ACOGNA: An Efficient Method for Protein-Protein Interaction Network Alignment

Ha Tran Ngoc
Thai Nguyen University of Education
hatn@tnu.edu.vn

Huan Hoang Xuan
Vietnam National University-Hanoi
huanhx@vnu.edu.vn

***Abstract* — Protein–protein interaction network alignment enables us to identify orthologous proteins, predict protein functions and build evolutionary relationships of species. This article introduces an algorithm for global alignment of protein-protein interaction network based on ant colony optimization (ACO) method. The experimental results offer outstanding advantages of the proposed algorithms.**

Keywords—Ant colony optimization, PPI network, network alignment.

I. INTRODUCTION

Protein-protein interaction alignment (PPI Networks) allows us to discover conserved regions amongst them, hence effectively keeping us informed and able to predict unknown functions or verify known ones of proteins and further understanding protein and evolutionary relations amongst species. [2,8,11,26].

Methods of network alignment (NA) are divided into two approaches: local network alignment (LNA) and global network alignment (GNA). The initial methods of NA belong to the LNA which compares to indentify topology/ sequence similarity subnets. (Read [9,13,14,21,25]). However, that the aligned subnets are overlapping leads to many-to-many ambiguous mappings. GNA compares two entire networks, so it can overcome this drawback of LNA and is being extensively studied [1,3-6,15,18,20,21,26-30].

PPI network alignment is a combination of sequence similarity and topography-similarity information to identify an injection from the source network to the target network, so we can recognize orthologs as accurately as possible. However, in real bio-data alignment, the right alignment has remained unknown while there is currently no common assessment measure of quality. Notably, EC measure was suggested by Kuchaiev et al. (2010), based on the edge proportion of source network conserved when aligning to assess the suitability of alignment based on the source network. Patro and Kingsford (2012) developed ICS measure to assess alignment quality based on the ratio of conserved edge number to the number of edges of the sub-graph in the network. Saraph et al (2014) [28] found that in case of difference in the edge density of two different networks, the two measures fail, resulting a suggestion of a more suitable S^3 measure. To combine with sequence similarity information, Adalag et al (2013) [1] suggested GNAS measure to balance the EC and Blast bit-scores.

GNA problem proved to be a hard NP problem [1]; therefore, the majority of GNA are heuristics algorithms [1,6,15,16,19,20,30]. Recently, Saraph et al [28] have proposed the algorithm called MAGNA by inheritance methods applied to randomly - initialized population combined with the solution proposed by state-of-the-art algorithms: IsoRank, MI-GRAAL and GHOST. MAGNA and the new version of MAGNA++ [29] use a quality measure of alignment - S^3, and evidenced by the experiments, they improved considerably the quality of solution for initializing algorithms.

On the other hand, using GNAS alignment quality measure, Adalag et al. [1] proposed algorithm of SPINAL which produces better solution than IsoRank and MI-GRAAL. Dong et al. [6] proposed FASTAn which is better than SPINAL in terms of both quality and running time. Both of two algorithms are heuristic.

This article introduces the ACOGNA algorithm based on ant colony optimization method, used GNAS correctness/measure for PPI network global alignment. Experiments have shown that the proposed algorithm gives better results than the SPINAL and FASTAn. Particularly, the solution to problems in which source graphs have big number of edges is of better quality as opposed to MAGNA ++ under S^3 correctness.

Apart from the conclusion, the remaining parts in the article are structured as follows: Section 2 – an introduction to problems of PPI network global alignment, as well as relevant definitions and tasks; Section 3 – novice algorithms; and Section 4 – experimental results in comparison with the state-of-the-art algorithms.

II. THE GLOBAL ALIGNMENT PROBLEM OF PPI NETWORKS AND RELATED WORKS

The problem of GNA stated as optimization problem with the objective function of GNAS as in [10]; however, each alignment is stated as an injection between the two networks as used in[28].

A. Problem Definition

Consider 2 graphs and describe 2 protein interaction networks, with V_1, V_2 respectively being sets of vertices corresponding to set of proteins in G_1 and G_2 networks; E_1, E_2 are set of edges describing the interactions between proteins in

G_1 and G_2 respectively. Without loss of generality we can make such an assumption where | V | symbolizes the element number of V.

Definition 1. (network alignment) Every injection from V_1 to V_2 is called an alignment of the two networks G_1 and G_2.

The problem of two-network GNA is to find an optimal alignment for an correctness assessing the given quality based on sequence similarity and topology similarity. Given a NA f denote $f(E_1) = \{(f(u), f(v)) \in E_2 : (u,v) \in E_1\}$ and $f(V_1) = \{f(v) \in V_2 : v \in V_1\}$.

Alignment quality assessment: This article will consider the problem of finding the maximum alignment of alignment correctness GNAS [1]:

$$\alpha \left| f\left(E_1\right) \right| + (1-\alpha) \sum_{(u,v) \in E_1} similar(u,v) \quad (1)$$

where $\alpha \in [0.1]$ is the parameter, *similar (u, v)* is a sequence similarity measure, for example, BLAST bit-scores or E-values. The advantage of GNAS is the ability to show both topology correlation and sequence similarity between two networks aligned.

Kuchaiev et al. proposed using EC measure and Patro proposed ICS measure.

$$EC = \frac{\left| f(E_1) \right|}{\left| E_1 \right|} \quad (2)$$

$$ICS = \frac{\left| f(E_1) \right|}{\left| E\left(G_2[f(V_1)]\right) \right|} \quad (3)$$

where $E\left(G_2\left[f(V_1)\right]\right)$ is the set of G_2 edges in the sub graph with a set of vertices $f(V_1)$. We see the first term in Eq (1) is proportional to the score of EC. Saraph et al [28] claimed that when the density of edges in two networks are different, then these two measure fail and that S^3 score should be used

$$S^3 = \frac{\left| f(E_1) \right|}{\left| E_1 \right| + \left| E(G_2\left[f(V_1)\right]) \right| - \left| f(E_1) \right|} \quad (4)$$

B. Related Works

Adalag et al. Spinal proposed heuristics algorithms of SPINAL (2013) to the problem of GNA with GNAS score. Its solution proves to be better than IsoRank and MI-GRAAL, as shown by the experiment. Dong et al. [6] proposed the two-phase FASTAn algorithm, with the former being procedures to construct raw alignment based on sequence similarity and the latter being to reconstruct reinforced solution. Experiments have shown that FASTAn is better than SPINAL in both solution quality and running time. A simplified algorithm version based on ant colony optimization has been introduced by authors of this article in [27], whose experimental results show a better result than FASTAn.

Saraph et al proposed the algorithm based on genetic MAGNA methods to the problem with the objective function S^3. In MAGNA solutions by such algorithms as IsoRank, MI-GRAAL and GHOST are used as the initial population. Experiments have shown that its solution is by far better than those by IsoRank, MI-GRAAL and GHOST.

Before the introduction of the new algorithm, we briefed the ant colony optimization (ACO) with local search.

ACO method

This method, recommended by Dorigo (1991), is the method of metaheuristics – a simulation of the way that real ants find pathsusing pheromone trailslike reinforcement learning information (see [7]). In the ACO, the original problem is translated into the problem of finding a solution on the structural graph G = (V, E, Ω, η, T), where V is set of vertices, E is set of edges, and η and T are respectively a set of heuristics information and pheromone trails (representing reinforcement learning information) which can be put at the vertices or the edges. Each acceptable solution is a path that satisfies Ω path, starts from a vertice in C_0 subset of V, then extends through random walk procedures based on heuristics information and pheromone trails. ACO algorithms used artificial N_{ant}. In each iteration, each ant finds a solution based on a random walk procedures on the structural graph, which in turn is evaluated and applied with local search procedures. Then it is evaluated again and updated with pheromore as reinforcement learning information for ants to find a solution in the next iteration.

Many algorithms use the same graph G (V, E). However, for information heuristics, because of difference in rules of pheromone updating and searching techniques, G (V, E) here after will be named as the structural graph.

In the application of ACO, 4 important factors greatly affect algorithm performance: 1) The structural graph and random step procedure to find a solution, 2) heuristicsinformation, 3) pheromone updating rule, 4) techniques of local search .

III. ACOGNA ALGORITHM

A. The components of ACOGNA

Given G_1, G_2 graphs; parameter α and similarity measures of vertice pairs $< u_i, v_j >$trong đó $u_i \in V_1, v_j \in V_2$. For each node set V_{12} of $V_1 \times V_2$, we denote $V_{12}^1 = \{u_i \in V_1 :< u_i, v_j > \in V_{12}\}, V_{12}^2 = \{v_j \in V_2 :< u_i, v_j > \in V_{12}\}$ (V_{12}^i is sets of vertices of node set V_i of the G_i aligned graph. ACOGNA algorithm's components are described as follows:

Construction Graph

The outline of an ACOGNA algorithm is given in Fig 1, including 2 layers, with the i^{th} showing G_i graph. The vertices of the upper layer are connected with all lower–layer vertices.

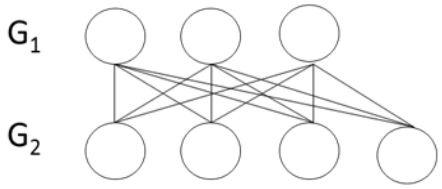

Fig 1. The Construction Graph of ACOGNA algorithm

Pheromone trails and heuristic information

The pheromone trail τ_j^i on (i,j) edge, aligned $u_i \in G_1$ with vertex $v_j \in G_2$ started up by τ_{max} and then updated again after each iteration using (7).

Heuristic information η_j^i is calculated using (5).

$$\eta_j^i = \alpha * M + (1-\alpha) * similar(u_i, v_j) \quad (5)$$

where α is a parameter in GNAS measure, M is the total number of edges preserved after alignment if u_i is aligned with v_j, $similar(u_i, v_j)$ is the similarity between the u_i and v_j

Random walk procedures for alignment construction

Constructing a solution, ants will iteratively to take an unaligned vertex in the 1st layer and align it one vertex in the 2nd layer using the eq (6) until all the vertices of G_1 are aligned.

$$p_j^i = \frac{(\tau_j^i)^a * [\eta_j^i]^b}{\sum_{k \in R\_V_2} (\tau_k^i)^a * [\eta_k^i]^b} \quad (6)$$

where $R\_V_2 = V_2 - V_{12}^2$ are vertices in G_2 that are not aligned yet.

After finding j vertex to be aligned from i^{th} vertex of the first-layer, the algorithm will select next first-layer vertex whose edges are connected with sets of vertices in the 1st aligned layer.

Pheromone updating rules

At each iteration, the best solution in loop by ants are applied with local search procedures to increase the solution quality. Then this best solution is used to update the pheromone on edges according to the Smooth Max-Min Ant System (SMMAS) updating rules of pheromone [10] and was given by Eq (7) and Eq (8):

$$\tau_j^i = (1-\rho)\tau_j^i + \Delta_j^i \quad (7)$$

$$\Delta_j^i = \begin{cases} \rho * \tau_{max} & (i,j) \in \text{ best solution} \\ \rho * \tau_{min} & (i,j) \notin \text{ best solution} \end{cases} \quad (8)$$

Where τ_{max} and τ_{min} are predetermined parameters, $\rho \in (0,1)$ is predetermied evaporation parameters.

Local search procedures

In each loop, after all ants have finished constructing a solution. Local search procedures are applied to the best explanation of the loop (I_{best}). Local search procedure is conducted as follows:

Step 1. Remain vertex n_{best} in the set A_{12} with the best score according to the criteria given by the equation (9):

$$score(u_i) = \alpha \times w(u_i) + (1-\alpha) \times similar(u_i, f(u_i)) \quad (9)$$

where $u_i \in V_1$ and $f(u_i)$ in V_2 vertices connected to u_i in A_{12}, $w(u_i)$ is the number of nodes $u_j \in V_1$ that $< u_i, u_j > \in E_1$ and $< f(u_i), f(u_j) > \in E_2$

Step 2. For $k = n_{best} + 1$ to $|V_1|$:

2.1. Procedures of **find_next_node** (): Find vertex $u_i \in V_1 - V_{12}^1$ having the maximum number of edges connecting to nodes in V_{12}^1.

2.2. Procedures of **choose_best_matched_node** (u_i) vertex $v_j \in V_2 - V_{12}^2$ so that a complemented $< u_i, v_j >$ into V_{12} leads to *GNAS(A12)* calculated by the biggest form (1) where A_{12} is the graph whose vertex is the V_{12} set and edges are induced by G_1, G_2.

2.3. Complementing $< u_i, v_j >$ in V_{12};

2.4. Updating E_{12} based on V_{12};

After each local search procedures, we have a new alignment as input A_{12} for the next iteration; the process is repeated until no improvement to GNAS (A_{12}) is made again.

In local search procedures, the value of n_{best} that determined the number of vertices aligned in A_{12} set is retained to construct the new alignment innext iterations. If n_{best} is too small, the reconstruction of alignment will take time and fail to save vertice pairs which even get appropriate alignment. Conversely, if n_{best} is too big, is is possible to save the not-well-aligned pairs of vertices, leading to a poor improvement for the solution quality. In algorithms of ACOGA, n_{best} value is fixed, and in ACOGNA algorithm, growing use is applied to this value. The purpose of this change is to save greater number of aligned pair of vertices after each local search procedure. This will increase the solution quality and reduce the time to implement Local search procedure after each reiteration.

B. ACOGNA algorithm

Step 1. Initialize pheromone trail matrix and a set A includes *m* ants

Step 2. While (stop conditions not satisfied) do

Each ant executes the follow steps.

2.1. Initialize V_{12} with a node pair $< u_i, v_j >$ of the maximum similarity score.

2.2. For *k= 2* to $|V_1|$

2.2.1. Find a node $u_i \in V_1 - V_{12}^1$ that has the most of number of edges connecting to the nodes belong to V_{12}^1;

2.2.2. Find node $v_j \in V_2 - V_{12}^2$ follow the random walk procedure which is described in part B as eq (6)

2.2.3. Add $< u_i, v_j >$ to V_{12};

2.2.4. Update E_{12} based on V_{12};

2.3. Call local search procedure with the best solution to improve the score.

2.4. Update the best solution.

2.5. Update the pheromone train with SMMAS rule base on the best solution.

Step 3. Save the best solution.

Note that, at 2.2.1 step, if finding more than one node $u_i \in V_1 - V_{12}^1$, the procedure will choose a random node among such.

GNAS score is used to select the best ant.

IV. EXPERIMENT

Experiments have been done on 4 benchmark data sets that had been used in [1]. ACOGNA algorithm was executed with different values of n_{best} parameter and number of ant at each iteration. The experimental results indicate that with n_{best} = 1% and the ant number is 6, ACOGNA gets the best performance.

Our experiments are performed on a computer with following configuration: CPU Intel Core 2 Duo 2.53Ghz, RAM DDR2 3GB and Windows 7 operating system.

Parameters are set as follows: The number of ants at each iteration is 6, $\alpha = \beta = 1$, $\tau_{max} = 1.0$ and $T_{\min} = \frac{1}{|V_1| + |V_2|}$.

A. Data

The dataset was used to compare the methods are 4 data sets were used to evaluate the quality of the Spinal and FASTAn solution. They are the protein-protein interaction networks: Saccharomyces cerevisiae (sc), Drosophila melanogaster (dm), Caenorhabditis elegans (ce), and Homo sapiens (hs). The interaction network are obtained from [22]. Description of this dataset is shown in Table 1. From the data set to create six pairs aligned network (ce-dm, ce-hs, ce-sc, dm-hs, dm-sc, hs-sc).

Table 1. Data description

Dataset	No. of proteins	No. of interactions
ce	2805	4495
dm	7518	25635
sc	5499	31261
hs	9633	34327

B. Experimental results

Experiments to compare ACOGNA and FASTAn according to GNAS score

ACOGNA, FASTAn and Spinal are algorithms which optimize GNAS score. The experiment results show that FASTAn outperformed Spinal in [6], so this experiment only compare the GNAS score of ACOGNA and FASTAn.

Because ACOGNA and FASTAN are random algorithm, they were executed 30 times for each pair of study PPI networks. The GNAS end E12 scores were averaged over those calculated from such 30 resulting alignments. The comparative results with α values respectively 0.3, 0.4, 0.5, 0.6 and 0.7 are shown in Table 2.

Table 2. COMPARISONS OF ACOGNA AND FASTAN ACCORDING TO GNAS AND E12 CRITERIA USING DIFFERENT VALUES OF THE PARAMETER α.

Datasets	α = 0.3		α = 0.4		α = 0.5		α = 0.6		α = 0.7	
	FASTAn	**ACOGNA**	**FASTAn**	**ACOGNA**	**FASTAn**	**ACOGNA**	**FASTAn**	**ACOGNA**	**FASTAn**	**ACOGNA**
ce-dm	778.46 2560.7	**833.14** **2749.2**	1109.92 2564.6	**1109.92** **2758.2**	1290.11 2567.2	**1368.35** **2726.10**	1545.86 2567.7	**1641.35** **2728.3**	1801.24 2567.6	**1930.84** **2753.7**
ce-hs	863.46 2842.8	**913.39** **3015.3**	1144.17 2838.1	**1207.94** **3001.1**	1429.89 2844.9	**1513.4** **3014.2**	1708.81 2838.0	**1824.69** **3033.0**	1994.87 2843.4	**2091.43** **2982.6**
ce-sc	834.79 2761.1	**876.78** **2902.9**	1109.93 2761.2	**1178.46** **2934.8**	1389.21 2769.7	**1457.65** **2907.9**	1663.39 2766.5	**1742.2** **2898.3**	1936.83 2763.1	**2064.12** **2945**
dm-hs	2260.31 7478.3	***2431.59*** ***8058.4***	3007.11 7481.9	**3226.76** **8038.7**	3755.36 7429.0	**4039.68** **8060.1**	4496.45 7478.2	**4828.29** **8034.29**	5242.32 7478.8	**5648.18** **8060.9**
dm-sc	1977.82 6569.7	**2108.13** **7008.7**	2631.85 6565.5	**2811.97** **7019.2**	3290.03 6570.7	**3518.87** **7030.2**	3950.16 6577.4	**4203.53** **7000.90**	4603.41 6572.3	**4908.90** **7009.6**
hs-sc	2268.21 7531.8	**2429.12** **8072.9**	3017.96 7528.5	**3256.54** **8182.8**	3772.96 7535.2	**3938.3** **7666.0**	4520.51 7527	**4895.45** **8153.4**	5279.88 7538.1	**5693.4** **8129.8**

Remark: The results in table 2 show that with all test set ce-dm, ce-sc, ce-hs, dm-sc, dm-hs, hs-sc and with all values of α, the GNAS and E12 scores of ACOGNA are better than FASTAn.

Experiments to compare ACOGNA and MAGNA++ according to EC and S^3 score

The EC, ICS and S^3 measures are used to assess the topology similarity of the alignment of two networks. This experiment will compare results of ACOGNA with MAGNA++ what is an algorithm optimized for S^3 measure. There are 3 optimal options in MAGNA ++ is EC, ICS and S^3.

Table 3 shows the results of the comparison of two algorithms with EC measure, ACOGNA algorithm was run with the values of α respectively 0.3, 0.4, 0.5, 0.6, 0.7. Its EC scores are shown in the corresponding column, MAGNA ++ algorithm was run with 3 options to optimize the EC, ICS and S^3. These results are shown in the corresponding column is EC, ICS and S^3. The best results of each measure are shown in bold format.

Table 3. COMPARISONS OF ACOGNA AND MAGNA++ ACCORDING TO S^3 SCORE

Datasets	ACOGNA				MAGNA++			
	α = 0.3	**α = 0.4**	**α = 0.5**	**α = 0.6**	**α = 0.7**	**EC**	**ICS**	**S^3**
ce-dm	0.6116	**0.6136**	0.6065	0.6070	0.6126	0.5217	0.07	0.2983
ce-hs	0.6708	0.6677	0.6706	**0.6747**	0.6635	0.3061	0.1288	0.3065
ce-sc	0.6458	0.6529	0.6469	0.6448	**0.6553**	0.6002	0.1449	0.2901
dm-hs	**0.3144**	0.3136	0.3144	0.3159	0.3138	0.1921	0.0885	0.1464
dm-sc	0.2242	0.2245	**0.2249**	0.2239	0.2240	0.1575	0.0569	0.1357
hs-sc	0.2582	0.26	0.2608	**0.2608**	0.2601	0.1680	0.0390	0.1439

*Remark***:** The results in Table 3 indicate that in all of the test set and all value of α parameter, the EC score of ACOGNA is always better than MAGNA++.

Table 4 shows the comparative results of two algorithms under measure S^3, the S^3 results of ACOGNA are shown in the column corresponding to α values respectively 0.3, 0.4, 0.5, 0.6, 0.7, and the results of MAGNA ++ algorithm which was run with 3 options to optimize the EC, ICS and S^3, are shown in the corresponding column as EC, ICS and S^3. The best results of each measure are shown in bold format.

Table 4. Comparison of the S^3 score

Datasets	ACOGNA				MAGNA++			
	α = 0.3	**α = 0.4**	**α = 0.5**	**α = 0.6**	**α = 0.7**	**EC**	**ICS**	**S^3**
ce-dm	0.1344	0.1123	0.1068	0.1338	0.1061	0.1580	0.07	**0.2597**
ce-hs	0.1265	0.0993	0.0953	0.0939	0.0909	0.2621	0.1284	**0.2639**
ce-sc	0.1063	0.0953	0.0925	0.0911	0.0922	0.1198	0.1446	**0.2573**
dm-hs	**0.1593**	0.1559	0.156	0.1567	0.1555	0.0988	0.0785	0.1088
dm-sc	**0.1446**	0.1417	0.1415	0.1407	0.1406	0.1030	0.0554	0.1081
hs-sc	**0.1501**	0.1452	0.1484	0.1446	0.1433	0.1043	0.0387	0.1166

*Remark***:** For S^3 measure, when running with datasets dm-hs, dm-sc, hs-sc with all values of **α** parameter, ACOGNA gets better score than MAGNA++ with 3 optimization options of MAGNA++. However, with 3 test set as ce-dm, ce-sc, ce-hs MAGNA++ get better score than ACOGNA.

V. CONCLUTION AND FUTURE WORKS

This paper proposes an efficient algorithm base on ant colony optimization for global alignment problem of PPI network. The experimental results showed that the proposed algorithms get better results than the state-of-the-art algorithm with GNAS, E12 and EC score in all cases. Although not using the S^3 criteria is the objective function but in case of the source graph is thick, the proposed algorithm get better S^3 score than MAGNA++.

Local search procedures are used in the algorithm depends on the values n_{best}. It currently selected manually. In the near future we will study to determine this parameter automatically.

As well as the other ACO-based algorithms, ACOGNA could be easily implemented as parallel to work with the large PPI networks and to decrease running time.

REFERENCES

[1] Aladag, A.E. and Erten, C. (2013), *SPINAL: scalable protein interaction network alignment*. Bioinformatics, Vol. 29 no 7, 917–924

[2] Bogdanov P. and Singh A. K. (2010), *Molecular function prediction using neighborhood features*, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), vol. 7, no. 2, pp. 208–217.

[3] Clark c. and Kalita J. (2014), *A compaison of algorithms for the pairwise alignment of biological networks*, Bioinformatics, Vol 30, No. 16, pp2351-2359

[4] Chindelevitch,L. et al. (2010), *Local optimization for global alignment of protein interaction networks*. In: Pacific Symposium on Biocomputing,Hawaii,USA, pp. 123–132

[5] Chindelevitch L. et al. (2013*), Optimizing a global alignment of protein interaction networks*, Bioinformatics, Vol. 29 no. 21,2765–2773.

[6] Do Duc Dong et al,. (2015), *An efficient algorithm for global alignment of protein-protein interaction networks*, Proceeding of ATC15, pp. 332-336

[7] Dorigo M.(2004), *Ant Colony Optimization*. MIT Press, Cambridge.

[8] Dutkowski,J. and Tiuryn,J. (2007), *Identification of functional modules from conserved ancestral protein–protein interactions*. Bioinformatics, 23, i149–i158.

[9] Flannick,J. et al. (2006) *Graemlin general and robust alignment of multiple large interaction networks*. Genome Res., 16, 1169–1181.

[10] Hoang Xuan H. et al. "*Solving the Traveling Salesman Problem with Ant Colony Optimization: A Revisit and New Efficient Algorithms.*" REV Journal on Electronics and Communications Vol. 2, No. 3–4, July – December,2012, 121-129

[11] B.H. Junker and F. Schreiber, *Analysis of Bological Networks*, wiley, 2008

[12] Kelley,B.P. et al. (2003), *Conserved pathways within bacteria and yeast as revealed by global protein network alignment*. Proc. Natl Acad. Sci. USA, 100, 11394–11399.

[13] Kelley,B.P. et al. (2004*), Pathblast: a tool for alignment of protein interaction networks*. Nucleic Acids Res., 32,83–88.

[14] Koyuturk,M. et al. (2006),*Pairwise alignment of protein interaction networks*. J. Comput. Biol., 13, 182–199.

[15] Kuchaiev,O. et al. (2010), *Topological network alignment uncovers biological function and phylogeny*. J. R. Soc. Interface., 7, 1341–1354.

[16] Kuchaiev,O. and Przulj,N. (2011) *Integrative network alignment reveals large regions of global network*

similarity in yeast and human. Bioinformatics, 27, 1390–1396.

[17] Kuhn HW: *The Hungarian Method for the assignment problem*. Naval Res Logistics Q 1955, 2:83-97.

[18] Liao,C.S. et al. (2009) *IsoRankN: spectral methods for global alignment of multiple protein networks*. Bioinformatics, 25, i253–i258.

[19] Memisevic,V. and Przulj,N. (2012), *C-graal: common-neighbors-based global graph alignment of biological networks*. Integr. Biol., 4, 734–743.

[20] Milenkovic,T. et al. (2010), *Optimal network alignment with graphlet degree vectors*. Cancer Inform.,Vol.9, 121–137.

[21] Narayanan,M. and Karp,R.M. (2007), *Comparing protein interaction networks via a graph match-and-split algorithm*. J. Comput. Biol., Vol. 14, 892–907.

[22] Park,D. et al. (2011) *IsoBase: a database of functionally related proteins across PPI networks*. Nucleic Acids Res., 39, 295–300

[23] Patro, R. and Kingsford, C. (2012). *Global network alignment using multiscale spectral signatures*. Bioinformatics, 28(23), 3105–3114.

[24] Remm,M. et al. (2001), *Automatic clustering of orthologs and in-paralogs from pairwise species comparisons*. J. Mol. Biol., 314, 1041–1052.

[25] Sharan,R. et al. (2005), *Conserved patterns of protein interaction in multiple species*. Proc. Natl Acad. Sci. USA, 102, 1974–1979.

[26] Singh,R. et al. (2008), *Global alignment of multiple protein interaction networks*. In: Pacific Symposium on Biocomputing. pp. 303–314.

[27] Tran Ngoc Ha, Hoang Xuan Huan (2015), *An Efficient Ant –based Algorithm for Global Alignment of Protein-Protein Interaction Networks*, Proc. of the 8 th national conf. on FAIR, Hanoi 2015, pp 471-477 (in Vietnamese).

[28] V. Saraph and T. Milenkovic (2014), *Magna: Maximizing accuracy in global network alignment*, Bioinformatics, vol. 30, no. 20, pp. 2931–2940.

[29] Vijayan V., Saraph V., and Milenkovic T.(2015), *Magna++: Maximizingaccuracy in global network alignment via both node and edge conservation*" Bioinformatics,1-3

[30] Zaslavskiy,M. et al. (2009) *Global alignment of protein-protein interaction networks by graph matching methods*. Bioinformatics, Vol.25, 259–267