Thomas Young, Christoph Uhl
Aaron Clauset
CSCI 3352

Preferential Attachment in Duplication Mutation

Background Material

There are two parts to our project. The first part involves looking at duplication mutation of proteins in gene networks. Duplication mutation is the process of copying existing sets of proteins along with their connections, then disconnecting and reconnecting them to pre-existing nodes with probability δ and α [1]. The second part of the project involves using various models of network formation to determine the features that guide the formation of duplication mutation networks. It is well believed in the network science community that scale-free networks are formed via (i) expanding “continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected” [2]. The duplication mutation networks mentioned before are characterized by a scale-free model.

Research Question

For this project we want to determine if preferential attachment is the only process at play in the creation of gene networks or if there are other processes guiding new node attachment. From this we have come to the question: Are the underlying network structures of duplication mutation networks caused by preferential attachment?

Anticipated Findings

We anticipate that preferential attachment will not be the only method of node attachment which accurately models gene network formation.

Description of Data and Algorithms

We will be implementing 3 methods for new node attachment: preferential attachment and two other methods based on different procedures, i.e. triadic closure, uniform attachment, age of node, or clustering coefficient [3]. Along with testing different forms of attachment we will create models of gene networks using the steps listed in the background knowledge section. Lastly, to determine if and how the modeled networks differ we are going to plot the rate of attachment over node degree. Then, by comparing the attachment rate plots from each network, we can determine if they have similar structural formation.

Citations

[1] Ricard V. Sol, RomualdoPastor-Satorras, Eric Smith, and Thomas B. Kepler,” A model of large-scale proteome evolution”, Advances in Complex Systems 05 (2002), no. 01, 43–54.

[2] Albert-Laszlo Barabasi, and Reka Albert. "Emergence of Scaling in Random Networks." Science 286.5439 (1999): 509-12. ProQuest. 8 Oct. 2019 .

[3] Jan Overgoor, Austin R. Benson, and Johan Ugander. “Choosing to grow a graph: Modeling network formation as discrete choice”, 2018.