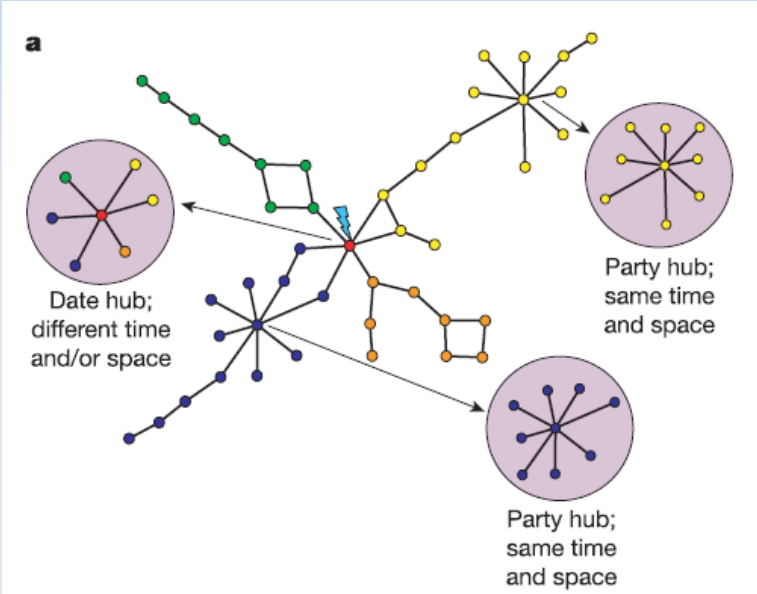
Ph.D. Overview: Roles in Protein Interaction Networks

- PINs are a means of representing all physical interactions between proteins in a cell; allowing for a bird's-eye view and an attempt to gain a systems-level understanding of how the cellular machinery operates.
- Hubs are proteins with many links; it was suggested that they are of two types, Date and Party, which play key roles in network organisation
- Idea was exciting because it suggested that network-based analysis could lead to genuine insights into biology
- Was controversial and prominently debated; we re-analysed data in several ways to find strong evidence that it was flawed, and that there is no simple dichotomy of hub roles
- Also showed that link roles may be of interest: centrality of a link correlates inversely with the functional similarity of the proteins it links

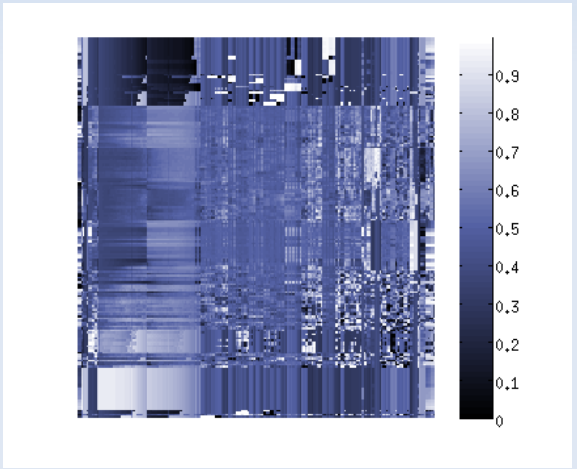


(Han et al., *Nature* 2004)

PLoS Comp Biol 6(6):e1000817 (2010)

Ph.D. Overview: High-Throughput Analysis of Networks

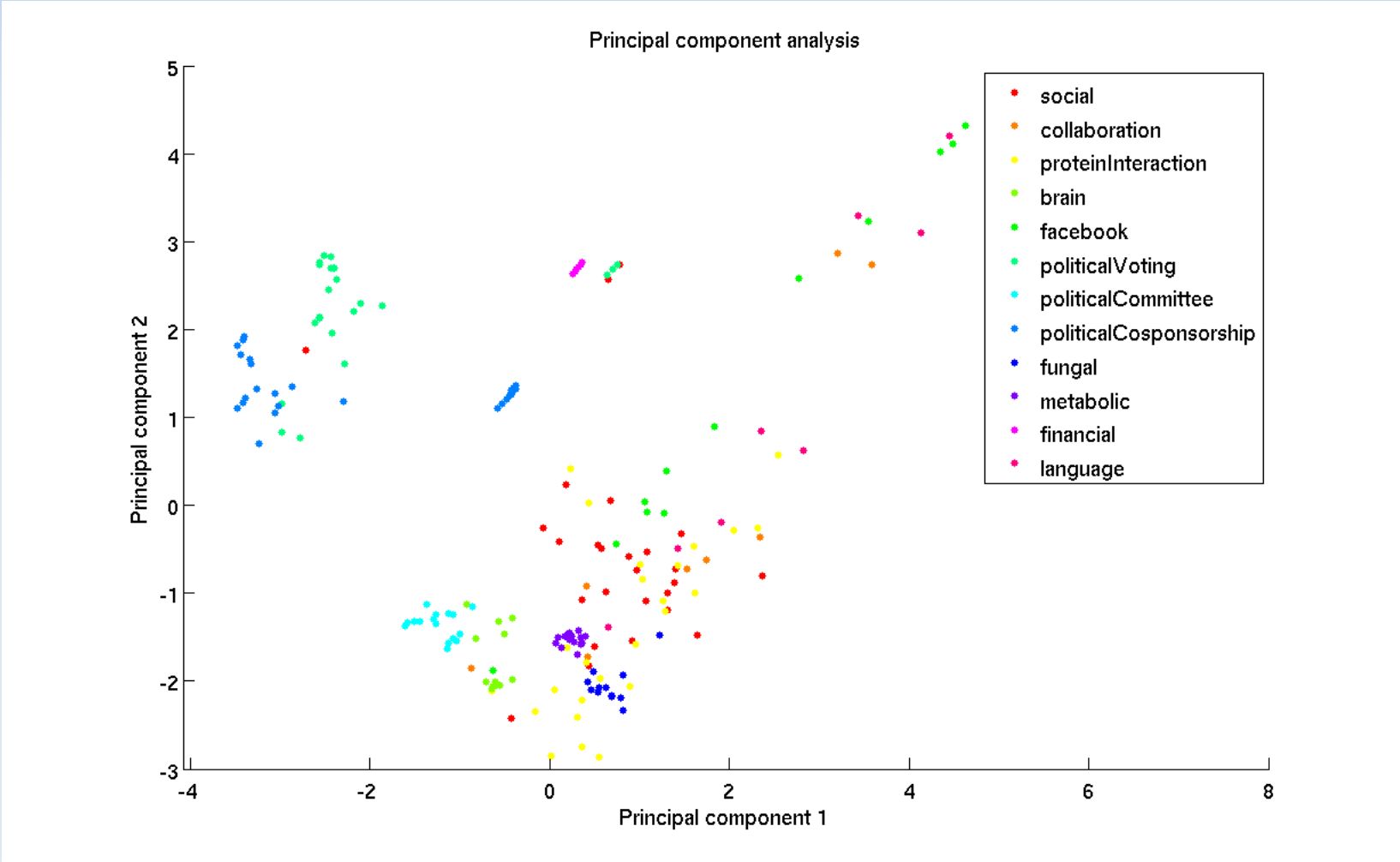
- Motivating questions: How does the structure of natural networks relate to their functionality? How do these networks emerge? What are the best ways of thinking about networks in a given context?
- People from a wide range of disciplines have studied graphs or networks: Mathematicians, Computer Scientists, Electrical Engineers, Sociologists, Physicists, Statisticians etc.; leading to a fragmented literature, with inconsistent terminology and reinvention of concepts and methodologies
- We aimed to use computing and data mining techniques to systematically investigate patterns of relationships between different kinds of networks and metrics/features
- We apply a large a number of diagnostics to a large number of networks, giving a 'data matrix' which can be mined in various ways



Proceedings of ACNE, ECML PKDD 2010;
Proceedings of MLSB 2010 (extended abstracts)
Paper in preparation

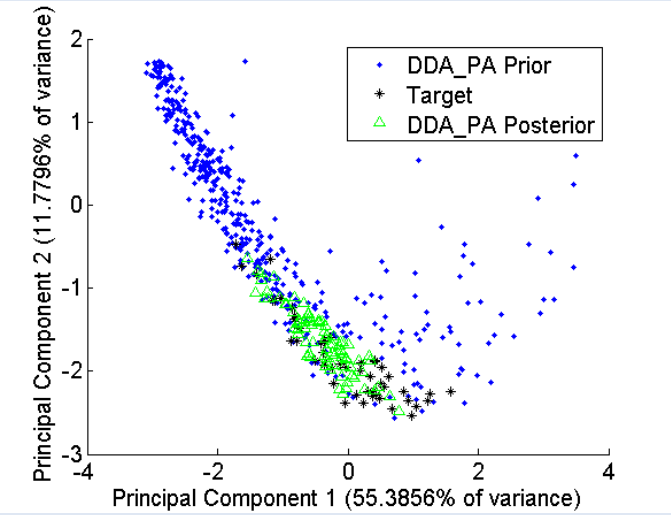
Ph.D. Overview: Network Clustering

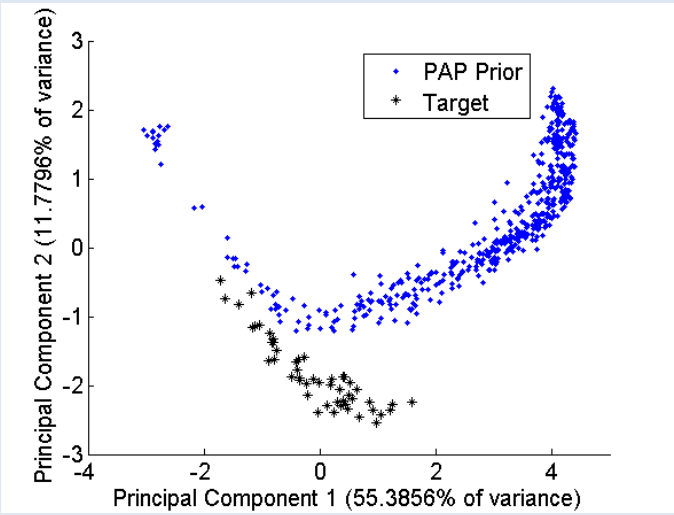
- Each row of data matrix represents a network as a vector of features; a point in a high-dimensional vector space which can be projected onto a low-dimensional one

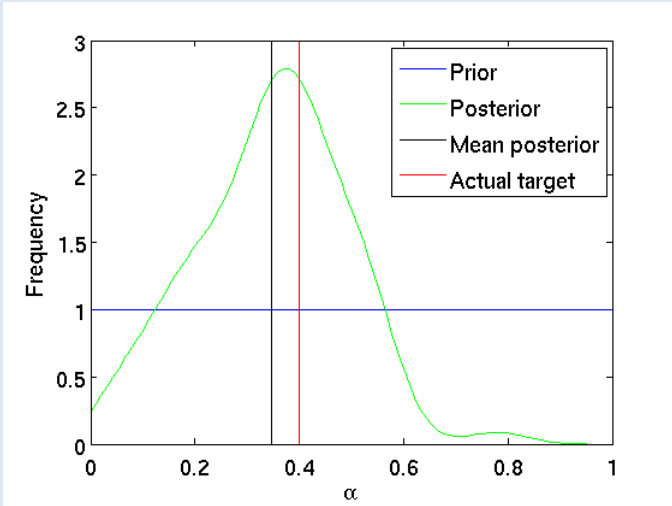


Ph.D. Overview: Comparing Models for PIN Evolution

- Is the structure we see in complex networks the outcome of relatively simple stochastic events, or does its explanation require more intricate mechanisms?
- We can use the feature space to compare given networks to model-generated ones, and thus assess different generative models and parameter settings
- We combine a statistical model-fitting approach called Approximate Bayesian Computing (ABC) with our network representation
- Addressing two key questions: How natural networks evolve, and how does one fit complex models to data

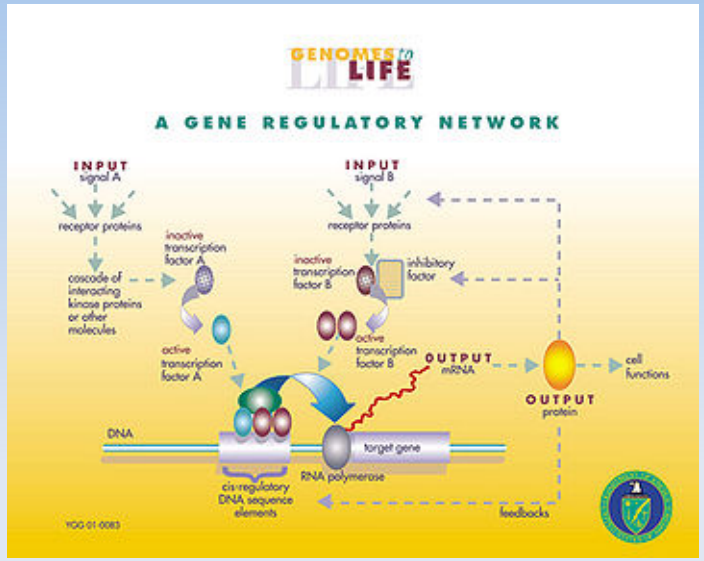




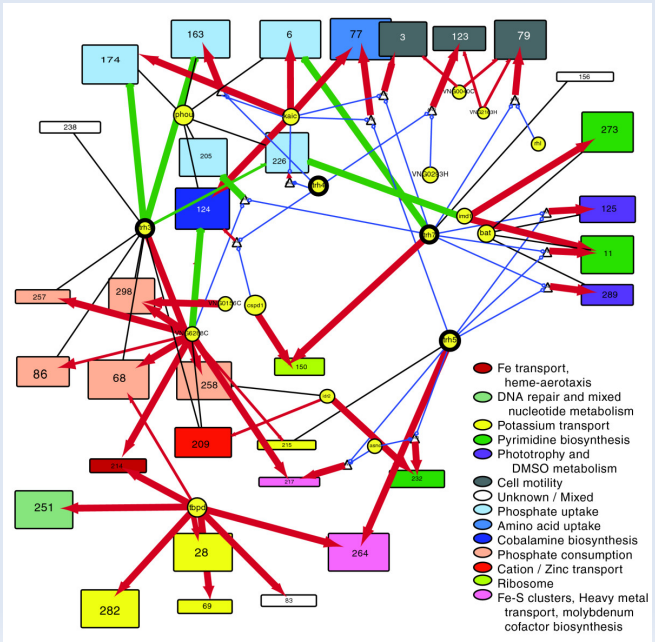


Project Proposal: Challenges To Address

- Scientific
  - To understand cellular circuitry, we need to look at dynamics and control in the interactions between genes/proteins: not merely static networks
  - Regulation and information flows; how systems-level behaivour emerges from individual components
- Technical
  - Much work on gene regulatory networks; but hard to choose from a plethora of models, as data leaves the problem underspecified
  - Integrating gene regulatory networks with protein interaction networks
- Medical/Sociological
  - Feeding back into wet lab and medical applications. I plan to collaborate with experimentalists at JNU/NCBS, focusing on under-studied organisms such as the amoebiasis parasite *E. histolytica* (responsible for ~100,000 deaths/year)



(NCBI)



(Bonneau et al., *Cell* 2007)

Project Proposal: Data and Methodology

- Public databases for well-studied organisms (yeast, humans) to construct initial models
- I hope to get access to data from experimental collaborators, in particular Prof. Alok Bhattacharya's lab at JNU, which has worked on *E. histolytica*
- Develop predictive mathematical models of subcellular circuitry and dynamics. Use structural signatures from high-throughput analysis to inform/constrain search
- Representation language which can incorporate interactions across different levels (genes, proteins) and also encode background knowledge like the structural constraints. Logic-based representations useful for encoding qualitative, intuitive rules and models; e.g., instead of modelling continuous varation in expression level, a gene might be represented as being in one of a few discrete states, say ON/OFF
- Automated machine learning to find rules/models that best explain data, given known constraints; for logical representation will mean Inductive Logic Programming (ILP)
- Generate testable predictions for novel stimuli, feed back into experiments
- My previous work on protein interaction networks, on the feature-based representation of network structures and some experience with ILP should help
- For the network modelling and machine learning aspects, I hope to collaborate with Dr. Nick Jones (Imperial College London) and Prof. Ashwin Srinivasan (SAU Delhi)

Collaboration and Integration

- I hope to establish collaborations with a number of people with similar interests: Alok Bhattacharya and Narinder Sahni (JNU), Ashwin Srinivasan (SAU), Mukund Thattai (NCBS), Nick Jones (Imperial)
- I intend to engage with the Centre for Molecular Parasitology at JNU, recently funded by the DBT, whose remit includes studying *E. Histolytica*
- Participation in an EPSRC (UK) grant enabling reciprocal visits for research collaboration between Imperial College London and Indian institutions. Visiting Imperial and inviting people from there to establish links with their large concentration of systems/computational biology researchers
- Teaching/supervision, for instance as part of Systems Biology M.Tech. programme at JNU. I have also discussed with Prof. Ashwin Srinivasan, Prof. Indira Ghosh (SCIS, JNU) and others the possibility of initiating a doctoral training centre in Systems Biology, along the lines of my alma mater at Oxford; I would like to explore the possibilities for this further