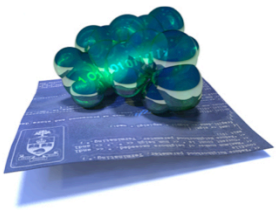
A
Bioinformatics
Course

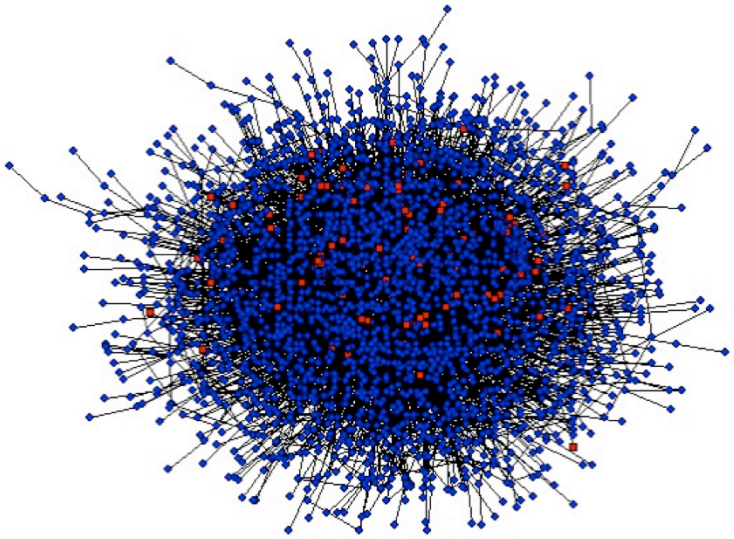
Analysis of PPI Networks



Boris Steipe

department of biochemistry – department of molecular genetics
University of Toronto

Interaction Hairball†



† also called: "Giant Component"

Human Prote in Interaction Network: This is still not the way the cell works. Biological interactions are tissue–, compartment–, *time*–, or otherwise context dependent.

Plotting genome-scale interaction data as networks usually results in one single, lareg component. This obscures the detailed systems of interactions that make up the function of the cell.

CONSTITUVE / OBLIGATE INTERACTIONS: not dynamic; “party hubs”; e.g. components of multi-protein complexes, structural interactions.

TRANSIENT INTERACTIONS: *reactive* or *programmed*; “date hubs”; usually involving post-translational modifications (but also regulated expression, differential splicing *etc.*)

Interactions are modulated over time but defining an appropriate reference time point for the purpose of comparing and categorizing interactions may be challenging. Time along the cell cycle? Time after onset of an external stimulus? Time after specific expression signal? Moreover, the “clock” of timed events may run at different speeds, depending on general aspects of the cell’s current state (Stress? Hunger? Resting? Replicating?).

Time resolved interaction studies are technically challenging, and the existence of data- and party- hubs has been a controversial topic in the literature.

Time Resolved Detection:
synchronization (or single-cell methods);
perturbation.

Issue: many(!) datapoints needed for statistical power.

Computational Analysis:
(i) clustering
(ii) detection of causal relationships:
lagged correlation ... (mere correlation or causation?).
(iii) significance and relevance.

Once a satisfactory network has been computed, we can apply standard tools of network analysis:

Degree distribution

Defines global properties of the network, possibly can be interpreted in terms of a generative process.

Topological Analysis

Identifies critical pathways (centrality measures)

Community Analysis

Identifies clusters, systems, complexes ...

Once a satisfactory network has been computed, we can annotate its nodes:

Node properties

In PPI networks all manners of protein annotations including name, pathway function, GO terms etc can help to identify a protein with a role that is suggested by its postion in the network.

Neighborhood and community properties

Neighborhoods of nodes, and communities may have particular properties enriched, givng an indication of their functional relationship.

Annotation transfer

Communities may have un-annotated nodes that can assume annotations from other nodes in their neighborhood or community via “label propagation” or “diffusion”.

http://steipe.biochemistry.utoronto.ca/abc

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA