Evidence for dynamically organized modularity in the yeast protein-protein interaction network

Sari Bombino

Helsinki 27.3.2007

UNIVERSITY OF HELSINKI

Department of Computer Science

Seminar on Computational Systems Biology

Seminar report

Abstract

Protein-protein interactions refer to the association of protein molecules and the study of these associations from the perspective of biochemistry, signal transduction and networks. In *scale-free protein-protein interaction networks* ('interactome' networks) most proteins interact with few partners. A small but significant proportion of proteins, the hubs, interact with many partners [HBH04]. In general, scale-free networks are tolerant to random node removal but they are very sensitive to the targeted removal of hubs. Knockouts of yeast genes encoding hubs are approximately threefold more likely to cause lethality than knockouts of non-hubs. This indicates that there seems to exist a connection between genetic robustness and the scale-free topology of protein-protein interaction networks.

Han *et al* investigated how the hubs might contribute to robustness and other cellular properties for protein-protein interactions dynamically regulated in time and space. They discovered two types of hubs: 'party' hubs and 'date' hubs. Party hubs interact with most of their partners at the same time, while date hubs bind their different partners at different times or locations. Both *in silico* studies of network connectivity and genetic interactions described *in vivo* support a model of organized modularity. In this model the date hubs connect biological processes (or modules) to each other and thus organize the proteome, while the party hubs function inside the modules.

Contents

1 Introduction

A *scale-free network* is an important kind of a complex network because many real-world networks, for example social networks and the World Wide Web, fall into this category. In scale-free networks, some nodes act as highly connected hubs (high degree), although most nodes are of low degree. Scale-free networks' structure and dynamics are independent of the system's size N, the number of nodes the system has. Their most distinguishing characteristic is that their *degree distribution* (the probability distribution of degrees in a complex network) follows a power law relationship

$$P(k) \sim k^{-\gamma}$$

i.e. the probability P(k) that a node in the network connects with k other nodes is roughly proportional to $k^{-\gamma}$. The coefficient γ varies approximately from 2 to 3 for most real networks.

In *scale-free protein-protein interaction networks* ('interactome' networks) most proteins interact with few partners, whereas a small but significant proportion of proteins, the hubs, interact with many partners [HBH04]. However, it has been claimed that there is some protein interaction data that do not exhibit power law statistics [TYD05].

2 Dynamically Organized Modularity in the Yeast Protein-Protein Interaction Network

Han *et al* discovered two types of hubs: 'party' hubs and 'date' hubs [HBH04]. The biological role of hubs might vary depending on the timing and location of the interactions they mediate (Figure 1). In this protein interaction network the proteins are coloured according to mutual similarity in their yeast mRNA expression patterns. Party hubs are highly correlated in expression with their partners, and presumably interact with them simultaneously. The partners of date hubs show more limited co-expression, and probably the physical interactions occur at different times or locations.

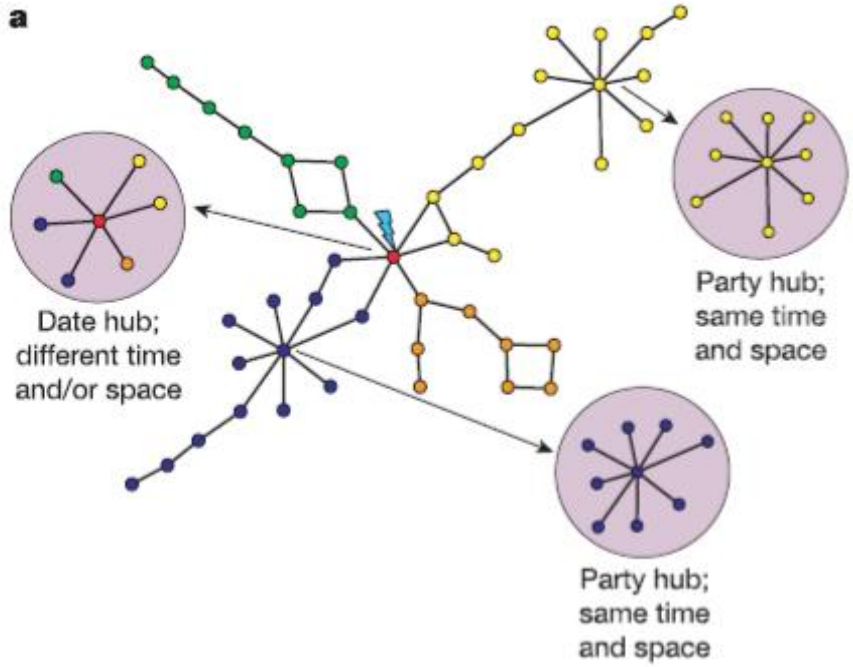

Figure 1. Date and party hubs.

2.1 Filtered Yeast Interactome (FYI) data set

Hubs connected by false positive interactions [MKS02] would be uncorrelated in mRNA expression with their partners, and would be similar to date hubs. To minimize false positives, a high-quality yeast interaction data set was generated by intersecting data generated by several different interaction detection methods. The resulting 'Filtered Yeast Interactome' (FYI) data set contains 2493 high-confidence interactions. The FYI network contains 1379 proteins with an average degree of 3.6 interactions per protein and a large connected component of 778 proteins. The degree distribution follows the power law that is characteristic for scale-free networks: $P(k) \sim k^{-\gamma}$ with $\gamma \approx 2$ (Figure 2).

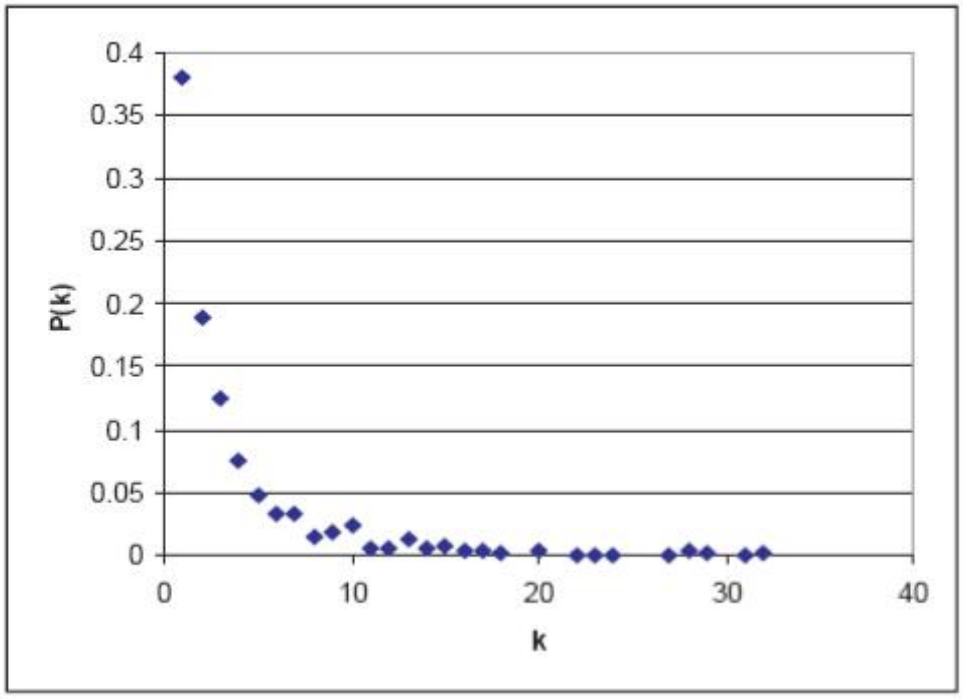

Figure 2. Scale-free organization of the FYI network: proportion of nodes with *k* partners P(*k*) versus node degree *k*.

FYI hubs were characterized with an expression-profiling compendium of 315 data points for most yeast genes across five different experimental conditions. For each hub, the average of *Pearson correlation coefficients* between the hub and its partners (i.e. the neighbouring nodes in the network) for mRNA expression was calculated. The Pearson correlation coefficient (PCC) is a measure of the correlation of two variables *X* and *Y* measured on the same object or organism, that is, a measure of the tendency of the variables to increase or decrease together. The coefficient ranges from −1 to 1. A value of 1 shows that a linear equation describes the relationship perfectly and positively, with all data points lying on the same line and with *Y* increasing with *X*. A score of −1 shows that all data points lie on a single line but that *Y* increases as *X* decreases. A value of 0 shows that a linear model is inappropriate, i.e. there is no linear relationship between the variables.

2.2 Splitting the hub population into party and date hubs

The hubs were defined as nodes (proteins) with degree $k > 5$. Suprisingly, the average of Pearson correlation coefficients calculated between each hub and its partners follow a *bimodal distribution* shown in Figure 3 (red curve). In statistics, a bimodal distribution is a continuous probability distribution with two different modes, which appear as distinct peaks (local maxima) in the probability density function.

The bimodal distribution suggests that hubs can be split into two distinct groups:

- party hubs with relatively high average PCCs, and
- date hubs with relatively low average PCCs.

Party hubs have in average a similar mRNA expression as their partner proteins. This indicates that party hubs interact with their partners simultaneously.

The partners of date hubs show more limited co-expression in average, so probably the interactions occur at different times.

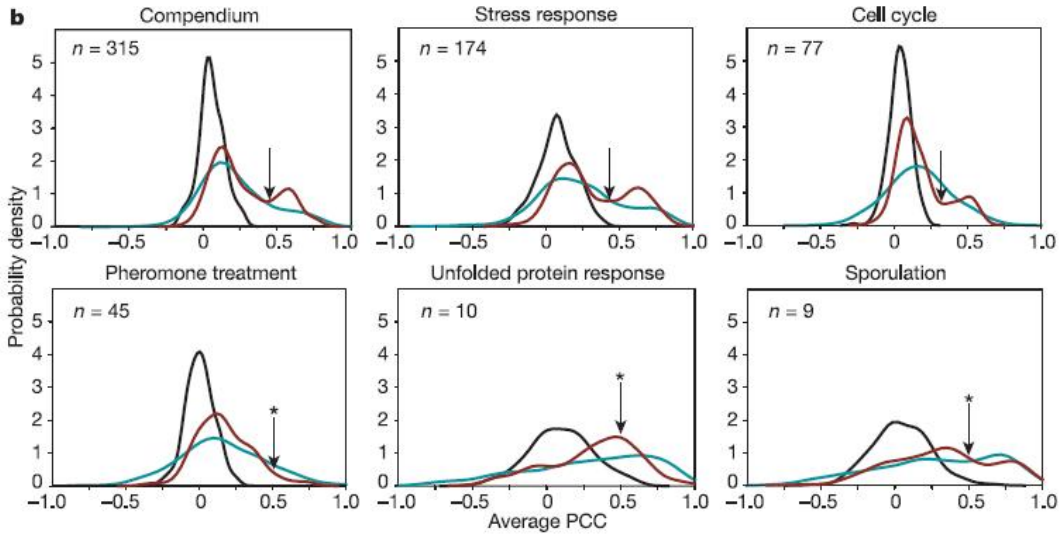

Figure 3. Probability densities of the average PCCs were calculated from a global expression profiling compendium (top left). Average PCCs were also calculated independently for each condition constituting the compendium. The number *n* refers to the number of data points for each gene for each condition. The red curve is the average PCCs for hubs and it shows a clear bimodal distribution in the top panels.

The bimodal distribution suggests a natural boundary for separating date hubs from party hubs. In Figure 3, the bimodal distribution is used to separate date and party hubs (located by the arrow) for the conditions shown in the top panels, i.e. 'stress response' and 'cell cycle'.

For the conditions in the bottom panels that do not show a clear bimodal distribution, an arbitrary average PCC cutoff of 0.5 was used.

In contrast, there is no bimodal distribution observed with the average PCCs of non-hub proteins, defined as nodes with degree $k < 5$ (blue curve). For hubs in randomized interactome networks the average PCCs also show a normal distribution centered on 0 (black curve).

For subsequent analysis, party hubs were defined as hubs with an average PCC higher than the threshold indicated by the arrow, in at least one of the five conditions in Figure 3. All other hubs were defined as date hubs. Using this criteria, Han *et al* found 91 date hubs and 108 party hubs in FYI.

The dynamics of interactome networks should be considered not only by expression timing but also spatial distribution, i.e. subcellular localization. Han *et al* estimated the localization diversity of partners of hubs by using a proteome-wide cellular localization data set [HFG03]. Partners of date hubs were clearly more diverse in spatial distribution than partners of party hubs. Therefore, the distinction between date and party hubs obtained from gene expression is repeated by protein localization data.

2.3 Removing the hubs from the network

In general, scale-free networks are tolerant to random node removal but they are very sensitive to the targeted removal of hubs.

When removed from the interactome network, party and date hubs have different effects on the topology. Han *et al* used an *in silico* strategy [AJB00] that simulates the effect of specifically removing (attacking) hubs in the FYI network on the *characteristic path length* of the main component of the network. The characteristic path length is defined

as the average distance between node pair, and it reflects the overall network connectivity [AJB00].

As expected, attacks against FYI hubs, without distinguishing between party and date hubs, have a significantly more damaging effect on the network integrity than the removal of random proteins.

However, this *in silico* experiment revealed an unexpected difference between party and date hubs. Removal of party hubs does not affect connectivity and thus resembles failures. In contrast, attacks against date hubs account for a vast majority of the effect observed when attacking all hubs. Thus, date and party hubs have clearly different global properties in the interactome network.

2.4 Subnetworks

It was discovered that the main component that remains after the removal of party hubs is significantly larger than the component remaining after the removal of date hubs (Figure 4). Conversely, the subnetworks obtained by date hub removals tend to be larger in size and number than those obtained by party hub removals.

To test whether FYI subnetworks obtained after the complete removal of date hubs corresponds to small interaction maps of specific biological processes, Han *et al* estimated the functional homogeneity by using annotations from the Munich Information Center for Protein Sequences (MIPS) database [MHK02]. In comparison with control networks of the same size distribution, most FYI subnetworks were more homogeneous in function. A 'most likely' function could be assigned to each subnetwork by determining the most enriched function category among all nodes over the entire FYI data set. Thus, the subnetworks often correspond to known biological modules.

Figure 4. The main component of the FYI network (top panel) splits into small subnetworks after the removal of date hubs (middle panel). The network stays almost undamaged after the removal of party hubs (bottom panel).

Subnetworks represent not only stable molecular machines or complexes, such as the ribosomal RNA synthesis complex, but also more loosely connected regulatory pathways, for example osmosensing. Protein pairs inside subnetworks corresponding to protein complexes tend to show high PCC values, whereas more loosely connected regulatory pathway modules tend to show lower PCC values (Figure 5).

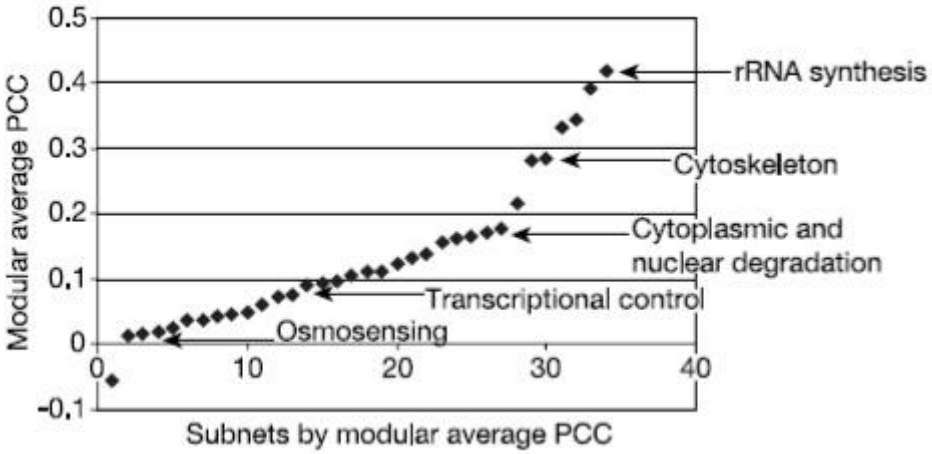

Figure 5. Subnetworks are probably both complexes and more loosely connected modules. The arrows indicate several examples.

2.5 The roles of party and date hubs in the model

These results support a model of organized modularity for the yeast proteome (Figure 6). In this model, date hubs are global, higher level connectors between modules. Party hubs function inside the modules, at a lower level of the organization of the proteome.

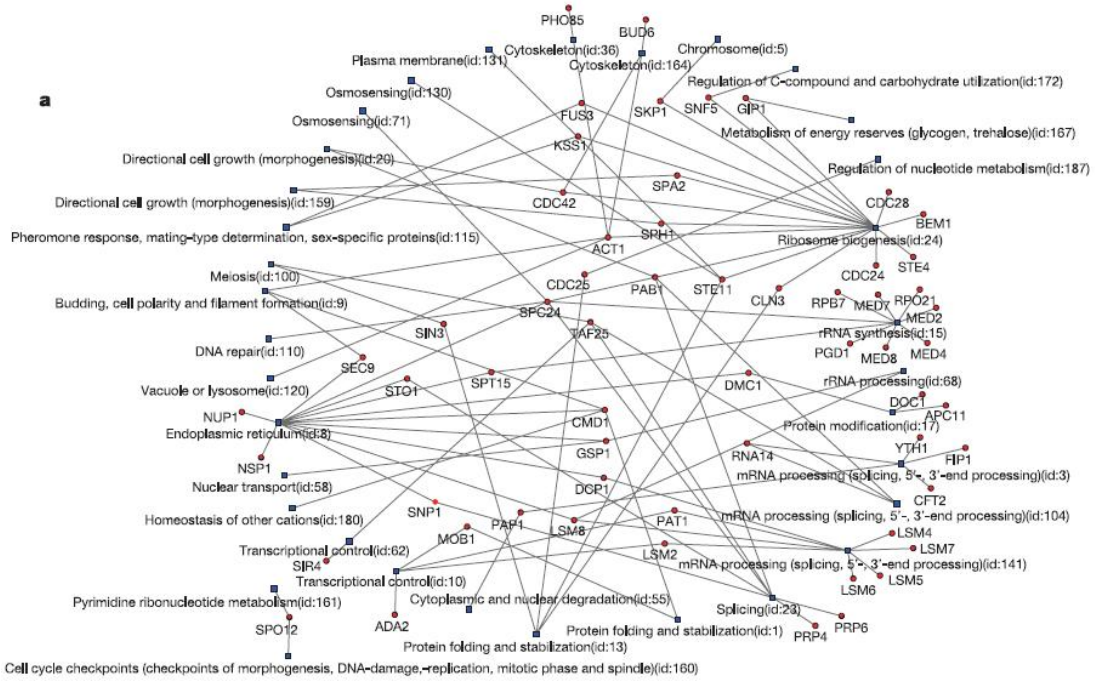

Figure 6. Organized modularity model.

For example, in Figure 7, the date hub calmodulin (Cmd1) connects four different biological modules: 'homeostasis of cations', 'protein folding and stabilization', 'budding, cell polarity and filament formation' and 'endoplasmic reticulum'. The party hubs Sec17, Sec22 and Vti1 all function inside the 'endoplasmic reticulum' module.

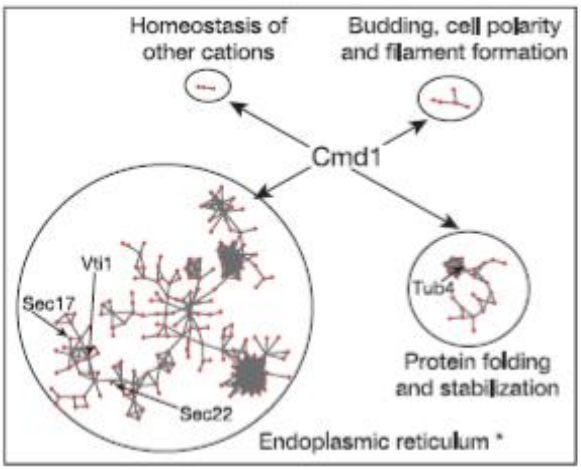

Figure 7. The date hub Cmd1 connects four biological modules at higher level, whereas the party hub Sec22 connects to eight proteins within the 'endoplasmic reticulum' module.

The organized modularity model predicts that experimental perturbations of date hubs *in vivo* should have different effects from perturbations of party hubs. In single-gene knockout experiments [WSA99, GCN02] similar proportions of party and date hubs score as essential genes. Although party hubs tend to function locally within modules, they can still mediate unique functions in essential modules and thus score as essential genes. This explains the similar essentiality rate between date and party hubs.

On the other hand, genetic perturbations of date hubs tend to make the proteome sensitive to other perturbations, more so than perturbations of party hubs. Among all genetic interactions curated in MIPS, genetic interactions involving date hubs are twice as prevalent as those involving party hubs or non-hub proteins. The higher rate of observed genetic interactions for date hubs suggests that they have a central role in organizing the modularity of the yeast proteome. In contrast, the lower rate of observed

genetic interactions for party hubs reflects their localized role within isolated regions of the proteome.

However, it has also been speculated that the coordination between the modules might be occurring in a pairwise fashion, rather than by way of high-degree hub proteins responsible for coordinating multiple modules [VaC06].

3 Conclusions

Hubs in the yeast interactome network can be classified into date and party hubs on the basis of their partners' expression profiles. This distinction suggests a model of organized modularity for the yeast proteome. Modules are connected through the date hubs which act as regulators, mediators or adaptors. Party hubs represent integral elements within the modules and tend to function at a lower level of the organization of the proteome, although they are important for the functions mediated by these modules.

Han *et al* propose that date hubs participate in a wide range of integrated connections required for a global organization of biological modules in the whole proteome network. Properties of the interactome network, such as genetic robustness and adaptability towards a wide range of external conditions, might be better understood by using such an organized modularity model.

Presuming that a modular network organization has selective advantages for reasons of stability and flexibility, similar partitioning might reveal modularity also in *metazoan* (multicellular animals) interactome networks [LAB04, GBB03]. Similar temporal or spatial dynamic analysis might also be applied to non-biological networks, such as the World Wide Web, epidemiological networks and social networks. It is also possible that discriminating between date and party hubs might help to define new therapeutic drug targets.

References

AJB00 R. Albert, H. Jeong and A.L. Barabasi, *Error and attack tolerance of complex networks,* Nature 406:378-382, 2000.

GCN02 G. Giaever, A.M. Chu, L. Ni *et al., Functional profiling of the Saccharomyces cerevisiae genome,* Nature 418:387-391, 2002.

GBB03 L. Giot, J.S. Bader, C. Brouwer *et al., A protein interaction map of Drosophila melanogaster,* Science 302:1727-1736, 2003.

HBH04 J-D. J. Han, N. Bertin, T. Hao *et al., Evidence for dynamically organized modularity in the yeast protein-protein interaction network,* Nature 430:88-93, 2004.

HFG03 W-K. Huh, J.V. Falvo, L.C. Gerke *et al., Global analysis of protein localization in budding yeast,* Nature 425:686-691, 2003.

LAB04 S. Li, C.M. Armstrong, N. Bertin *et al.***,** *A map of the interactome network of the metazoan C. elegans,* Science 303:540-543, 2004.

MKS02 C. Mering, R. Krause, B. Snel *et al., Comparative assessment of large-scale data sets of protein-protein interactions,* Nature 417:399-403, 2002.

MHK02 H.W. Mewes, K. Heumann, A. Kaps *et al., MIPS: a database for genomes and protein sequences,* Nucleic Acids Res. 30:31-34, 2002.

TYD05 R. Tanaka, T-M. Yi and J. Doyle, *Some protein interaction data do not exhibit power law statistics,* FEBS Letters, Vol. 579, Issue 23, 5140-5144, 2005.

VaC06 A.X.C.N. Valente and M.E. Cusick, *Yeast Protein Interactome topology provides framework for coordinated-functionality,* Nucleic Acids Res. Vol. 34, No. 9, 2812-2819, 2006.

WSA99 E.A. Winzeler, D.D. Shoemaker, A. Astromoff *et al., Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis,* Science 285:901-906, 1999.