Classifying Aging- and Non-Aging-Related Genes in a Dynamic Protein-Protein Interaction (PPI) Network

Jon Genty, Khalique Newaz, Tijana Milenkovic
Department of Computer Science and Engineering, University of Notre Dame

Introduction

Network science studies relationships between objects, representing each object as a node and a relationship between two nodes as an edge.

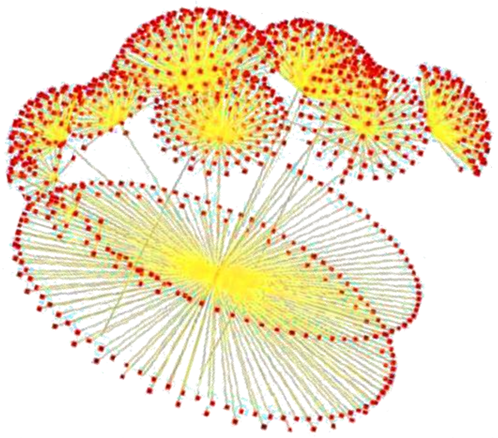

Figure 1: Visualization of a protein-protein interaction (PPI) network. Image taken from - https://www3.nd.edu/~cone/2.png

In a PPI network, displayed in Figure 1, nodes represent proteins (gene products) and edges represent bindings between proteins.
Specifically, we are interested in studying the human PPI network and changes in its structure and thus cellular functioning with age. This is important, because the incidence of serious diseases often increases with age. Therefore, network science can serve to predict and potentially identify aging related targets for treatment.

Condition-Specific Data

The current PPI network of humans is static, as it spans many different biological contexts.
However, we can introduce context-specific information by integrating the static PPI network with dynamic, aging-related gene expression data.

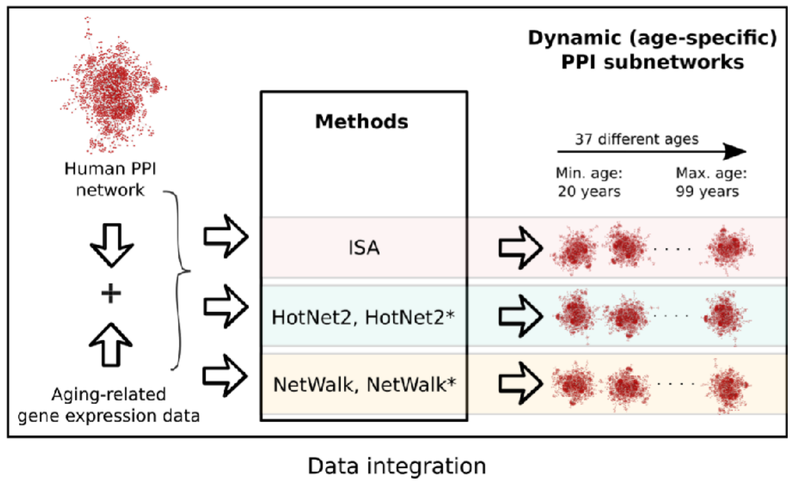

Figure 2: Method used to produce dynamic PPI networks from static PPI networks and gene expression data. Image taken from - https://arxiv.org/abs/1807.05637

Static to Dynamic Networks

We can assign age-specific weights, derived from gene expression data, to nodes or edges and then declare the highest-weighted network regions as active at the given age. The active network regions then correspond to the part of the whole network that is active at the age in question.

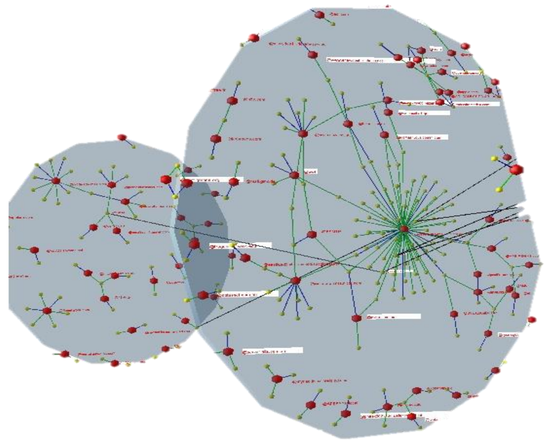

Figure 3: A dynamic network changing with time. Image taken from - http://www.it.usyd.edu.au/~shhong/temporal.html

We use five existing methods for data integration to do this, which results in five corresponding dynamic, aging-related PPI networks of human: Induced, HotNet2, HotNet2*, NetWalk and NetWalk*.

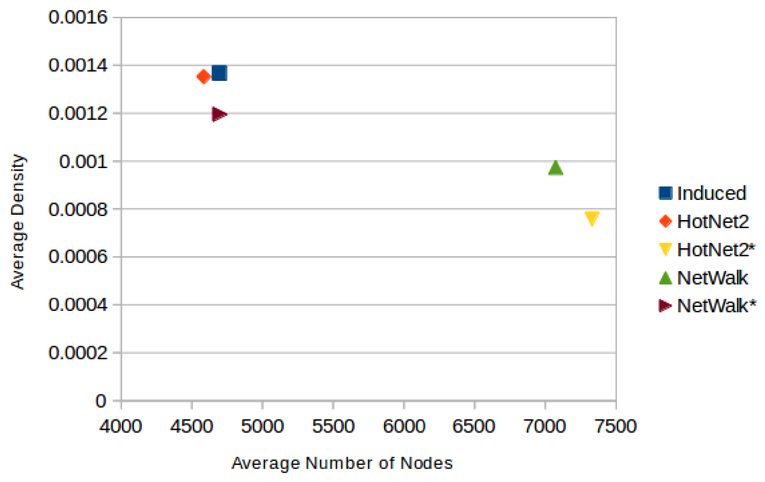

Figure 4: Properties of the five dynamic networks that were produced using gene expression data.

Centralities

We calculate network positions (centralities) of every node at every age.
Previously, nodes whose centralities significantly correlate (i.e., increase or decrease) with age are predicted to be aging-related.
However, nodes could express other trends. A node's centrality may periodically increase and decrease with age, and this is a behavior that we would like to capture.

Coefficient of Variance

We evaluate whether one of the two gene groups shows higher or lower variation (fluctuation of a node's centrality overtime) than the other group. This is measured using a coefficient of variance, described by Equation 5.

$$CV = \frac{\sigma}{\mu}, \quad \sigma = standard\ deviation$$
$$\mu = mean$$

Equation 5: Coefficient of variance formula

Results & Discussion

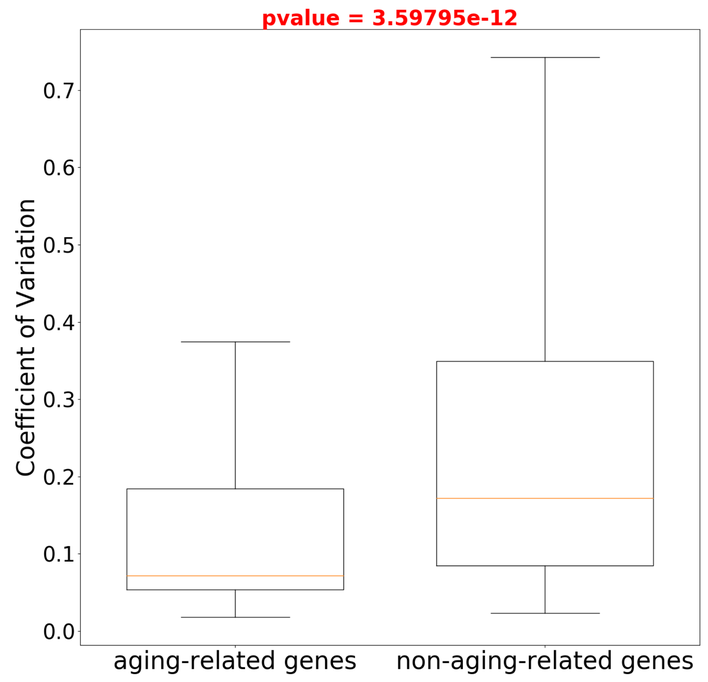

Figure 6:Boxplot of aging- and non-aging-related genes, with respect to their corresponding coefficients of variance.

As Figure 6 indicates, aging-related genes fluctuate less than non-aging-related genes.
Additionally, aging-related genes have higher centrality values (averaged over all ages) than non-aging-related genes, matching findings obtained from existing static network analyses of aging.
We will continue to evaluate the fluctuation of a gene's centrality as a method of predicting aging-related genes, based on these initial results.

References

K. Newaz and T. Milenkovic, "Improving inference of the dynamic biological network underlying aging via network propagation", 2018. Available: https://arxiv.org/abs/1807.05637
Fazle E. Faisal and Tijana Milenkovic (2014), "Dynamic networks reveal key players in aging", *Bioinformatics*, 30(12): 1721-1729.
This work was supported by National Science Foundation grant IIS-1560363: "REU Site: Data Intensive Scientific Computing"