An Efficient Ant Based Algorithm For Global Alignment Of Protein-protein Interaction Networks

Trần Ngọc Hà[1], Hoàng Xuân Huấn[2]
[1]Thai nguyen University of Education, [2]Vietnam National University, Hanoi, Vietnam.

Introduction

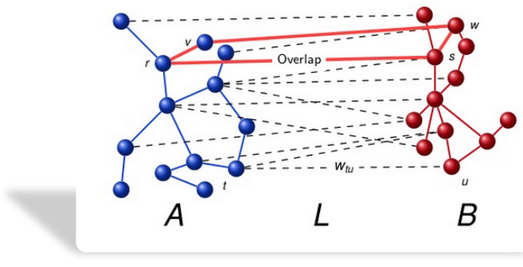
- Aligning two protein-protein interaction networks is an essentially important task in bioinformatics. It is a challenging and widely studied research topic in recent years. Accurately aligned networks allow us to identify functional modules of proteins and/ororthologous proteins from which unknown functions of a protein can be inferred
- PPI network alignment methods fall into two approaches: **local alignment** and **global alignment.**
- **Local network alignment**
  - ✓ The goal is to identify from the input PPI networks, subnetworks that closely match in terms of network topology and/or sequence similarities.
  - ✓ Typically many overlapping subnetworks from a single PPI network are provided as part of the local alignments; this gives rise to ambiguity, as a protein may be matched with many proteins from a target PPI network.
- **Global network aligment**
  - ✓ The objective is to avoid the ambiguity as in local alignment by drawing an injection between proteins in two different networks.
  - ✓ Global alignment of two networks was proven to be NP-hard
  - We introduce an ant based global network alignment algorithm called ACOGA. The experiments show that the method that we proposed get better results than the introduced methods recently.

Global alignment problem of PPI networks

- Denote two protein-protein interaction networks by G_1 =(V_1,E_1) and G_2 =(V_2,E_2)
  - V_1, V_2 indicate sets of nodes corresponding to proteins in the network G_1, G_2
  - E_1, E_2 indicate sets of edges corresponding to protein-protein interactions in G_1, G_2

Without losing the generality we can assume that $|V_1| < |V_2|$.

- Network alignment aims at finding an injection from V_1 into V_2 which is the best according to specific evaluation criteria.
- The graph A_{12} is considered as an alignment of two network if and only if:
  - i. Each node $<u_i, v_j> \in V_{12}$ corresponds a pair of nodes $u_i \in V_1$ and $v_j \in V_2$.
  - ii. Two distinct nodes $<u_i, v_j>$ and $<u_i', v_j'>$ of V_{12} imply $u_i \neq u_i'$ and $v_j \neq v_j'$.
  - iii. The edge $(<u_i, v_j>, <u_i', v_j'>)$ belong to E_{12} if and only if $(u_i, u_i') \in E_1$ and $(v_j, v_j') \in E_2$



- **Alignment quality measure**

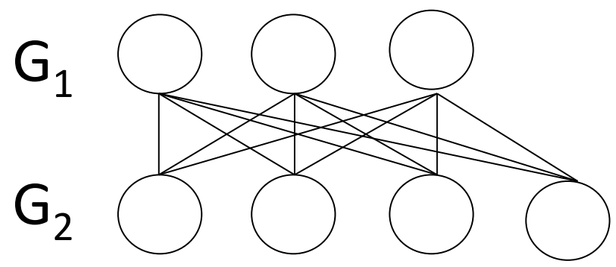
$$GNAS(A_{12}) = \alpha |E_{12}| + (1-\alpha) \sum_{\forall <u_i, v_j>} similar(u_i, v_j)$$

Proposed Method

- **ACOGA algorithm**

Input: Graph 1: $G_1 = (V_1, E_1)$; Graph 2: $G_2 = (V_2, E_2)$;
Similarities of node pairs: *Similar*
Balancing parameter α
Output:Alignment network $G_{12} = (V_{12}, E_{12})$
Begin
Initialize; // initialize pheromone trail matrix and m ants (A);
while (stop conditions not satisfied) do
for each ant a $\in$ **A do**
$V_{12} = \{< i, j >\}$ //The best similar pair $< i, j >$
for k=2 **to**$|V_1|$**do**
i =findMaxRelate (k $\in V_1 - V_{12}^1$)
j=antMove(i, $V_2 - V_{12}^2$)
$V_{12} = V_{12} \cup < i, j >$
Update(E_{12})
end-for
end-for
Update pheromone trail follows *SMMAS* rule;
Local search
Update the best solution;
End while
Save the best solution;
End.

- **Construction Graph**



- **Heuristic Information**

$$\eta_j^i = \alpha * M + (1-\alpha) * similar(i, j)$$

- **Raldom walk procedure**

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

- **Pheromone Update Rule**

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

vì i

$$\Delta_j^i = \begin{cases} \rho * \tau_{max} & < i, j > \in V_{12} \\ \rho * \tau_{min} & < i, j > \notin V_{12} \end{cases}$$

- **Local Search Procedure**

Input: Graph 1: G_1=(V_1,E_1) Graph 2: G_2=(V_2,E_2)
Alignment network A;
Output: Better Alignment network
Begin
Keep n_{best} pair <u,v> of V_{12}
For=n_{best}+1 to $|V_1|$ do
i= find_next_node();
j= choose_best_matched_node(i);
V_{12}=V_{12} ∪ <i,j>
Update(V_{12})
end-for
End

Results

- **Data**
  - ✓ Used 4 benchmark datasets that had been used to evaluate SPINAL performances.
  - ✓ They are datasets of protein-protein interactions on: Saccharomyces cerevisiae (sc), Drosophila melanogaster (dm), Caenorhabditis elegans(ce), and Homo sapiens (hs)

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

- **Test configuration**
  - ✓ The number of ants at each iteration is 6
  - ✓ ρ=0.3
  - ✓ τ_{max} = *1.0* and $\tau_{min} = \tau_{max}/(|V_1|+|V_2|)$
  - ✓ Local search procedure is appied with the best solution of ants of each iteration.
- ✓ **Comparison of the GNAS and $|E_{12}|$ score**

Datasets	α = 0.3		α = 0.4		α = 0.5	
	FASTAn	ACOGA	FASTAn	ACOGA	FASTAn	ACOGA
ce-dm	778.46 2560.7	**798.67** **2629.20**	1034.20 2564.6	**1057.34** **2622.9**	1290.11 2567.2	**1327.15** **2642**
ce-hs	863.46 2842.8	**885.47** **2916.1**	1144.17 2838.1	**1177.49** **2922.40**	1429.89 2844.9	**1461.74** **2909.1**
ce-sc	834.79 2761.1	**857.45** **2837.3**	1109.93 2761.2	**1144.56** **2849.4**	1389.21 2769.7	**1435** **2861.6**
dm-hs	2260.31 7478.3	***2315.78*** ***7663***	3007.11 7481.9	**3052.08** **7597**	3755.36 7429.0	**3803.79** **7584.3**
dm-sc	1977.82 6569.7	**2023.60** **6721**	2631.85 6565.5	**2653.53** **6619**	3290.03 6570.7	**3337.87** **6666.6**
hs-sc	2268.21 7531.8	**2300.31** **7640**	3017.96 7528.5	**3048.78** **7609.12**	3772.96 7535.2	***3838.3*** ***7666.0***

Datasets	α = 0.6		α = 0.7	
	FASTAn	ACOGA	FASTAn	ACOGA
ce-dm	1545.86 2567.7	**1601.13** **2660.4**	1801.24 2567.6	**1861.08** **2653.4**
ce-hs	1708.81 2838.0	**1758.37** **2921.1**	1994.87 2843.4	**2049.1** **2921**
ce-sc	1663.39 2766.5	**1688.11** **2808**	1936.83 2763.1	**1996.96** **2849**
dm-hs	4496.45 7478.2	**4574.12** **7607.8**	5242.32 7478.8	**5319** **7588.6**
dm-sc	3950.16 6577.4	**3989.68** **6643.30**	4603.41 6572.3	**4651.2** **6641.1**
hs-sc	4520.51 7527	**4640.28** **7726.90**	5279.88 7538.1	***5422.18*** ***7742***

Conclusion

- We proposed a novel algorithm called ACOGA for global alignment of two protein-protein interaction networks.
- Experimental results demonstrated the advancement and efficacy of the proposed algorithm in global alignment of protein-protein interaction network in terms of GNAS, EC criteria and running time as well.
- Finally, the procedure *Local search* depends on a critical parameter called $n\downarrow best$, which is a number of nodes with top scores in the previous alignment retained after each repetition. Getting the optimal value of this parameter automatically going to study in the future.
- The proposed algorithm can be parallelized to reduce running time