An Efficient Ant Based Algorithm For Global Alignment Of Protein-protein Interaction Networks
Trần Ngọc Hà1, Hoàng Xuân Huán2
1Thai nguyen University of Education, 2Vietnam 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/or orthologous 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 alignment
  - 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.

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 \( Similar \), \( Similar \) is a set of pairs \( \{ (u_1, v_1), (u_2, v_2), ... \} \) and \( \forall u_i \in V_1, v_i \in V_2 \), \( Similar \) is a function that returns the similarity between \( u_i \) and \( v_i \).

  Output: \( \text{Alignment network } G_{12} = (V_1 \cup V_2, E_{12}) \).

  Initializer: \( \text{Initialize } \tau_{ij} = \text{Similar}(u_i, v_j), \text{ for } \forall u_i \in V_1, v_j \in V_2 \).

  while (termination condition is not satisfied) do
    for each ant \( a \) do
      \( u_i \leftarrow \text{randomly select a node from } V_1 \) ; \( v_j \leftarrow \text{randomly select a node from } V_2 \)
      \( \text{The best similar pair } i < j \rightarrow \text{max} \text{Sim}(u_i, v_j) \)
      \( \text{Update } \tau_{ij} \).
    end-for
    \( \text{Update } \tau_{ij} \text{ follow SSMAS rule} \).
    \( \text{Local search} \).
    \( \text{Update the best solution} \).
    \( \text{Save the best solution} \).
  end-for
  

- Construction Graph

- Heuristic Information

  \[ \eta_j = \alpha M + (1 - \alpha) \text{Similar}(i, j) \]

- Randon walk procedure

  \[ p^t_j = \sum_{k \in \mathcal{V}_j} (\tau_k^j)^{\alpha} \eta_k^j \]

- Pheromone Update Rule

  \[ \tau_{ij}^t = (1 - \rho) \tau_{ij}^{t-1} + \Delta_{ij} \]

  \( \Delta_{ij} = \begin{cases} \rho \tau_{max} & \text{if } i, j \in \mathcal{V}_1 \\ \rho \tau_{max} & \text{if } i, j \notin \mathcal{V}_1 \\ 0 & \text{otherwise} \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 \( B \).

  Begin
    Keep \( \text{pair } u_i, v_j \) of \( V_{12} \)
    For each \( v_i \) do
      if \( \text{find } v_i \) do
        if \( \text{choose } v_i \) do
          \( \text{Update } V_i \leftarrow v_i \leftarrow u_j \).
          \( \text{end-for} \)
    \( \text{end-for} \)
  \( \text{End} \).

- Alignment quality measure

  \[ \text{GNAS}(A) = \alpha |E_1| + (1 - \alpha) \sum_{u_i v_j \in \mathcal{E}_{12}} \text{Similar}(u_i, v_j) \]

- Datasets

  \begin{array}{|c|c|c|}
  \hline
  \text{Dataset} & \text{No. of proteins} & \text{No. of interactions} \\
  \hline
  \text{ce} & 2805 & 4495 \\
  \text{dm} & 7518 & 25635 \\
  \text{sc} & 5499 & 31261 \\
  \text{hs} & 9633 & 34327 \\
  \hline
  \end{array}

  \begin{array}{|c|c|c|c|c|}
  \hline
  \text{Dataset} & \alpha = 0.3 & \alpha = 0.4 & \alpha = 0.5 \\
  \hline
  \text{ce-dm} & \text{FASTAn} & 776.48 & 1034.20 & 1290.10 \\
  & ACOGA & 1657.34 & 1327.15 & 2642.00 \\
  \text{ce-hs} & \text{FASTAn} & 883.46 & 1144.17 & 1429.89 \\
  & ACOGA & 2484.4 & 2369.44 & 2809.79 \\
  \text{ce-sc} & \text{FASTAn} & 634.79 & 1109.93 & 1369.21 \\
  & ACOGA & 2694.94 & 2765.78 & 2861.06 \\
  \text{dm-hs} & \text{FASTAn} & 2260.31 & 3052.08 & 3755.36 \\
  & ACOGA & 2764.56 & 3830.79 & 3783.69 \\
  \text{dm-sc} & \text{FASTAn} & 1977.82 & 3025.17 & 3290.03 \\
  & ACOGA & 3511.23 & 3337.87 & 3666.66 \\
  \text{hs-sc} & \text{FASTAn} & 2288.21 & 3048.78 & 3772.96 \\
  & ACOGA & 2763.1 & 3838.37 & 3783.83 \\
  \hline
  \end{array}

Results

- Test configuration
  - The number of ants at each iteration is \( \alpha = 0.3 \)
  - \( \tau_{min} = 1.0 \) and \( \tau_{max} = |V_1| + |V_2| \)
  - Local search procedure is applied with the best solution of ants of each iteration.

- Comparison of the GNAS and \( |E_1| \)

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 criterion and running time as well.
- Finally, the procedure Local search depends on a critical parameter called \( n_{best} \), 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.