Supplementary Material For: Global Network Alignment In The Context Of Aging

Fazle Elahi Faisal, Han Zhao, and Tijana Milenković

S1 Supplementary Methods

S1.1 Topological properties of our predictions

As explained in Section 3.5.2 in the main paper, we analyze the topological position of a node in the network with respect to *seven* node centrality measures as follows.

Degree centrality (DEGC) measures the degree of a node in the network, i.e., the number of the node's neighbors. The higher the degree of a node, the more central the node according to DEGC.

Clustering coefficient centrality (CLUSC) measures, for a given node, how many pairs of neighbors of the node are connected by an edge, out of all pairs of the node's neighbors. Intuitively, the more interconnected the neighborhood of the node, the more central the node is according to CLUSC.

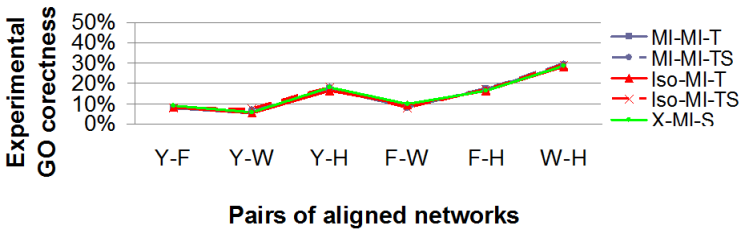
K-core of a network is a maximal subset of nodes in the network such that each node is connected to at least k others in the subset. *K-coreness centrality* (KC) of a node is k if the node is in k-core.

Graphlet degree centrality (GDC) measures how many graphlets a node participates in, for 2-5-node graphlets [1]. Intuitively, the more graphlets a node touches, the more central the node is according to GDC. Since it captures the *extended* network neighborhood of a node, GDC is a sensitive measure.

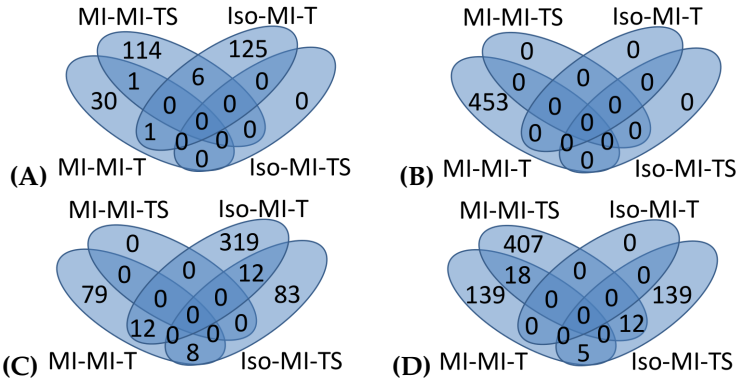
Betweenness centrality (BETWC) measures the involvement of a node in the shortest paths in the network. Intuitively, nodes that occur in many shortest paths have high centrality according to BETWC. BETWC of node v, $C_{\text{betwc}}(v)$, is: $C_{\text{betwc}}(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where V is the set of nodes in the network, σ_{st} is the number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ is the number of shortest paths between s and t that go through v.

Closeness centrality (CLOSEC) measures the "closeness" of a node to all other nodes in the network. Intuitively, nodes with small shortest path distances to all other nodes have high centrality according to CLOSEC. CLOSEC of node v, $C_{\text{closec}}(v)$, is: $C_{\text{closec}}(v) = \frac{1}{\sum_{u \in V} \sigma(u,v)}$, where $\sigma(u,v)$ is the shortest path distance between nodes u and v. In a PPI network, CLOSEC of a protein indicates the "likelihood" of the protein to reach or be reachable from all other proteins [2].

Eccentricity centrality (ECC) is very related to CLOSEC, except that it measures the "closeness" of a node *only* to the *farthest* node in the network [3]. Intuitively, nodes with small shortest path distances to the furthest node in the network have high centrality according to ECC. ECC of node v, $C_{\text{ecc}}(v)$, is: $C_{\text{ecc}}(v) = \frac{1}{\max_{u \in V}\{\sigma(u,v)\}}$.



Supplementary Fig. S1. Biological alignment quality of different cost functions under MI-GRAAL's alignment strategy. Experimental GO correctness of MI-MI-T, MI-MI-TS, Iso-MI-T, Iso-MI-TS, and X-MI-S is shown when they are used on each pair of networks of yeast (Y), fly (F), worm (W), and human (H). In all figure, the higher the values, the better the alignment quality.



Supplementary Fig. S2. Overlap of new predictions from different aligners under MI-GRAAL's alignment strategy in **(A)** yeast, **(B)** fly, **(C)** worm, and **(D)** human.

Supplementary Table S1

Statistical significance (in terms of p-values) of the difference between centrality values of Complement and centrality values of each of the following: 1) Novel-All, 2) ExpressionAge, 3) DyNetAge, and 4) GenAge, for each centrality measure (BEWTC, CLOSEC, CLUSC, DEGC, ECC, GDC, and KC).

	Data set	BETWC	CLOSEC	CLUSC	DEGC	ECC	GDC	KC
1	Novel-All	3.5×10^{-19}	1.5×10^{-13}	5.2×10^{-14}	4.6×10^{-24}	4.1×10^{-06}	1.8×10^{-17}	2.1×10^{-21}
2	ExpressionAge	8.8×10^{-3}	0.03	3.2×10^{-3}	5.5×10^{-4}	0.11	8.7×10^{-3}	8.9×10^{-4}
3	DyNetAge	1.8×10^{-11}	1.4×10^{-5}	1.7×10^{-4}	1.4×10^{-11}	1.3×10^{-3}	2.4×10^{-7}	2×10^{-10}
4	GenAge	9.1×10^{-24}	9.9×10^{-15}	3.3×10^{-15}	8×10^{-26}	1.7×10^{-10}	1.1×10^{-18}	6.3×10^{-20}

Supplementary Table S2

The number of GO terms ("N") enriched in Novel-All, ExpressionAge, DyNetAge, GenAge, and Complement, according to any evidence code ("any") or experimental evidence codes only ("exp").

Data set	N (any)	N (exp)
Novel-All	105	26
ExpressionAge	190	37
DyNetAge	153	28
GenAge	768	191
Complement	8	0

REFERENCES

[1] T. Milenković, V. Memišević, A. Bonato, and N. Pržulj, "Dominating biological networks," *PLOS ONE*, vol. 6, no. 8, p. e23016, 2011.
[2] G. Scardoni, M. Petterlini, and C. Laudanna, "Analyzing biological network parameters with centiscape," *Bioinformatics*, vol. 25, no. 21, pp. 2857–2859, 2009.
[3] S. Wuchty and P. F. Stadler, "Centers of complex networks," *Journal of Theoretical Biology*, vol. 223, no. 1, pp. 45–53, 2003.