *BMC Bioinformatics.* 2016;17(1):17–30.
- Song B, Gao J, Hu X. Identifying conserved protein complexes across multiple species via network alignment. In: Proceedings of the 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017); 2017. p. 1008–1009.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

