Signed vs. Unsigned Topological Overlap Matrix
Technical report

Peter Langfelder*
*Peter.Langfelder@gmail.com

December 5, 2013

Abstract

I describe the "signed" Topological Overlap Matrix introduced in the WGCNA R package.

1 Review of TOM and signed vs. unsigned networks

The Topological Overlap Measure (TOM) was introduced in [1] to make networks less sensitive to spurious connections or to connections missing due to random noise. While the original work [1] considered unweighted networks, the authors of [2] generalized TOM to a weighted network.

The central idea of TOM is to count the direct connection strengths as well as connection strengths "mediated" by shared neighbors. The standard, or "unsigned" TOM assumes that neighbor-mediated connections can always be considered as "reinforcing" the direct connection. This may not always be the case, and the *signed* TOM is an attempt to take this into account.

Before explaining the last point, I briefly review the notion of *signed* and *unsigned* correlation network (adjacency or connection strength). In a *signed* correlation network, nodes with negative correlation are considered unconnected (their connection strength is zero or very close to zero). In contrast, in *unsigned* correlation networks, nodes with strong negative correlations have high connection strengths: the unsigned network adjacency is based on the absolute value of correlation, so positive and negative correlations are treated equally.

2 Unsigned networks may contain triplets with anti-reinforcing correlations

Because unsigned networks lose information about the sign of the underlying correlations, it is possible to find triplets of nodes (i.e., groups of 3 nodes) such that the 3 correlations between the node pairs have signs $(+, +, -)$ or $(-, -, -)$, that is, one or all three of the correlations are negative. Denote the nodes x_i, x_j, x_k and consider the topological overlap measure of nodes i and j, TOM_{ij}. The idea of TOM is to "reinforce" the direct connection between the nodes i and j by the connection mediated via the node k. However, in this case, the underlying correlations do not appear to be reinforcing: while the correlation of nodes i and j is positive, the positive correlation between i and k and the negative correlation between j and k together suggest that the correlation of nodes i and j should be negative. In this sense, triplets of nodes with correlation signs $(+,+,-)$ or $(-,-,-)$ are not reinforcing; they suggest that at least some of the three correlations are affected by noise and I call them *anti-reinforcing*. It is important to note that such triplets can only occur in *unsigned* networks - in a signed network, the connection strength of nodes with negative correlation is zero or very close to zero, so this situation cannot arise.

3 Taking into account anti-reinforcing correlations: signed TOM

The standard or "unsigned" TOM does not distinguish anti-reinforcing connections from the reinforcing ones and counts all as reinforcing. In contrast the **"signed TOM"** subtracts anti-reinforcing mediated connections from the direct (and the reinforcing) connections. To be able to do this, signed TOM needs as input not only the connection strengths a_{ij}, but also the sign of the underlying correlations. This can be onveniently achieved by defining a modified adjacency matrix $\tilde{a}_{ij}$:

$$\tilde{a}_{ij} = a_{ij} \times \text{sign}\left(\text{cor}(x_i, x_j)\right) . \tag{1}$$

The *signed TOM* is then defined as

$$TOM_{ij}^{signed} = \frac{\left| a_{ij} + \sum_{u \neq i,j} \tilde{a}_{iu} \tilde{a}_{uj} \right|}{\min(k_i, k_j) + 1 - |a_{ij}|} , \tag{2}$$

where k_i and k_j denote the connectivities of nodes i and j:

$$k_i = \sum_{u \neq i} |\tilde{a}_{ui}| . \tag{3}$$

In contrast, *unsigned TOM* is defined as follows (note the difference in the placement of absolute values in the numerator):

$$TOM_{ij} = \frac{|a_{ij}| + \sum_{u \neq i,j} |\tilde{a}_{iu} \tilde{a}_{uj}|}{\min(k_i, k_j) + 1 - |a_{ij}|} . \tag{4}$$

The minimum in the denominator of Eqs. 2 and 4 can optionally be replaced by mean which improves the behaviour of TOM in certain special situations.

The take-home message from these notes is this: *signed TOM* takes into account possible anti-reinforcing connection strengths that may occur in *unsigned networks*. Since the anti-reinforcing connection strengths (practically) cannot occur in *signed networks*, in *signed networks* the signed and unsigned TOM are (practically) identical.

References

[1] E Ravasz, A L Somera, D A Mongru, Z N Oltvai, and A L Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–5, 2002.

[2] Bin Zhang and Steve Horvath. General framework for weighted gene coexpression analysis. *Statistical Applications in Genetics and Molecular Biology.*, 4(17), 2005.