**Subnetwork hierarchies of biochemical pathways**

**T. pallidum**

The hierarchical clustering tree of *T. pallidum* shows the metabolic network of *T. pallidum*, the bacterium causing primary syphilis. The tree is divided into three main clusters: the first cluster (red) represents the pyruvate metabolism, the second cluster (green) the lipoamide metabolism, and the third cluster (blue) the nucleoside metabolism. Each node in the tree represents a subnetwork, and the height of the node corresponds to the number of shortest paths between pairs of substances passing through it.

**Motivation**

The richness and complexity of the biochemical networks that have been mapped out by modern genomics calls for decomposition into subnetworks.

Such networks can have inherent non-local features that require the global structure to be taken into account in the decomposition procedure.

Basic questions such as what define the network (graph theoretically) can and should be handled by schemes that subnetworks are little studied.

**The networks**

The networks we use are 43 organisms from the WIT database (aarch, 32 bacteria, and 9 eukaryotes). We represent the data as a directed bipartite graph G = (A,E), where A is the set of substances, E is the set of reactions (an important property of complex and it is links from elements of A to elements of R or vice versa.

**The decomposition algorithm**

The decomposition algorithm starts from the full network and decomposes it by recursively deleting reaction nodes of high betweenness centrality. $C_B(C)/B$ of a reaction node $i$ is the number shortest paths between pairs of substances passing through $i$.

The decomposition algorithm consists of the following steps repeated until no reaction nodes remain:

1. Calculate the effective betweenness for all reaction nodes.
2. Remove the reaction node with highest betweenness and all its in- and out-going links.
3. See information about the current state of the network.
4. How many clusters there are, and what nodes that belongs to a specific cluster.

**Schematic picture of the two different orderings**

We consider two small subnetworks. Our method of ordering is based on the hierarchical organization of a network. For comparing the hierarchical decomposition trees of subnetworks, we compute the number shortest paths between pairs of substances passing through a reaction node.

**Statistics of the hierarchical clustering**

For comparing the hierarchical clustering trees of different organisms we measure the relative half-height $h_{rel}$, the height where the largest cluster is half of its original size divided by the usual height of the tree and also the maximal size of the second largest cluster relative to the original. We can conclude that $h_{rel}$ has a close to universal value of 0.79 for metabolic, and 0.76 for whole cell networks. A high $h_{rel}$ implies a very robust network, and these values are indeed high (corresponding values for social networks are of the order 0.2).

**Conclusions**

We propose an algorithm to decompose biochemical networks into subnetworks based on the global network structure (a development of an algorithm of Girvan and Newman, PNAS 2002).

With the algorithm we avoid strong categorization of e.g. degree-one substances.

We emphasise the use of hierarchy trees to get a categorisation of e.g. degree-one substances.

The function of two extreme types of hierarchical ordering: community- and shell-type ordering are discussed.

**General shape of the trees**

Schematic picture of the two different orderings in hierarchy trees. (a) Community-type ordering, all same level core-clusters connected by outer parts of the network. (b) Shell-type—a sequence of outer clusters contained in each other. The squares symbolises the reaction nodes that is deleted at the height marked by the arrow. In (b), three subnetworks of similar sizes get disconnected when the reaction node is removed. In (a), many individual metabolic nodes (squares) get isolated.

**Detected subnetworks**

Subnetworks of the whole cellular networks are more functionally distinct than the metabolic-only networks. One of the subnetworks is a part of the bacterial photosynthesis system, the function of which is to import carbohydrates into the cell. Each of these enzymes is specific for a certain kind of carbohydrate; in (a), we see enzymes specific for mannitol, glucose, sucrose and fructose, respectively. The other network (c) has to do with DNA replication. The DNA replication subnetwork (c) is controlled around a reaction node with high degree (local centrality), but relatively low betweenness (global centrality). Thus local, degree-based, algorithm would have difficulties identifying such a subnetwork.

**Uncle cell network of H. pneumoniae**

For comparing the hierarchical clustering trees of different organisms we measure the relative half-height $h_{rel}$, the height where the largest cluster is half of its original size divided by the usual height of the tree and also the maximal size of the second largest cluster relative to the original. We can conclude that $h_{rel}$ has a close to universal value of 0.79 for metabolic, and 0.76 for whole cell networks. A high $h_{rel}$ implies a very robust network, and these values are indeed high (corresponding values for social networks are of the order 0.2).

**Online:**

This poster: http://www.tp.umu.se/forskning/networks/meta/poster.pdf

**Contact:**

Mikael Huss

Petter Holme

Hauong Jeong