IPSJ Transactions on Bioinformatics Vol.6 2–8 (Mar. 2013)

[DOI: 10.2197/ipsjtbio.6.2]

Original Paper

Global Network Alignment Method Using Node Similarity Based on Network Characteristics

Hitoshi Afuso[1,a)] Takeo Okazaki[2] Morikazu Nakamura[2]

Received: September 1, 2012, Accepted: November 1, 2012, Released: March 25, 2013

Abstract: Various methods to compare given biological networks have been proposed to date. For an instance, MI-GRAAL [8] is one of such popular methods. However, the method uses only local structural information to calculate a similarity among nodes. Owing to this limitation, the resulted alignment may not reflect the global features of the given networks. In social network analysis certain measurements, so-called *network characteristics* are used to capture some features of nodes in graphs. And some of these reflect global features of nodes in networks. In this paper, we proposed a network alignment method using a node similarity based on network characteristics so that resulted alignment would reflect the global structural features more than the traditional method. We compared our proposed method with traditional network alignment method, MI-GRAAL, to demonstrate the effectiveness of our proposal. The experiment was carried out through protein-protein interactions (PPI) networks of yeast and human. The results showed that proposed method led to better alignment in view of topological quality than MI-GRAAL.

Keywords: global network alignment, network characteristics, node similarity

1. Introduction

Improvements of the techniques to observe massive gene expressions in a time allow us to obtain various knowledge about biological networks in various species [1]. Understanding their structures and corresponding biological functions is one of the most challenging issues in the post-genomic era. Especially, transcriptional regulatory networks (TRNs) and protein-protein interactions (PPI) networks have vital role because their structures may explain underlying processes of the development of diseases. To date, various methods to estimate TRNs from gene expression profile data have been proposed [2], [3].

By comparing biological networks in different species, it is expected that we could reveal not only evolutionaly connections, but also conserved functions among them. Furthermore, such comparison is important for the evaluation of methods to estimate biological networks. To compare given biological networks, there are two major approaches traditionally, the method based on graph kernels and the network alignment. The methods based on graph kernels [4], [5] define the kernel function. Its value denotes certain similarity among graphs in a high-dimensional space. To construct such function, they use the information about shortest paths or random walks over given graphs. However, they directly calculate the similarity among graphs. So it becomes difficult to capture conserved substructures in given networks.

Unlike the method based on graph kernels, the network alignment tries to find a matching among nodes in different networks. The matching obtained from the network alignment shows us conserved subnetworks directly. In addition, we can also define a similarity between given biological networks based on the result of network alignment. As of now, various methods to utilize the network alignment among biological networks have been proposed [6], [7]. Like as the sequence alignment, the methods for network alignment are divided into two categories, local network alignment and global one. Local network alignment tries to find small subnetworks corresponding to pathways or protein complexes conserved in compared biological networks. In the result of local network alignment, certain node may be matched to some nodes in another network. In contrast, global network alignment matches one node in smaller network exactly once to one in larger. In other words, global network alignment constructs unique matching between networks.

In the early date of network alignment, local network alignment was considered as more valuable than global one while it was believed that conserved subnetwork is small across different species. However, recent report showed that large conservation in the PPI network of yeast and human [8]. Consequently, the global network alignment draws attention. As recent studies of global network alignment, we can cite the method by Terada et al. [9] and the one by Kuchaiev et al. [8]. Terada et al. proposed a global network alignment method, so-called *ALICE*, based on *abstract graph* that represents rough structure of given PPI networks. The abstract graph consists of some clusters of nodes that represent functional modules. In ALICE, two given PPI networks are transformed into abstract graphs and searching the optimum matching between clusters across the graphs are utilized. Although the method resulted biologically plausible alignment between PPI networks in nematoda and vinegar fly, it requires a parameter denotes the number of clusters in a abstract graph in advance and

[1] Graduate School of Engineering and Science, University of the Ryukyus, Nakagami, Okinawa 903–0213, Japan
[2] Faculty of Information Engineering, University of the Ryukyus, Nakagami, Okinawa 903–0213, Japan
[a)] lim_bay@ms.ie.ryukyu.ac.jp

there is no systematic method to determine it yet. On the other hand, Kuchaiev et al. [8] proposed a method, named *MI-GRAAL* uses a similarity among nodes in given networks. To calculate the similarity, the method enumerates small and connected subnetworks in compared networks. The enumerated small subnetworks are called *graphlet* and a vector that its each element denotes the frequency of certain graphlet is referred as *graphlet degree* [10]. Several studies reported the effectiveness of graphlet degree for capturing the similarity among nodes in PPI networks [11]. Using similarity based on the graphlet degree and *seed-and-extend* strategy, MI-GRAAL tries to find optimal matching between networks. As a consequence, it showed remarkable result, the large conservation of substructure in PPI networks between yeast and human. Despite its impacting result, MI-GRAAL has some difficulties. Firstly, some studies pointed out that the calculation of graphlet degree is computationally expensive. Secondly, it uses random break of ties in the similarity when it tries to find the seed of matching. Then, it may lead different matching from same input networks among different runs. In other words, results obtained by MI-GRAAL might be unstable among the different executions of alignment. And the most difficult point is that the graphlet degree only focuses on the subnetworks that includes only less than five nodes. Then it is difficult to say whether the obtained similarity and alignment reflect global structural feature accurately or not.

On the other hand, some measurements to capture features of nodes have been proposed in the area of social network analysis. Such measurements are called *network characteristics*. A social network is usually modeled as a graph that actors and social relationships among them are represented as nodes and edges, respectively. We can see clustering coefficient [12] and closeness centrality [13] for instances of network characteristics. Using such characteristics, we can obtain information about roles that each node has in a particular social network. Especially, some of the network characteristics can measure the global features of nodes in given graphs. From above consideration, by using the network characteristics, it is possible to define certain similarity among nodes considering the global features.

In this paper, we proposed new method to utilize pairwise global network alignment using node similarity based on network characteristics. Firstly, we defined a similarity among nodes using network characteristics. And next, new network alignment method was proposed based on the node similarity. Finally, to show the effectiveness of our proposed method, we compared it with traditional network alignment method, MI-GRAAL via PPI network in yeast and human. The results showed that our proposal could find larger subnetwork between yeast and human PPI networks than MI-GRAAL.

2. Materials and Methods

2.1 Outline of the Proposed Method

In our proposed method, we used a node similarity similar to MI-GRAAL because it does not require vital parameter in advance unlike ALICE. To solve the difficulty that traditional method has, it is required that proposed node similarity must reflect the global features of nodes. The proposed network alignment was designed according to *seed-and-extend* strategy similar to MI-GRAAL. To use such strategy, we have to avoid the random break of ties in the similarity when we determine the seed of matching for stable result of alignment. To summarize above discussions, following requirements must be met. The first point is that proposal node similarity must reflect the global feature of networks. And second point is that proposal could avoid the random break of ties to determine the seed of matching. In following sections, we provided solutions to meet the requirements stated above. Firstly, we explained about the network characteristics that we used to define a similarity among nodes.

2.2 Network Characteristics for Node Similarity

We showed a brief introduction of six network characteristics that we used, degree, namely clustering coefficient, closeness centrality, eccentricity centrality, betweenness centrality and PageRank.

Firstly, we explained local characteristics, degree and clustering coefficient. Degree denotes that the number of neighbors of particular node in a graph. This is the simplest local characteristic. However, its distribution in given network sometimes is used to check whether given network has some global structural property, so-called *Scale-free* property [14]. Clustering coefficient [12] is an index to measure how many neighbors of focal node are connected each other. This characteristic reflects the local complexity around a focal node. In the definition of it, if a node has only one neighbor, then we cannot calculate this characteristic. In this research, we assigned zero as clustering coefficient for such node.

As an examples of characteristics that reflect the global features in networks, we can see various types of centrality. In the area of social network analysis, the idea of centrality is used to measure how much each member has "central role" in given social network [13]. There are some variations of centrality according to the definition of "central role". In closeness centrality, a node that the other nodes can reach to it with smaller steps is considered as the central. Closeness centrality, $Closeness(v)_G$ of node v in graph G is calculated as the mean of lengths of shortest paths as follows.

$$Closeness(v)_G = \frac{|V| - 1}{\sum_{u \in V} d_G(u, v)} \quad (1)$$

where V denotes a node set in graph G and $d_G(u, v)$ is the length of shortest path between node u and v in graph G. This centrality is calculated from shortest paths that could reach to focal node v only.

In another definition of centrality, one may consider a node that could reach to the other nodes with smaller steps as the central. According to such view of centrality, eccentricity centrality is defined. It is calculated as the maximum length of shortest path. However, using this definition, one cannot handle disconnected graph. Then, in this research we modified it as follows.

$$Eccentricity(v)_G = \sum_{u \in V} \frac{d_G(v, u)}{|V| - 1} \quad (2)$$

In Eq.(2), $d_G(v, u)$ has same value to the number of nodes in G when node v cannot reach to u.

Betweenness centrality denotes how many times the shortest paths among nodes through the focal node. In other words, in betweenness centrality, a node that connects many other nodes directly or indirectly is considered as the central. This index is defined as the number of shortest paths that go through the focal node.

As another type of characteristics that capture the global features of nodes, we can cite PageRank [15]. It had been proposed by Page et al. to measure relative importance among Web pages. The basic concept of PageRank is that Web page that has links from other important Web pages is also important. It can be considered as the number of visits of the users that surf among Web pages according to hyper-links at random for enough long time. Node that has been visited by such *random surfing* users in many times has relatively high PageRank value.

Using these network characteristics shown above, we defined a similarity among nodes considering both local and global features of nodes in given networks.

2.3 Vector Representation and Node Similarity

Some definitions of similarities among nodes based on network characteristics explained above is provided. The definition consists of two parts, vector representation of nodes and definition of a similarity between the vectors. In this paper, we used four similarity measurements.

We defined a vector representation $\mathbf{f}_G(v)$ of node v in undirected graph G as follows.

$$\mathbf{f}_G(v) \rightarrow \begin{pmatrix} Degree_G(v) \\ ClusteirngCoefficient_G(v) \\ Closeness_G(v) \\ Eccentricity_G(v) \\ Betweenness_G(v) \\ PageRank_G(v) \end{pmatrix} \tag{3}$$

Next, using this vector representation in Eq.(3), we defined four node similarities. Suppose that two undirected graphs G_1 and G_2 are given, and u and v are nodes in G_1 and G_2, respectively. In such case, we can calculate the vector representation $\mathbf{f}_{G_1}(u)$ and $\mathbf{f}_{G_2}(v)$ according to Eq.(3). Traditionally, to measure the similarity between vectors, correlations and distances have been used. According to this convention, we defined three node similarities: $S_{pea}(u,v)$, $S_{spe}(u,v)$ and $S_{euc}(u,v)$ as follows.

$$S_{pea}(u,v) = \frac{\sum_{i=1}^{M}\left(f_{G_1}^{(i)}(u) - \overline{\mathbf{f}_{G_1}(u)}\right)\left(f_{G_2}^{(i)}(v) - \overline{\mathbf{f}_{G_2}(v)}\right)}{\sqrt{\sum_{i=1}^{M}\left(f_{G_1}^{(i)}(u) - \overline{\mathbf{f}_{G_1}(u)}\right)^2}\sqrt{\sum_{i=1}^{M}\left(f_{G_2}^{(i)}(v) - \overline{\mathbf{f}_{G_2}(v)}\right)^2}}$$

$$S_{spe}(u,v) = 1 - \frac{\sum_{i=1}^{M} D_i}{N^3 - N} \tag{4}$$

$$S_{euc}(u,v) = \frac{1}{1 + \sqrt{\sum_{i=1}^{M}\left(f_{G_1}^{(i)}(u) - f_{G_2}^{(i)}(v)\right)^2}} \tag{5}$$

where M denotes the dimensions of vector representations. As shown in above expressions, S_{pea} and S_{spe} denote Pearson's product-moment correlation and Spearman's rank correlation among samples, respectively. $f_G^{(i)}(u)$ denotes i-th element of the vector $f_G(u)$. $\overline{\mathbf{f}_G(u)}$ is a mean of the value in vector $\mathbf{f}_G(u)$. D_i and N represent the diffenrence of ranks between i-th element in the vector and N denotes the number of pairs of the elements.

On the other hand, Kuchaiev et al. proposed a similarity among nodes. It is called *"confidence score."* We used this as fourth similarity among ndoes. In this similarity, each network characteristics are treated as agents that have individual opinion about the similarity between nodes. And this similarity summarizes up each agents' opinion to determine focal pair of nodes should be matched or not. Calculation steps of confidence score consist of following steps.

(1) Calculate the difference of each network characteristics between nodes i and j in graph G_1 and G_2. Arranging these results to (i, j) element in a matrix, we can obtain differential matrix $D_{X_k}(G_1, G_2)$ where X_k denotes the characteristic that used to calculate this matrix. In Kuchaiev et al. [8], five characteristics were used.

(2) Calculate $conf_{X_k}(i, j)$ from differential matrices $D_{X_k}(G_1, G_2)$. The $conf_{X_k}(i, j)$ represents the fraction of elements in the i-th row of difference matrix $D_{X_k}(G_1, G_2)$ that are strictly greater than $D_{X_k}(G_1, G_2)_{\{i,j\}}$. By arranging $conf_{X_k}(i, j)$, we can construct a matrix $conf_{X_k}(G_1, G_2)$.

(3) Sum up $conf_{X_k}(G_1, G_2)$ to confidence matrix $Conf(G_1, G_2)$. Then, confidence matrix $Conf(G_1, G_2)$ is calculated as $Conf(G_1, G_2) = \sum_{X_i} Conf_{X_i}(G_1, G_2)$. It is said that confidence score is robust to minor error in individual difference matrix because the index is based on simple majority vote.

Using these measurements, we can calculate the similarity between undirected graphs.

Based on above four similarities, we proposed a global network alignment method.

2.4 Network Alignment based on the Proposed Node Similarity

We proposed new global network alignment method, called *GrAliNe* (Graph Aligner based on Network Characteristics) using node similarity defined above.

Traditionally, some different formulation of the global alignment problem have been proposed by Flannick et al. [16] and Zaslavskiy et al. [17]. Unlike the case of sequence alignment, any reasonable formulation of this problem makes it computationally hard. The reason is that problem contains *subgraph isomorphism* problem as its subproblem. Given two graphs, subgraph isomorphism problem asks which one graph is contained as exact subgraph of the other. This problem is known to belong to NP-complete class [18]. In this paper, we used standard definition of the global network alignment problem as follows. Suppose two networks $G_1(V_{G_1}, E_{G_1})$ and $G_2(V_{G_2}, E_{G_2})$ are given, the problem is to find a bijective mapping $m : V_{G_1} \rightarrow V_{G_2}$. Mapping m is called *total* if all nodes in V_{G_1} will be mapped into some nodes in V_{G_2} and *injective* if it doesn't map different nodes in V_{G_1} to identical node in V_{G_2}. Total and injective mapping is called *bijective*.

Similar to MI-GRAAL, GrAliNe is based on *seed-and-extend* strategy. It consists of two steps, selection of the seed of matching in given graphs and extension of current matching. The main algorithms of GrAliNe were shown in **Fig. 1** and **Fig. 2**. In Fig. 1, the graph G raised to power p is defined as $G^p = (V(G), E^p)$, where $E^p = \{(u_1, u_2) : d_G(u_1, u_2) \leq p\}$. As the power p is in-

```
1: procedure GrAliNe(Undirected graph G, H, Similarity matrix S)
2:     Give cannonical labeling to G.
3:     alignedPairs ← φ
4:     while There are unaligned nodes in G do
5:         find maximal similar pair (u, v) from similarity matrix S.
6:         alignedPairs ← alignedPairs ∪ {(u, v)}
7:         newAlignedPair ←AlignLocally(u, v, G, H, S)
8:         alignedPairs ← newAlignedPair
9:         if There are still unaligned nodes in G then
10:            raise the graph to next power.
11:        end if
12:    end while
13:    return alignedPairs
14: end procedure
```

Fig. 1 Main procedure of GrAliNe.

```
1: procedure AlignLocally(node u0, v0, graph G, H, similarity matrix S)
2:     newAlignedPairs ← φ
3:     nextProcessingPairs ← {(u0, v0)}
4:     finishedFlag ← false
5:     while !finishedFlag do
6:         temporalyPairs ← φ
7:         finishedFlag ← true
8:         for all (u, v) in nextProcessingPairs do
9:             neighborsU ← neighbors of u that not aligned yet
10:            neighborsV ← neighbors of v that not aligned yet
11:            if neighborsU and neighborsV are not φ then
12:                newPairs ← FindMaximalMatching(neighborsU,
    neighborsV, S)
13:                temporalPairs ← temporalPairs ∪ {newPairs}
14:                finishedFlag ← false
15:            end if
16:        end for
17:        nextProcessingPairs ← temporalPairs
18:        newAlignedPairs ← newAlignedPairs ∪ {temporalPairs}
19:        if nextProcessingPairs is not φ then
20:            sort matchings in nextProcessingPairs by their similarity
    value.
21:        end if
22:    end while
23:    return newAlignedPairs
24: end procedure
```

Fig. 2 Subroutine of GrAliNe.

creasing, indirect descendent of nodes in current matching would be included the candidates of next matching. In this sense, the rising a graph G to the power p corresponds to the insertion of gaps that lengths are $p-1$ in the matching just like the gap insertions in sequence alignment. Unlike MI-GRAAL, our method transforms given graph G to canonical form to obtain the consistent result for identical input graphs even though some pairs of nodes have same similarity value. Using the transformation to canonical form of graphs, we can avoid the instability of alignment results via unique input networks. In Fig. 2, we used the Hungarian algorithm [19] to find the maximal matching between candidates in each graph. Using this algorithm, we can utilize the global network alignment considering global features of nodes.

2.5 Indices to Measure the Quality of Alignment

We used two indices to measure the quality of the alignment result, edge correctness and the size of largest common connected subgraph same as Kuchaiev et al. Edge correctness denotes that how many edges in smaller graph are preserved in larger graph by the alignment. Edge correctness score EC of an alignment m between graph $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ is calculated as follows.

$$EC = \frac{|(u,v) \in E_1 \wedge (m(u), m(v)) \in E_2|}{min(|E_1|, |E_2|)} \times 100\% \quad (6)$$

where m denotes a matching between node sets V_1 and V_2. In global network alignment, each node in smaller graph should be mapped to a node in larger graph injectively. Then, in the calculation of EC score, the smaller number of edges is used as the denominator. As another measurement for the topological quality, largest common connected subgraph ($LCCS$) also has been used in various studies. Since large and contiguous subgraph is preferred than small and disconnected ones, greater size of $LCCS$ is desirable in a result of network alignment. Unlike to the discussion in Kuchaiev et al. [8], we did not assess the biological validity of the alignment because any information about biological similarity among proteins was not used in our research. Only structural quality measures were considered in following experiments.

Table 1 Summary of PPI nework in yeast and human.

Species	# of Nodes	# of Edges	Density of Edges
Yeast	1994	15819	0.00398
Human	8934	41341	0.00051

2.6 Benchmark Data Setup

As benchmark data, we used the PPI network data in yeast from Collins et al. [20] and also the one in human from Radivojac et al. [21]. These networks contain one large connected component and a number of small components. Since we used only structural information not biological one, small components isolated from large one don't have much information about the structure. Then, only largest connected components in these networks were used in our experiments. The summary of each PPI network were shown in **Table 1**. Using these PPI networks, the comparison with MI-GRAAL was demonstrated.

3. Results and Discussion

To show the effectiveness of GrAliNe, we compared it with traditional network alignment method, MI-GRAAL. The reason why we used MI-GRAAL for comparison was that the method is the first global network alignment method in the world showed the importance and potential of global network alignment.

3.1 Features of PPI Networks in Yeast and Human

From our definition of network alignment problem, if one tries to utilize the global network alignment between very large and complex network and relatively simple and small one, any matching between two networks will be agreed as a plausible result in view of the topological quality, even though that matching were obtained at random. In other words, such comparison may corrupt global network alignment problem itself. To confirm the complexity in our comparisons, some rough analyses were demonstrated in advance.

Firstly, we checked the distribution of clustering coefficient in

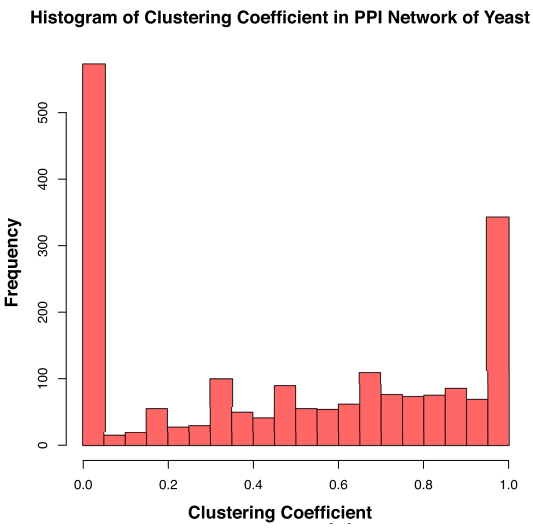

Fig. 3 Histogram of clustering coefficient in yeast.

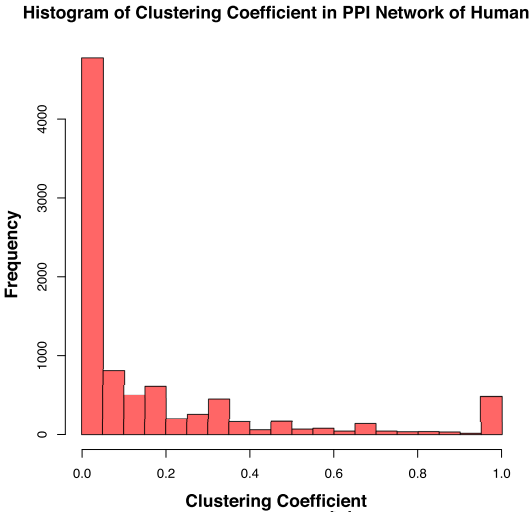

Fig. 4 Histogram of clustering coefficient in human.

each PPI network to compare the local complexity around each node. Histograms of clustering coefficients in each PPI network were shown in **Fig. 3** and **Fig. 4**. In Fig. 3 and Fig. 4, width of each bin is 0.05. As we can see from these figures, PPI network in yeast has greater number of the nodes that have high clustering coefficient value than human's. This fact shows that PPI network in yeast has relatively complex local structures compared to human's.

Next, we demonstrated the analysis about degree distributions in each PPI network to compare the rough global structures. Degree distributions in each PPI network were shown in **Fig. 5**. In Fig. 5, each axe was log-scaled. The shape of both degree distributions showed that each PPI network has scale-free property. In this perspective, these two networks have similar global structure. However, even though PPI network in human have many nodes than yeast's, the one in yeast has greater ratio of the nodes that have high degree. This denotes that some parts of global structure in PPI network in yeast don't match with the one in human.

From above discussions, we can see that finding the matching between PPI networks in yeast and human is enough complex task that random matching or poor strategic approach could not

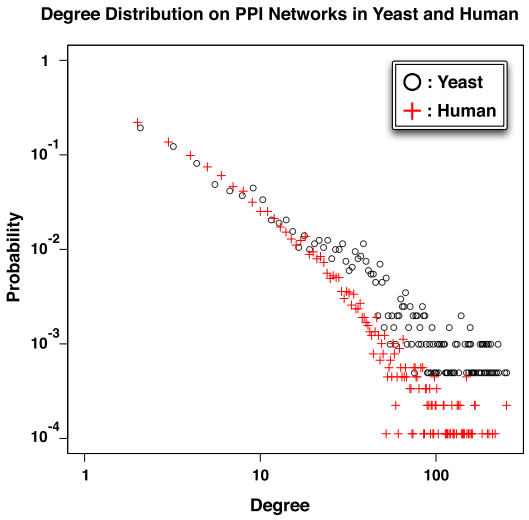

Fig. 5 Comparison of degree distribution between PPI network in yeast and human.

Table 2 Results with gap insertions.

Method	*EC* Score	Size of *LCCS*
S_{pea}	20.20 %	1917
S_{spe}	16.13 %	1879
S_{euc}	17.42 %	1892
$Conf$	13.23 %	1661
MI-GRAAL	18.68 %	1853

Table 3 Results without gap insertions.

Method	*EC* Score	Size of *LCCS*
S_{pea}	20.38 %	1917
S_{spe}	16.34 %	1879
S_{euc}	17.61 %	1892
$Conf$	14.12 %	1661

obtain plausible solution.

3.2 Characterization of GrAliNe and MI-GRAAL

Two experiments were demonstrated in the rest of this section to show the effectiveness of our proposed method, GrAliNe. In the first experiment, we compared alignments obtained by GrAliNe with four different similarities shown in Section 2.2 and MI-GRAAL, in two different situations, taking into account the gap and otherwise. Because it is very difficult problem to determine the timing of gap insertion, then we utilized comparisons in such simple two situations. And the second experiment was demonstrated to investigate the effect of the gap insertions to GrAliNe, especially in *LCCS* score. In the experiment, the numbers of obtained connected components were compared. The results in the first comparison experiment were shown in **Table 2** and **Table 3**. They showed the results in the case with or without gap insertions, respectively. In the Table 3, the result of MI-GRAAL is not shown because we simply referred the value shown in the Kuchaiev et al. [8]. Results in Table 2 showed that except the result obtained from confidence score, GrAliNe lead better results than MI-GRAAL in *LCCS* score. Especially, in the case of GrAliNe with similarity by Pearson's correlation lead the best *EC* score among compared methods. And also, comparing the results in Table 2 and Table 3, we can see that *EC* score had little improvement and *LCCS* scores are same in every case. From this result, the gap insertions was not effective for the improvement of

Table 4 The number of obtained components with or without gap insertion.

Similarity Measure	With Gap Insertion	Without Gap Insertion
S_{pea}	54	74
S_{spe}	71	114
S_{euc}	62	102
$Conf$	313	334

EC score for GrAliNe.

In the second experiment, we investigated the effects of the gap insertions for GrAliNe in *LCCS* score. In this experiment, the numbers of connected components were compared. The results were shown in **Table 4**.

From Table 4, we can see that the gap insertion resulted greater number of the connected components. However, as shown Table 2 and 3, the size of largest common connected subgraph were not changed in both cases. This showed that the gap insertion broke small connected components in the case without the gap insertion into more smaller ones. This fact denotes that the gap insertion effects only minor parts of the alignment, but not the major one for GrAliNe.

To sum up above results, our proposed method GrAliNe could obtain larger conserved subnetwork between two PPI networks in yeast and human than the traditional method.

4. Conclusion

We proposed a network alignment method, GrAliNe, using the node similarity based on network characteristics that could consider both local and global features in given networks. Comparison with a traditional network alignment method, MI-GRAAL, showed that GrAliNe could find larger substructure in PPI networks between yeast and human.

By modifying the network characteristics that we used in this paper to the ones can deal with directed graphs, we can define new similarity among nodes and global network alignment method for directed graphs based on it. As future tasks, we see such extension of proposed method and application to the comparison among biological networks modeled as directed graphs.

References

[1] Lee, I., Li, Z. and Marcotte, E.M.: An Improved, Bias-Reduced Probabilistic Functional Gene Network of Baker's Yeast, *Saccharomyces cerevisiae, PLoS ONE*, Vol.2, No.10, e988 (online), DOI: 10.1371/journal.pone.0000988 (2007).

[2] Mahdi, R., Madduri, A.S., Wang, G., Strulovici-Barel, Y., Salit, J., Hackett, N.R., Crystal, R.G. and Mezey, J.G.: Empirical Bayes Conditional Independence Graphs for Regulatory Network Recovery, *Bioinformatics*, Vol.28, No.15, pp.2029–2036 (2012).

[3] Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. and Guthke, R.: Gene Regulatory Network Inference: Data Integration in Dynamic Models-a Review, *Biosystems*, Vol.96, No.1, pp.86–103 (2009).

[4] Borgwardt, K.M. and Kriegel, H.P.: Sortest-Path Kernels on Graphs, *Proc. 5th IEEE International Conference on Data Mining* (2005).

[5] Neuhaus, M. and Bunke, H.: A Random Walk Kernel derived from Graph Edit Distance, *Structural, Syntactic and Statistical Pattern Recognition*, Lecture Notes in Computer Science, Vol.4109, pp.191–199 (2006).

[6] Kelley, B.F., Yuan, B., Lewitter, F., Sharan, R., Stockwell, B.R. and Ideker, T.: PathBLAST: A Tool for Alignment of Protein Interaction Networks, *Nucl. Acids Res.*, Vol.32(Suppl 2), pp.W83–W88 (2004).

[7] Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M. and Ideker, T.: Conserved Patterns of Protein Interaction in Multiple Species, *PNAS*, Vol.102, No.6, pp.1974–1979 (2005).

[8] Kuchaiev, O. and Pržulj, N.: Integrative Network Alignment Reveals Large Regions of Global Network Similarity in Yeast and Human, *Bioinformatics*, Vol.28, No.10, pp.1390–1396 (2011).

[9] Terada, A. and Sese, J.: Global Network Alignment using Graph Summarization for Comparison Gene Function, *SIG-BIO*, Vol.24, No.12, pp.1–7 (2011).

[10] Pržulj, N.: Biological Network Comparison using Graphlet Degree Distribution, *Bioinformatics*, Vol.23, No.2, pp.177–183 (2006).

[11] Kuchaiev, O., Milenković, T., Memišević, V., Hayes, W. and Pržulj, N.: Topological Network Alignment Uncovers Biological Function and Phylogeny, *Journal of the Royal Society Interface*, Vol.7, No.50, pp.1341–1354, (2010).

[12] Liu, Z., Wang, C., Zou, Q. and Wang, H.: Clustering Coefficient Queries on Massive Dynamic Social Networks, *Proc. 11th International Conference on Web-age Information Management*, pp.115–126 (2010).

[13] Eppstein, D. and Wang, J.: Fast Approximation of Centrality, *Journal of Graph Algorithm and Applications*, Vol.8, No.1, pp.39–45 (2004).

[14] Khanin, R. and Wit, E.: How Scale-Free are Biological Networks, *Journal of Computational Biology*, Vol.13, No.3, pp.810–818 (2006).

[15] Brin, S. and Page, L.: The Anatomy of a Large-scale Hyper Textual Web Search Engine, *Computer Networks and ISDN System*, Vol.30, Issues1-7, pp.107–117 (1998).

[16] Flannick, J., Novak, A., Srinivasan, B.S, McAdams, H.H. and Batzoglou, S.: Graemlin: General and Robust Alignment of Multiple Large Interaction Networks, *Genome Res.*, Vol.16, No.9, pp.1169–1181 (2006).

[17] Zaslavskiy, M., Bach, F. and Vert, J.P.: Global Alignment of Protein-Protein Interaction Networks by Graph Matching Methods, *Bioinformatics*, Vol.25, No.12, pp.259–267 (2009).

[18] Cook, S.A.: The Complexity of Theorem-Proving Procedures, *Proc. 3rd Ann ACM Symp on Theory of Computing*, pp.151–158 (1971).

[19] Ozaki, H. and Shiraishi, I.: Theory of Graphs and Networks, CORONA publishing Co. Ltd (1973).

[20] Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S. and Krogan, N.J.: Toward a Comprehensive Atlas of the Physical Interaction of Saccharomycess Cerevisiae, *Molecular and Cellular Proteomics*, Vol.6, No.3, pp.439–450 (2007).

[21] Radivojac, P., Peng, K., Clark, W.T., Peters. B.J., Mohan, A., Boyle, S.M. and Mooney, S.D.: An Integrated Approach to Inferring Gene-Disease Associations in Humans, *Proteins*, Vol.72, No.3, pp.1030–1037 (2008).

Appendix

This appendix describes the outline of procedure, FindingMaximalMatching in Fig. 2. This procedure is based on Hungarian algorithm. In the first of the procedure, to make all elements in the similarity matrix positive, the minimal similarity value is subtracted from each element in similarity matrix. Next, the maximal value of each row is subtracted from each row. After that, the maximal value of each column is subtracted for each column. By these transformations, we could obtain the modified similarity matrix that its elements corresponding to matching candidates are zero, otherwise negative values. If it is possible to determine the matching from candidates, we construct it according to the order of nodes in the first graphs. Otherwise, we cover all zero elements in the modified matrix with the minimal number of horizontal or vertical lines. Although there are some candidates of covering, we can choose any candidate. After the covering, we obtain the maximal value that uncovered with the lines and subtract it from the elements that also uncovered. The value is added to intersection points of the horizontal and vertical lines. Iterating this operation until each column or row has at least one zero element, we can obtain the maximal matching.

Hitoshi Afuso received his B.S. and M.S. degrees in Information Engineering from University of the Ryukyus in 2005 and 2008, respectively. He belongs to doctoral course in same university. His research area is Bioinformatics. Especially, he is interested in stochastic model of genetic expression and alignment of biological network such as TRNs.

Takeo Okazaki took B.Sc. and M.Sc. from Kyushu University in 1987 and 1989, respectively. He had been a research assistant at Kyushu University from 1989 to 1995. He has been a lecturer at University of the Ryukyus since 1995. His research interests are statistical data normalization for analysis, statistical causal relationship analysis.

Morikazu Nakamura took B.E. and M.E. from University of the Ryukyus in 1989 and 1991, respectively. He took Ph.D. from Osaka University in 1995. He has been a professor at University of the Ryukyus. His research interest includes design and analysis of parallel and distributed algorithms.

(Communicated by *Jun Sese*)