Detection of statistically significant networks changes in complex biological networks Supplementary Information

June 14, 2016

1 Cosine similarity to calculate the edge weights

The one-step topological overlap measure used to estimate the edge weights is defined as:

$$a_{ij} = \frac{\sum_{l \neq i,j} A_{il} A_{lj} + A_{ij}}{\min(\sum_{l \neq i} A_{il} - A_{ij}, \sum_{l \neq j} A_{lj} - A_{ij}) + 1} \tag{1}$$

In this work we use the cosine similarity to calculate the edge weights a_{ij}. The cosine similarity takes into consideration one-step neighbourhood of nodes i and j while constructing the edge weight and is very efficient to calculate for sparse matrices. The weights a_{ij} are estimated as follows:

$$a_{ij} = \frac{\sum_l A_{il} A_{jl}}{\sqrt{\sum_l A_{il}^2} \sqrt{\sum_l A_{jl}^2}} \tag{2}$$

where A_{ij} represents the adjacency matrix.

We perform an experiment to calculate the correlation between the one-step topological measure and the cosine similarity measure. For this experiment, we generated 250 random geometric networks using $N = 250$ and the connectivity parameter $d = 0.15$. Figure 1 shows that the cosine similarity metric is nearly perfectly correlated (pearson correlaton = 0.952) to the topological overlap measure.

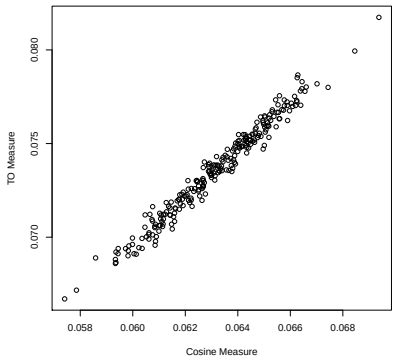


Figure 1: Correlation between topological overlap and cosine similarity.

2 Fast Approximation Algorithm

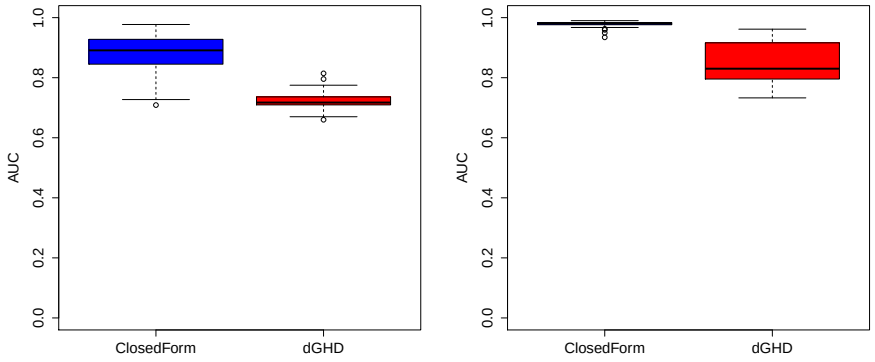
Algorithm 1: Fast Approximation

Data: Graphs A and B with N vertices V.
Result: Subset V^* representing the set of nodes which comprise the differential sub-network & p-values for GHD measure.

$V^* = \{\}$ `// Empty Set for differential sub-network nodes.`
$V_K = V$ `// Initialize a copy of the set of vertices` V.
$p^* = \{\}$ `// Empty Set for p-values.`
$\Delta_{V_K} = \{\}$ **forall the** $i \subset V_K$ **do**
  $t =$**GHD(**$A(V_{K|i}, E_A), B(V_{K|i}, E_A)) - \mu_{V_{K|i}}$.
  `// Estimate cGHD value after removal of node` i.
  Add t **to** Δ_{V_K}. `// Perform in parallel.`
Sort Δ_{V_K} **in descending order and keep in** $\mathcal{O}$.
while $N > 3$ **do**
  $z = \frac{\text{GHD}(A(V_K, E_A), B(V_K, E_B)) - \mu_{V_K}}{\sigma_{V_K}}$.
  Calculate p-value using z and append p-value to p^*.
  if *p-value* $> \theta$ **then**
    $\Delta_{V_K} = \{\}$ **forall the** $i \subset V_K$ **do**
      $t = (\text{GHD}(A(V_{K|i}, E_A), B(V_{K|i}, E_A)) - \mu_{V_{K|i}})$.
      Add t to Δ_{V_K} `// Perform in parallel.`
    $n^* = \max_i \Delta_{V_K}$
    `// Select that node after removal of which cGHD becomes maxmimum.`
    Remove node n^* from V_K and $\mathcal{O}$
  else if *p-value* $< \theta$ **then**
    $n^* =$ **max**$_i(\mathcal{O})$ `// Select node in the sub-network with least contribution.`
    Remove node n^* from $\mathcal{O}$.
  if *p-value* > 0.01 **then**
    Append n^* to V^*.
  $N = N - 1$
Adjust the p-values for false-discovery rate.

3 Boxplots of AUC

We make box-plots for the AUC metric and observe from Figure 2a that the dGHD method has lower variance w.r.t. to the evaluation metric in comparison to Closed-Form approach in case of permuted differential sub-network. However, in case of denser differential sub-network, the Closed-Form approach has smaller variance in comparison to dGHD algorithm w.r.t. AUC metric as depicted in Figure 2b. From Figure 2 we can conclude that the performance of Closed-Form technique is significantly better than dGHD method when differential sub-networks are formed either using permuted nodes or higher density. In order to test for significance we performed the Student's t-test under the null that the difference in the mean values of the two distributions is zero i.e. $\mu_{AUC_A} - \mu_{AUC_B} = 0$. At significance level of 5%, we obtain p-value of 3.16×10^{-9} in case of permuted sub-network, thereby rejecting the null. In the case of paired networks with a denser differential sub-network (i.e. $d' = 0.5$), we obtain p-value of 3.14×10^{-14} for the Student's t-test.



(a) Permuted sub-network ($d = 0.3$) AUC (b) Dense sub-network ($d = 0.3, d' = 0.5$) AUC

Figure 2: Comparison of proposed Closed-Form approach and dGHD method w.r.t. AUC metric.