BST281: Genomic Data Manipulation, Spring 2017

Monday 03: Introduction to Biological Networks

This lecture introduces basic concepts of network description and analysis, especially as applied to molecular biological networks, their analysis and reconstruction.

## **(Biological) networks**

## A network is collection of “things” (represented as nodes) and their relationships/interactions (represented as edges). In biology, networks enrich high-throughput molecular “parts lists” to include interactions among the parts. This approach can lead to a better understanding of biological mechanisms and systems.

## Network examples

Some biological networks represent physical interactions among molecules, such as *protein-protein physical interaction* (PPI) networks or *transcription factor (TF)-target gene (TG) binding* networks. Other networks are more abstract, such as *genetic interaction* networks, which capture epistatic relationships between perturbed genes.

*Node properties*

Network nodes are characterized by their “degree,” the number of edges in which they participate. Nodes that participate in many shortest paths are said to have high “betweenness.”

*Edge properties*

Network edges can be *directed* (arrows) or *undirected* (segments) depending on the type of action/relationship they capture. Edges can be additionally signed (activating/repressing) or weighted.

*Network properties*

Many real-world networks exhibit “scale-free” behavior: characterized by many low-degree nodes and a small number of high-degree nodes. Similarly, many real-world networks exhibit “small-world” behavior, where the longest shortest path is objectively small (see: the “Kevin Bacon game”).

*Identifying biological interactions*

PPIs can be identified by 1) affinity capture, in which a target protein (bait) is immunoprecipitated in complex with its binding partners (prey); or 2) two-hybrid approaches, where modified proteins produce measurable readouts (e.g. change in gene expression) upon interaction. TF-TG relationships are also determined by affinity capture-like methods, with bound DNA serving as the prey. Genetic interactions are identified by a number of experimental techniques (e.g. double gene knock-outs).

*Sources of biological interaction data*

BioGRID and EBI’s IntAct are among the largest sources of biological interaction data, especially for PPIs. Specialized databases are often required for other interaction types. Interaction data are represented as lists of interacting gene/protein pairs (often decorated with other useful information, e.g. literature citation).

# Suggested textbook reading

* Pevsner, Chapter 14, p670-685

# Related literature

* [Kim, Philip M., et al. "Relating three-dimensional structures to protein networks provides evolutionary insights." Science 314.5807 (2006): 1938-1941.](http://science.sciencemag.org/content/314/5807/1938)
* [Rolland, Thomas, et al. "A proteome-scale map of the human interactome network." Cell 159.5 (2014): 1212-1226.](http://www.sciencedirect.com/science/article/pii/S0092867414014226)