Modularity and node roles in protein interaction networks

Sumeet Agarwal

Supervisors: Charlotte Deane, Nick Jones, Mason Porter

Protein Interaction Networks

- Data on physical protein-protein interactions (the interactome) gathered from multiple sources: yeast-2-hybrid (Y2H), affinity purification (AP/MS) etc.
- Can be represented as an unweighted, undirected graph (network)
- Much focus recently on using network analysis tools to understand global interactome properties

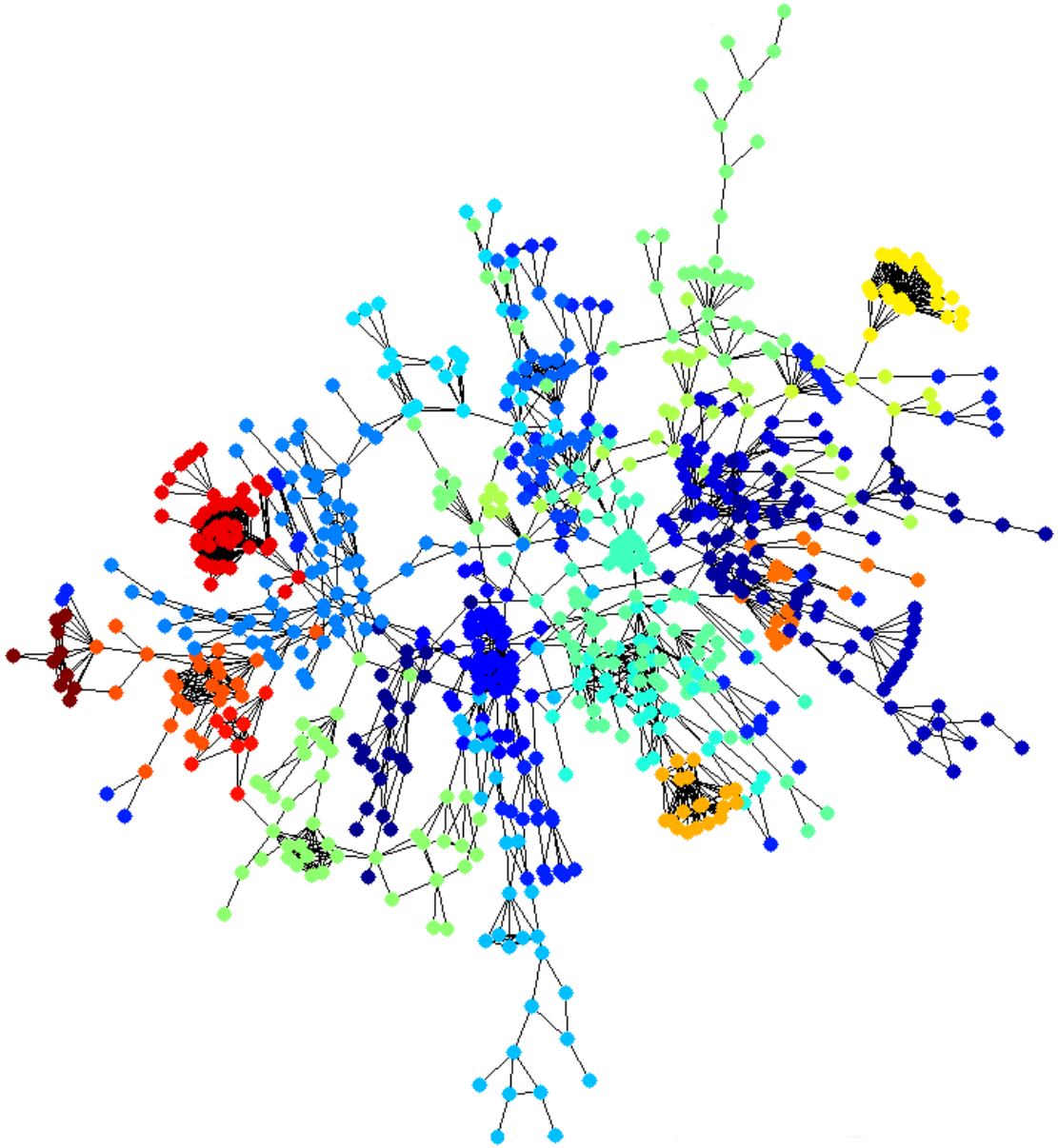
Community Detection

- Most real-world networks display some modularity: the nodes fall into clusters or communities with higher intra-cluster connectivity than inter-cluster
- A commonly used formal metric:

$$Q = \frac{1}{2m} \sum_{k=1}^{N} \sum_{i,j \in C_k} (A_{ij} - \frac{k_i k_j}{2m})$$

- One way of detecting meaningful communities is to find a partition that maximises *Q*

Modules in the yeast interactome

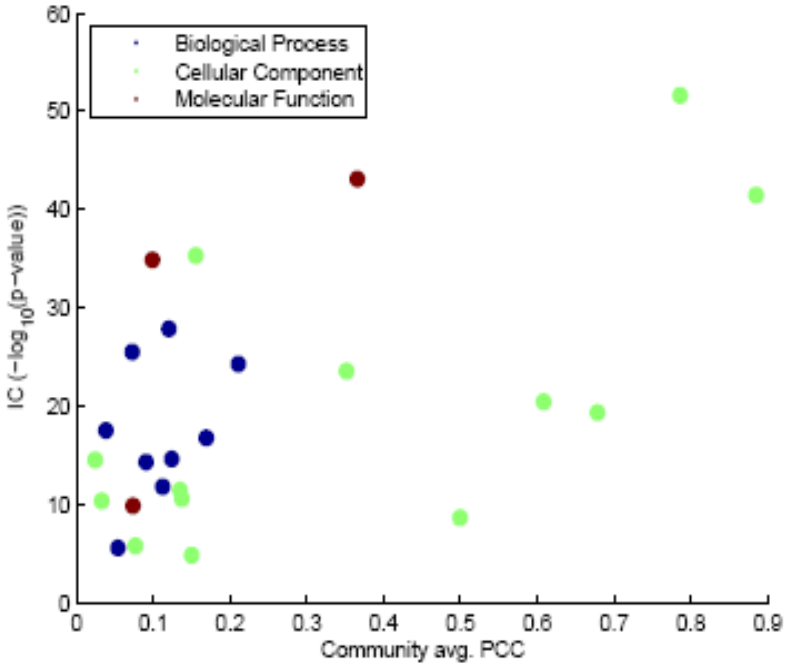


Using expression data

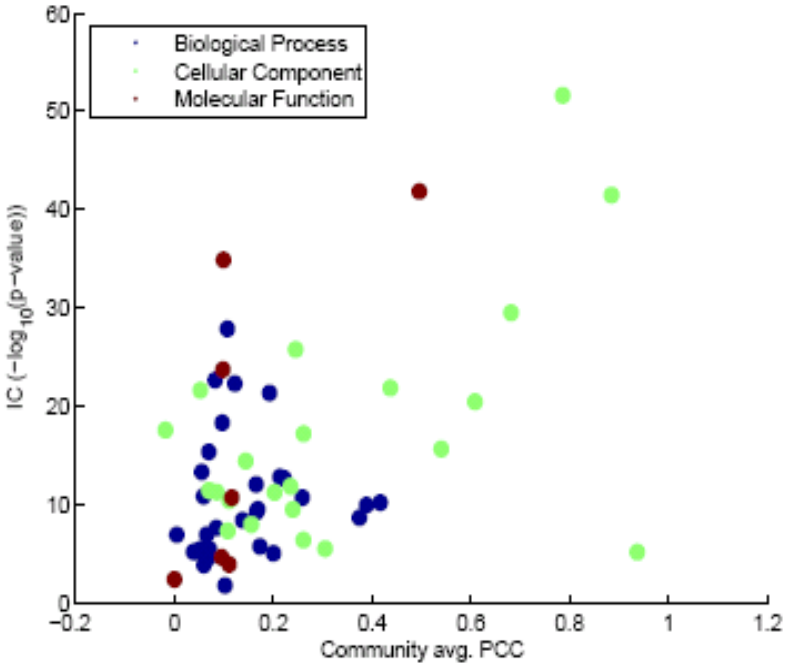
- Microarray expression data also potentially provides information about functional relationships between proteins: co-expressed proteins are likely co-regulated
- We can add this information to the interaction network by weighting edges by the expression correlation coefficients
- A weighted version of the modularity metric can be used to find communities

Community Quality

- We use term enrichment in GO as a measure of functional homogeneity



(a) FYI interaction network (25 communities, mean IC = 20.15)



(b) FYI hybrid network (59 communities, mean IC = 13.83)

Node roles in the interactome

- We use communities detected from interaction data only, which seem better
- Topological properties of nodes can be used to define ‘roles’ (Guimerà & Amaral, 2005):

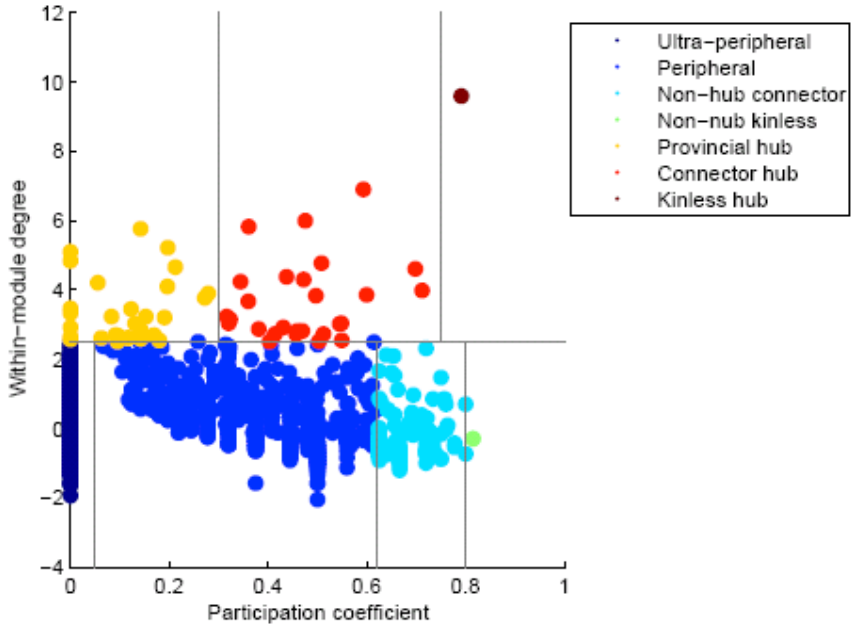
$$z_i = \frac{\kappa_i - \bar{\kappa}_{s_i}}{\sigma_{\kappa_{s_i}}} \qquad P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{\kappa_{is}}{k_i} \right)^2$$

- Within-module degree z_i distinguishes ‘hubs’ from ‘non-hubs’, whereas participation coefficient P_i indicates to what extent connections are spread across communities

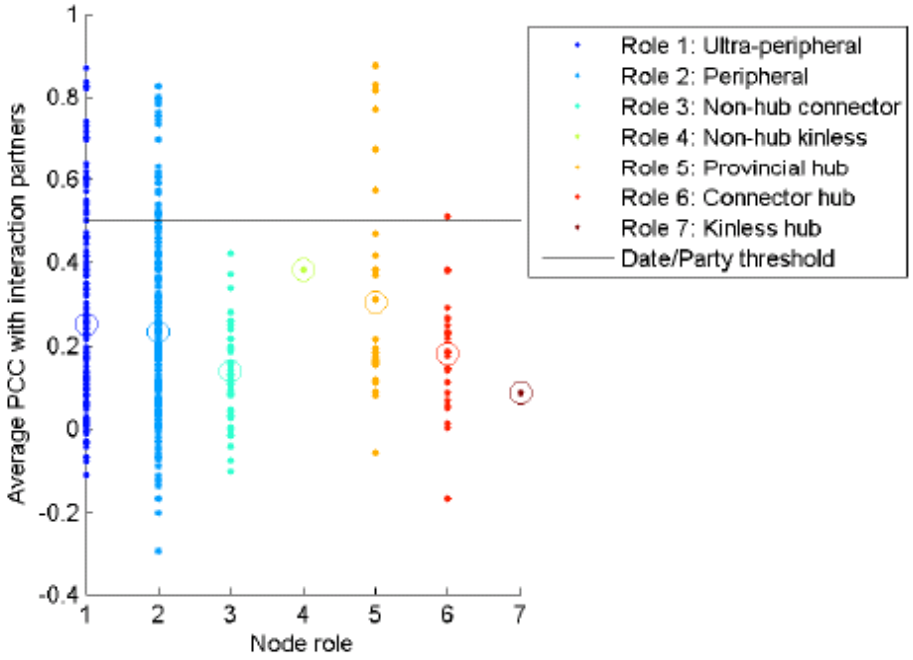
Date and Party hubs

- It has been proposed (Han *et al.* 2004) that highly connected proteins fall into 2 classes, based on average coexpression with their interaction partners
- ‘Party hubs’ interact with many proteins at once, and are said to coordinate a specific biological function or process
- ‘Date hubs’ interact with different proteins at different times, and are said to act as interfaces between distinct functional modules

Correspondence of node roles and Date/Party hubs



(a) FHC (2,233 nodes, 63 communities)



(a) FHC (553 hubs with min. degree = 7)

Experimental technique issues

- Date/Party dichotomy has little topological significance, but may be due to varied sources of interaction data
- Examination of single-source datasets revealed ~100% date hubs in Y2H data, ~80% party hubs in AP/MS data
- There is very little overlap between these datasets, suggesting that they detect different kinds of interactions

Conclusions

- Interaction data can give meaningful communities; not clear if combining with coexpression data helps
- Date/Party hub distinction appears too simplistic: nodes fall into various roles, and partner coexpression is not a very good predictor of topological properties
- Interaction networks obtained from different experimental sources have widely varying properties

Current and Future Work

- We are looking to explore other measures of node importance, in particular ‘centrality’ measures, to try and propose a better categorisation of ‘hub’ proteins
- We are also looking at using hierarchical structure to help in defining node roles
- Examining how best to reconcile Y2H, AP/MS and other data types, and also combining these with other information like genetic interaction data

References

- Roger Guimera and Luis A. Nunes Amaral. Functional cartography of complex metabolic networks. Nature, 433:895–900, 2005.
- Jing-Dong J. Han et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 430:88–93, 2004.