Subnetwork hierarchies of biochemical pathways

Petter Holme
Mikael Huss
Hawoong Jeong

Umeå Universitet
KTH
KAIST

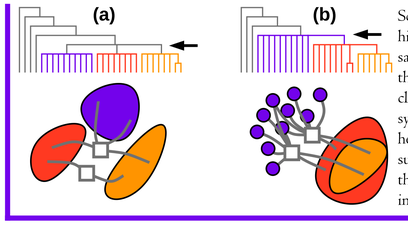
Motivation

- The vastness 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 to what extent the network (graph theoretically) can be said to be built by distinct subnetworks are little studied.

The networks

The networks we use are 43 organisms from the WIT database (6 archea, 32 bacteria and 5 eukaryotes). We represent the data as a directed bipartite graph $G = (S,R,L)$, where S is the set of substrates, R is the set of reactions (or temporary reaction complexes) and L is links from elements of S to elements of R, or vice versa.

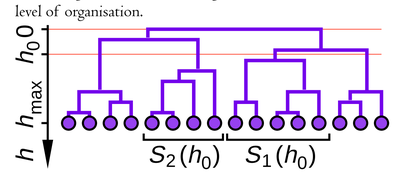
General shape of the trees



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

Hierarchy trees

Hierarchical clustering trees is meant to give an overview of the modular organisation of a network: Nodes (substrates in metabolical networks) connected below a specific horizontal line at height h of the tree belongs to the same subnetwork at that level of organisation.



The decomposition algorithm

The decomposition algorithm starts from the full network and deconstructs it by successively deleting reaction nodes of high betweenness centrality C_B. ($C_B(r)$ of a reaction node r is the number shortest paths between pairs of substrates passing through r.)

Since all in-neighbours of a reaction node needs to be present for a reaction to occur, we rescale the betweenness by the reaction node's in-degree to obtain an "effective betweenness."

The algorithm consists of the following steps iterated until no reaction nodes remains.

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. Save information about the current state of the network (how many clusters there are, and what nodes that belongs to a specific cluster).

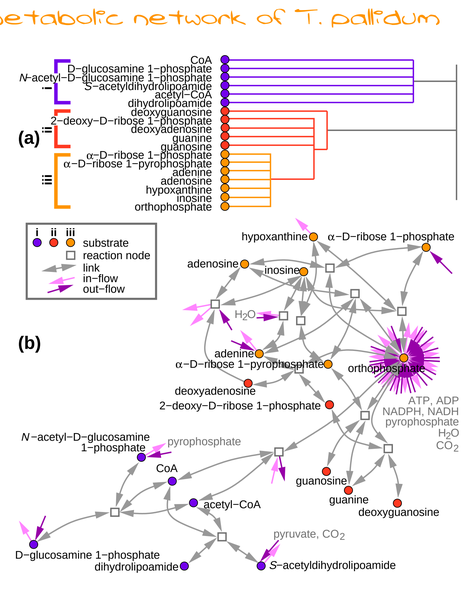
Detected subnetworks

As examples of the detected subnetworks we consider the metabolic network of *T. pallidum* (the pathological agent of syphilis) and the whole cellular network of *M. pneumoniae* (causing respiratory tract infections).

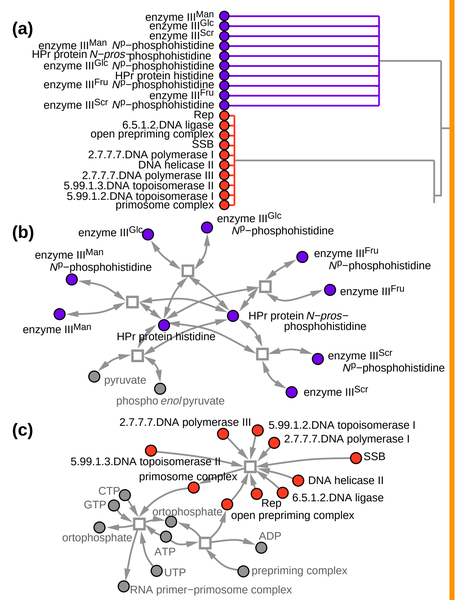
We consider two small subnetworks at $h = 40$ of the hierarchy tree of *T. pallidum*'s metabolic network. These subnetworks contains reactions associated with purine metabolism and pyruvate/acetyl-CoA conversion. The pyruvateacetyl-CoA part is a tightly interconnected, fairly independent subnetwork, while the purine metabolism part consists of an outer shell (ii) encapsulating a smaller core (iii), which is centred around orthophosphate and has to do with interconversion between adenosine and related nucleosides. Deoxyadenosine ends up in the outer shell (ii) because there are two reactions involving adenosine and orthophosphate, but only one with deoxy-adenosine and orthophos-phate.

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 phospho-transferase 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 (b), 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 centred around a reaction node with high degree (local centrality), but relatively low betweenness (global centrality). Thus local, degree-based, algorithms would have difficulties identifying such a subnetwork.

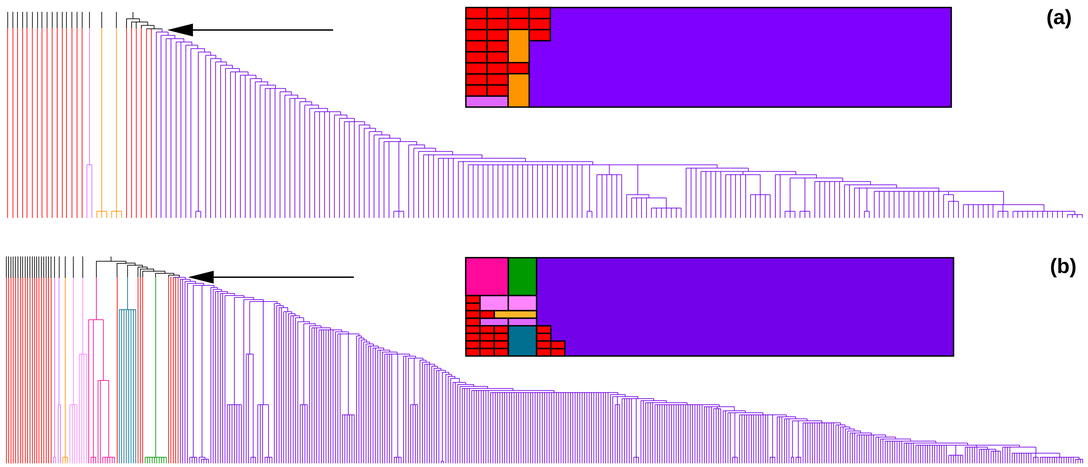
metabolic network of T. pallidum



whole cell network of M. pneumoniae



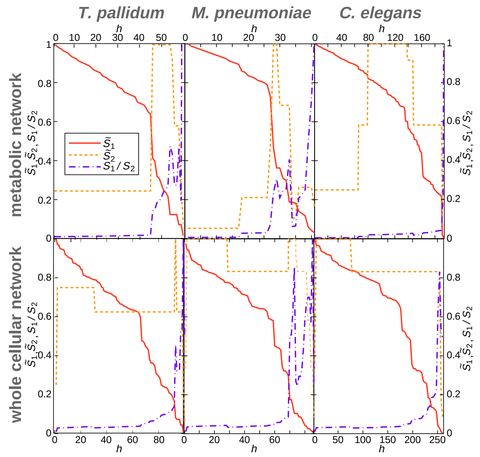
T. pallidum (an example)



The hierarchical clustering trees of *T. Pallidum*. (a) shows the tree for the metabolic network, (b) shows the whole cellular network. The squares represents the subnetwork configuration at $h = 0.1\ h_{\max}$ (the height indicated by the arrow). Sizes of the squares are proportional to the size of the clusters they represent.

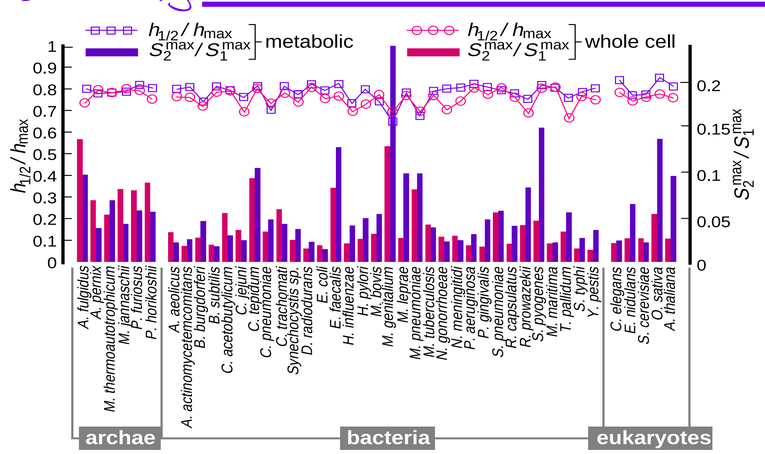
The trees are dominated by a few core clusters centred around the most connected substances (to the right). Outer substances are forms shells of weaker and weaker connections to the core, with a few well-defined subnetworks as exceptions (especially in (b)).

Statistics of the hierarchical clustering



To get a quantitative picture of the hierarchical ordering displayed through the hierarchical clustering trees we measure the size of the largest (S_1) and second largest (S_2) components; and also the S_1/S_2 ratio. At the peak of S_1/S_2 well defined subnetwork centred around the most highly connected substances exists.

For comparing the hierarchical clustering trees of different substances we measure the relative half-height $h_{1/2}$, the height where the largest cluster is half of its original size divided by the total height of the tree; and also the maximal size of the second largest cluster relative to the original S_1. One can conclude that $h_{1/2}$ has a close to universal value of 0.79 for metabolic, and 0.76 for whole-cell networks. A high $h_{1/2}$ implies a very robust network, and these values are indeed high [corresponding values for social networks are of the order $O(0.2)$].



Conclusions

- We propose an algorithm to deconstruct biochemical networks into subnetworks based on the global network structure (a development of an algorithm of Girvan and Newman, PNAS 2002).
- With this algorithm we avoid strange categorisation of e.g. degree-one substances.
- We emphasise the use of hierarchy trees to get a picture of the hierarchical organisation of subnetworks of subnetworks.
- The function of two extreme types of hierarchical ordering: community- and shell type ordering are discussed.
- The overall shape of the hierarchy trees are quantitatively studied through the relative half-height and ratio between largest and second largest connected components.
- The relative half-height is close to constant for all 43 organisms we study.

online:
Article: http://xxx.lanl.gov/abs/cond-mat/0206292
Project homepage: http://www.tp.umu.se/forskning/networks/meta/
This poster: http://www.tp.umu.se/forskning/networks/meta/poster.pdf

contact:
Petter Holme
Department of theoretical physics
Umeå University
901 87 Umeå, Sweden
tel. +46-90-786 7760
fax. +46-90-7869556
email: holme@tp.umu.se

Mikael Huss
SANS, NADA
Royal Institute of Technology
100 44, Stockholm, Sweden
email: hussm@kth.se

Hawoong Jeong
Department of physics
KAIST
Taejon 305-701, Korea
email: hjeong@nd.edu