Properties of Interaction Networks

Based on the following papers:

Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs
   Yu et al.

Evidence for dynamically organized modularity in the yeast protein-protein interaction network
   Han et al.

Introduction

A. Motivation

Protein interactions are ubiquitous and essential for cellular functions. For example, the processes such as signal transduction, metabolic pathway and transcription regulation are done by interaction of proteins.

Here we explain briefly how interactions between proteins play the role in these processes.

I. Cell signaling

PIP2 cleavage to IP3 and DAG initiates intracellular calcium release and PKC activation

One example for a cell signaling system is phosphatidylinositol (PIP2) metabolism and lipid signaling system. Phospholipase C is a key enzyme in this system that hydrolyzes PIP2, into two second messagers, inositol triphosphate (IP3) and diacylglycerol. During cellular signaling it regulates the activity of downstream proteins.
II. Metabolic Pathway

There are several metabolic pathways; what we have chosen here, is Pre-mRNA Processing (Splicing). During RNA processing, snRNPs loop the intron and bind to it to form a spliceosome. Then the intron is separated from the exons and is removed from the pre-mRNA. Then exons are spliced together and produce the translatable mRNA. This mature mRNA exits the nucleus and is translated in the cytoplasm.

III. Transcription Regulation

The enzyme for transcription process needs some proteins associated with RNA polymerase, called transcription factor. The transcription factors bind to RNA polymerase or to another transcription factor or to cis-acting DNA sequences. A transcriptional regulatory network is a collection of these regulatory proteins associated with genes across a genome.

A common transcription factor used for yeast two-hybrid screening is GAL4.

In these examples we did not have that many interactions and we could use a tutorial diagram to visualize the interactions. But it is not possible to show all the interactions that occur within a cell in one diagram and the alternative way is to use the graph representation to convey all the interactions that occur within a cell. In graph representation, nodes are proteins and each of the edges represents an interaction.
There are two definitions for the interaction:

1. **Physical Interactions**, in which two proteins contact each other
2. **Complex Interactions**; it is the super set which means it contains all the physical Interactions; but it also contains the interactions in which two proteins interact but not directly.

In some networks they show only physical interactions but they may show all the complex interactions in a network.

**B. Importance of Protein Interaction Networks**

Studying protein interaction network architecture allows us to:

- Assess the role of individual proteins in the overall pathway:
  You may know the protein's function or how it is involved in a biological process, nevertheless you can not know the mechanisms by which the protein function until you have the entire network. By looking at the entire network you can realize which proteins are correlating.

- Evaluate redundancy of network components:
  Suppose we have two proteins that communicate via a third protein; but they maybe can communicate without the third protein; but we can not know that until we have an overal picture of the entire network.

- Identify candidate genes involved in genetic diseases:
  If I have a disease and I know some genes involved in the disease but I do not have the entire set of associated genes to that disease; If one protein interacts with a high percentage of these genes, say %75, then we can say this protein is involved in that disease also.

- Sets up the framework for mathematical models:
  If you know the entire interaction network a sort of sets of the frameworks you can write with differential equations simulating the individual modules within the network to see how regulation occurs within the network.

For complex systems, the actual output may not be predictable by looking at only individual components.

**C. Protein Interaction Data**

Previously, experiments tested proteins one by one; but nowadays we have high-throughput experiments. Two famous high-throughput experiments are **Yeast 2 Hybrid Screens** and **Co-IP**.

---

**Yeast 2 Hybrid Screen (Cytotrap System)**

![Yeast 2 Hybrid Screen Diagram](image_url)
The yeast two-hybrid method is used to study the pairwise protein-protein interactions. Gal4 protein is a transcriptional factor composed of two parts, a DNA-binding domain that interacts with promoter and an activating domain which interacts with polymerase. If two proteins interact physically, DNA-binding domain of GAL4 activates transcription. Detection of protein-protein interactions is based on the fact that what hSos requires to be recruited to the membrane. A yeast strain contains a temperature-sensitive mutation in the yeast Ras guanyl nucleotide exchange factor, cdc25-2. The target protein, is fused to hSos, and cDNA is fused to a membrane-localization signal. The cells only grow at 25°C but not at 37°C. When these hybrid proteins are co-expressed in a cdc25-2 yeast strain, allows growth at 37°C. If the two fusion proteins interact, hSos can recruit to the membrane.

However, different sources of errors can affect the accuracy of these high-throughput experiments:
- False positives / False negatives
- Self-activators: proteins are naturally recruited to the cell membrane
- Promiscuous proteins
- Protein concentration differences
- Lack of benchmark

### D. Network cross-comparison

Pairs of proteins have been binned according to their shortest path in networks generated from Y2H and Co-IP data. The false-color map indicates bins with more (red) or fewer (blue) interactions than expected by chance. Bins enriched for true positives, false positives and true noninteractors are indicated.

In this figure, we compare the results in terms of interaction networks, the results from CO-IP network vs. the results from Y2H network.

![Network Cross-comparison](image)

It is very likely to detect an interaction using one method and see that the distance in terms of edges between these two potential interactors is very high.

### E. Validation

These are some of the methods that besides the experimental data, people use to validate the interaction data from experiments.
1. mRNA co-expression

You are taking a set of potential protein interactors that have been found from the experimentation and for each of these proteins, you look at the expression profiles and compare them for these proteins. So for example here we have two different genes that we think whose proteins are interacting. We make two external conditions, heat shock and stress; the ideal scenario is that we see the same expression level for both of these two genes. So the idea is that if two proteins interact they have similar mRNA expression.

2. Genetic interactions

Studying genetic interactions can reveal gene function, the nature of the mutations, functional redundancy, and protein interactions. Because protein complexes are responsible for most biological functions, genetic interactions are a powerful tool. If it is true that two proteins interact, it is very likely that their corresponding genes have interaction with each other.

All of these methods are for validation test but they can also be used to find the interactions. Although these do not correspond necessarily to physical interactions this is a general protein functional associations. All of these pieces of evidence say that those two genes are involved in some way in a single protein.

_analysis based on validation studies show that only 30 – 50 % of high-throughput interactions are valid._

Joint analysis of physical and genetic interactions

Genetic interactions have been used as anchors to mine the physical interaction network. The physical interactions can be validated by genetic interactions. this is a supportive way to verify the recognized interactions.

F. Network conservation across species

If we look at the interactome or protein interaction network in each of the species there is a very high degree of conservation. Here we have three species that we have the most protein interactions data for them, yeast, worm, and fly. For highly conserved cellular processes that all organisms need them to survive such as protein folding, RNA metabolism and etc. there are similar modules of protein interactions that are highly conserved among the species. By similar interactions we mean that the proteins that map across the species are similar at the sequence level.
Representative conserved network regions. Shown are conserved clusters (a–k) and paths (l and m) identified within the networks of yeast, worm, and fly. Each region contains one or more overlapping clusters or paths. Proteins from yeast (orange ovals), worm (green rectangles), or fly (blue hexagons) are connected by direct (thick line) or indirect (connection via a common network neighbor; thin line) protein interactions. Horizontal dotted gray links indicate cross-species sequence similarity between proteins (similar proteins are typically placed on the same row of the alignment).

In this lecture we go through two papers on Properties of Interaction Networks:

**Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs**

*Yu et al.*

The idea of this paper is that Interactions can be mapped from one genome to another through comparative genomics with some degree of confidence using the sequence homology

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

*Han et al.*

Unlike the first paper that look at the interactions with the perspective of individual interaction, in this paper it is tried to realize that how individual highly connected proteins contribute to the overall topology of the network. The idea of this paper is that by correlating gene expression profiles for a hub and its partners, we can predict whether it’s a date or party hub.

**Interolog Mapping**

**A. Background**

Interolog mapping is the transfer of interaction annotation from one organism to another using comparative genomics.
I. Homology-based function annotation

If two sequences are similar they probably produce proteins with similar structures and similar functions.

Sequence similarity → structural similarity → functional similarity

It is difficult to evaluate the relationship between sequence homology and function, because there is no clear measure of functional similarity between two proteins. In addition, protein function is a vague term and difficult to compare. The functions are very broad in terms of how proteins are classified, where proteins are localized to, timing, their expressions, the process it is involved in, etc. This paper focuses on one aspect of protein function that is interactions with other proteins and it defines a measure of sequence similarity so that we can actually assess how well we can transfer protein interactions based on sequence similarity. This is a sequence similarity across the two species not between two proteins that are interacting. They are two proteins interacting in some species and they are similar at the sequence level to the pair of proteins in another species. They examine the accuracy of comparing sequences to extrapolate protein interactions.

\[
\text{Functional similarity } = f(\text{Sequence similarity})
\]

\[
\text{Protein interactions } = f(\text{Joint sequence similarity of interaction pair})
\]

II. Protein Homology

Homologs are proteins with significant sequence similarity (E-value<=10\(^{-10}\)). They are classified into Paralogs and Orthologs. Paralogs are proteins in the same species that arose from gene duplication. So Paralogs have different functions. Orthologs are proteins in different species that evolved from a common ancestor by speciation. So Orthologs have the same function.

In-Paralogs are Paralogs that were duplicated after the speciation and hence are orthologs to a cluster in the other species and out-Paralogs are Paralogs that were duplicated before the speciation, not necessarily in the same species.

So there are two types of homologs, paralogs and orthologs but in this paper, they are interested in orthologs; because we are trying to transfer the functional aspects of protein interactions. so for each interacting pairs you want to find species who is in orthologs with another species because we know A and B interact in the source organism it is likely that their corresponding orthologs A' and B' in another species will also interact in order to carry out the same functions.
Interologs

If interacting proteins A and B in one organism have interacting orthologs A' and B' in another species, the pair of interactions A–B and A'–B' are called interologs.

Before we identify where the orthologs are in another species, we first define how to detect the orthologs. There are two definitions:

1. Loose definition: Top-blast hit
2. Stringent definition: Reciprocal top-blast hit

However, not all orthologs can be found using above definitions

III. Interaction Transfer

Two methods that have been used previously for interaction transfer were:

- Best-match mapping
- Reciprocal best-match mapping

Disadvantages of these methods:

The main disadvantage that targets both these methods is that it really limits to what we can test. By that we mean, say if we have 100 interactions in one species, the fact that we can probably only find orthologs in both proteins in about 30 pairs; that is really going to limit what we can actually test. We can test only 30; we cannot really labelize the accuracy in confidence; because the data we use is very small. So the coverage of total set of interactions is low.

Not only that, the prediction accuracy is also low.

There are some limitations for interaction transfer that target not only these two methods but also all the methods:

- Some networks are more complete than others
- Proportion of proteins that is annotated
- Proportion of protein interaction partners recorded

B. New Method: Interolog Mapping
Generalized Interolog Mapping

In this method we search for all homologs of each interacting protein to construct a homolog family. Generalized interologs are any protein from one family to any protein from another family.

Proteins A₁', A₂', A₃', and A₄' in the target organism are all homologs of protein A in the source organism. These proteins form the A’ family. Likewise, protein B’s homologs (B₁', B₂', B₃') form the B’ family in the target organism. If we know that protein A interacts with B, we can predict that the A’ family and the B’ family are interacting families. All possible pairs between these two families are considered as the generalized interologs (shown as black, dashed lines with arrows).

Sequence Similarity Measures

In order to assess the accuracy of transferring interactions based on sequence similarity, we first define some quantity that measures the sequence similarity(J).

Joint Sequence Similarity

There are many ways to define joint sequence similarity. In this paper, 2 definitions are used:

Joint Sequence Identity

\[ J_I = \sqrt{I_A \times I_B} \]

It is the geometric mean of individual percent identities. If protein A interact with protein B, \( I_A \) represents the individual sequence identity of protein A and its homolog and \( I_B \) is the individual sequence identity of protein B and its corresponding homolog.

Joint E-Value

\[ J_E = \sqrt{E_A \times E_B} \]

It is the geometric mean of the individual E-values. \( J_E \) is less biased in shorter sequences than \( J_I \); the shorter the sequence is, the higher the chance of randomly finding similar sequences. While measuring homology by percent identity the length of the matching sequences is not considered.

Prediction Accuracy vs. \( J_E \) and Prediction Accuracy vs. \( J_I \) plots convey similar trend. In order to test the accuracy of interolog mapping we need to define a standard.
Gold Standard Positives P

The set of known interactions as positives in the target organism is called gold standard positives. Here the target organism is S. cerevisiae. The previous standard reference used for S. cerevisiae is the MIPS complex catalog; it contains 8250 unique interacting protein pairs. So we consider MIPS interactions as the gold standard positives. Here the loose definition for interaction is used; interacting proteins do not have to have physical interaction; they can interact via a complex association. However, proteins in the same complex do not necessarily interact with each other directly. Interacting via complex association means that two protein subunits may belong to the same quaternary complex but not physically interact.

Gold Standard Negatives N

The set of noninteracting proteins as negatives in the target organism is called gold standard negatives. It is extracted or estimated from the knowledge of protein localization. If we consider pairs of proteins in different subcellular compartments as noninteracting pairs, there are 2,708,746 non-interactions in yeast.

Schema

We have two organisms, a source organism and a target organism. If you have a bunch of interactions known in the source organism and you want to transfer these interactions over to the target organism by comparing the similarity of each of these proteins between two species. The way it works is that starting from the source interactions we can then have a set of potential interactions existing in the target organism and then we can compare them with gold standard positives and gold standard negatives to validate how many of those interactions are true.

Comparison with the gold standards. After the interactions in the source organism are mapped onto the target organism, the predictions (i.e., generalized interologs) are compared with the gold standard positives and negatives. True positives are the predictions that overlap with the gold standard positives. False positives are those that overlap with the gold standard negatives. In this paper they have used four organisms as their source organisms, C. elegans(worm), D. melanogaster(fly), S. cerevisiae(yeat) and H. pylori(bacteria). They map the interactions in these organisms onto the S. cerevisiae genome. Because S. cerevisiae is an organism for which we have a large interaction data sets.
Quantitative Parameters
We need to define some parameters to assess the accuracy of our prediction with gold standard positives and negatives. We have two assessment parameters:

I. Verification
It is your true interactions over the entire set of possible interactions.

We define a threshold $J$, so that all interactions in the source organism with joint similarities larger than that, are considered as possible interactions in the target organism.

$V(J)$ is the percentage of verified predictions among generalized interologs using $J$,

$$V(J) = \frac{|T(J)|}{|G(J)|} \times 100\%$$

$G(J)$ is the set of generalized interologs in the target organism at a certain joint similarity level ($J$).
$T(J)$ is the set of the true positives in $G(J)$.
$T(J)$ is known as the sensitivity and $G(J)$ is known as the specificity of the experiment.

II. Likelihood Ratio
True positives over the entire set of positives

$L(J)$ is the likelihood that a generalized interolog is a true prediction

$$L(J) = \frac{P(J|\text{pos})}{P(J|\text{neg})} = \frac{TP}{|P|} \frac{|P|}{FP} \frac{TP}{|N|}$$

$L(J)$ is a function of $J$, the joint similarity level. So it depends on how we define $J$.

TP is the number of true positives and FP is the number of false positives.

$|P|$ is the number of the entire positives and $|N|$ is the number of the entire negatives.

$O_{\text{prior}}$ is the "prior" odds of finding a positive and $O_{\text{posterior}}$ is the "posterior" odds of finding a positive given that, in another organism, its generalized interolog with a joint similarity $J$ is a known interaction.

The likelihood ratio relates prior and posterior odds according to Bayes' rule:

$$O_{\text{post}} = L(J)O_{\text{prior}}$$

For each organism $O_{\text{prior}}$ is constant; so, $O_{\text{post}}$ is proportional to $L(J)$ i.e. for the higher likelihood ratio, the accuracy of the prediction is higher.

In Naïve Bayesian network, if there is no correlation between features, we can use different $L$’s iteratively.
Sequence Similarity and Interaction Transfer

Weighted Average of all 4 mappings

Conservation of protein–protein interactions between homologous protein pairs

We define thresholds for the joint sequence identities (JI) and the joint E-values (JE) of the two proteins that interact. By looking at these threshold values, we know that anything above these threshold are interactions with a high degree of confidence. In this figure, the relationship between V and JI is the weighted average (based on the total number of true positives in each data set) of the relationships in all four mapping processes. This indicates that all protein pairs having JI >= 80% with a known interacting pair will interact with each other, whereas few pairs interact at JI < 40%.

The plot for joint E-values (JE) indicates that more than half of the protein pairs with JE <= 10–70 indeed bind to each other. Therefore, JE of 10–70 could be used as a good threshold to reliably transfer the annotation of interactions.

Comparison of Interolog Mapping to other methods

To compare different mapping methods, we choose C. elegans as the source organism, and map its interactions onto S. cerevisiae genome by three different mapping methods. The results of the comparison predicted interologs from the different methods to the gold standard positives and negatives are as follows:
Best-Match | Reciprocal Best-Match | Generalized Interolog (all) | Generalized Interolog (top 5% $J_E$)
--- | --- | --- | ---
Predicted | 84 | 33 | 9317 | 112
Validated | 25 | 18 | 162 | 35
Accuracy | 30% | 54% | 2% | 31%

Trade-Offs

For the method that is covered in this paper, Interolog Mapping method, anything that could be paired so it includes everything. This would be the case that we use some kind of threshold value for $J_E$ to filter out everything on the left of this threshold value and we look at the accuracy of only those protein interactions. If you do not really care of how accurate the data is and you want a high predictive power you want to pretty blast the $J_E$ value so you can predict a lot of interactions.

**Increase in $J_E$ corresponds to increase in accuracy and decrease in predictive power.**

![Graph showing distribution of generalized interologs as a function of joint E-value ($J_E$).](image)

Distribution of the number of generalized interologs as a function of joint E-value ($J_E$). The dashed line represents the number of all predictions above a given $J_E$. The solid line represents the number of true positives above a given $J_E$.

Experimental Verification

PIE (Probabilities Interactome Experimental) = 4 large-scale yeast interaction data sets
ROC curves compare generalized interolog mapping with PIE. The coverage and accuracy of generalized interolog mapping is comparable to PIE.

Summary
The higher joint sequence similarity, the higher accuracy of protein interaction transfer. We can use interolog mapping method developed in paper to predict interactions in model organisms with less-complete interaction networks.

Interaction Network Modularity

A. Background
Interaction networks are scale-free i.e. most of the proteins in the network interact with a small number of partners and a few proteins interact with many partners that are called hubs. It is resistant to random node removal i.e. the overall connectivity of the network does not change; and it is also sensitive to targeted hub removal that are few in the network.

There are two types of Hubs:

1. Party Hubs
Each hub interacts with most of their partners simultaneously, at the same time and at the same location. They perform specific functions inside module

2. Date Hubs
Each hub interacts with different partners at different times or locations and they connect modules (biological processes) together.

B. Network Construction
To recognize the hubs we need to construct the network first.

We have a bunch of interactions from different experiments and computational predictions.
Input Methods
- High-throughput yeast-2-hybrid projects
- Co-IP
- Computational predictions
- MIPS protein complexes
- MIPS physical interactions

In this procedure we want to extract high-confidence interactions in yeast; by high confidence we mean interactions should be observed by at least 2 different input methods

Results
Filtered yeast interactome’ (FYI) data set contains 2,493 high-confidence interactions, each observed by at least two different methods. The FYI network contains 1,379 proteins with an average degree of 3.6 interactions per protein and a large connected component of 778 proteins. Its degree distribution follows the power law that characterizes scale-free networks

Hub Identification
Any node (protein) with more than 5 edges \( (k > 5) \) is considered as a hub.

In this figure proteins are colored according to mutual similarity in their mRNA expression patterns. 'Party' hubs are highly correlated in expression with their partners, and presumably interact with them at similar times. The partners of 'date' hubs exhibit more limited co-expression, and presumably the corresponding physical interactions occur at different times and/or different locations.

Hub Characterization

Data Source
Filtered Yeast Interactome(FYI) hubs were characterized with mRNA gene expression profiles under five different conditions.
They take a hub and they take all of its interactions and they look at the mRNA correlation for those two proteins and this is just basically we sum the expression of gene 1 and gene 2 and then divide by the total number of experimental conditions in your mRNA gene expression profile.

**Pearson Correlation Coefficients (PCC)**

They calculate the PCC for each pair-wise interaction and average them for each hub. The distribution of PCC is very different depending on whether it is hub or it is non-hub.

\[
\rho = \frac{\sum (x_i y_i)}{N}
\]

**Hub vs. Non-Hub**

For each hub we calculate the average of Pearson correlation coefficients between the hub and each of its respective partners for mRNA expression (see Methods). Strikingly, the average PCCs of hubs, defined as nodes (proteins) with degree \( k \) greater than 5, follow a bimodal distribution in the whole compendium (Fig. 1b, red curve). In contrast, the average PCCs of non-hubs, defined as nodes with degree \( k \) of 5 or less, show a normal distribution centred on 0.1 (Fig. 1b, cyan curve, and Supplementary Fig. 4). In randomized interactome networks of the same topology (Supplementary Methods), the average PCCs of hubs also show a normal distribution centred on 0 (Fig. 1b, black curve). This bimodal distribution suggests that hubs can be split into two distinct populations: one with relatively high average PCCs (party hubs) and the other with relatively low average PCCs (date hubs) (Supplementary Information). Furthermore, the bimodal distribution suggests a natural boundary for separating or partitioning date hubs from party hubs.

**PCC Distribution**

We apply different conditions to the cell.

![Probability densities of the average PCCs were calculated from a global expression profiling compendium. Average PCCs were also independently calculated for each condition constituting the](image-url)
compendium. The number $n$ in each panel refers to the number of data points for each gene for each condition. Average PCCs for hubs in the FYI (red curve) show a clear bimodal distribution that is used to separate date and party hubs (located by the arrow) for the conditions shown in the top panels. 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. No bimodal distribution is observed with the average PCCs of non-hub proteins (cyan curve) or for hubs in randomized networks (black curve).

Party hubs are those with an average PCC higher than the threshold. All other hubs were defined as date hubs.

**Prediction of Date vs. Party Hub**

### Yeast Expression Compendium

Superset of data for all external conditions

Bi-modal: suggests we can partition date hubs from party hubs

#### Yeast Expression Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Data Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pheromone treatment</td>
<td>45</td>
</tr>
<tr>
<td>Sporulation</td>
<td>10</td>
</tr>
<tr>
<td>Unfolded protein response</td>
<td>9</td>
</tr>
<tr>
<td>Stress response</td>
<td>174</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>77</td>
</tr>
</tbody>
</table>

**Absence of clear bi-modal**

**Presence of clear bi-modal**

#### Date/Party Partition

Party Hubs = nodes with average PCC > cutoff in $\geq 1$ conditions

**In Silico Node Removal**

Knowing that we can look at the different effects on the network topology due to party hub removal or date hub removal. The idea is that when you remove a party hub the connectivity of the graph is not really affected.

- **Effect on Path Connectivity**

  Characteristic path length is the average shortest path length or average of the number of edges between node pairs. If one node is removed, we observe change in characteristic path length. When you remove party hubs, the connectivity of the network is not affected; but when you remove date hubs, the connectivity of the network decreases because they are the high level connectors of individual modules.

- **Effect on Remaining Components**

  Another thing that they want to test is that they have a hypothesis that everything close to a party hub
participate in the same biological process. After node removal for party hubs, main component are much larger than the remaining main components after removing a date hub.

The effects on the characteristic path length of the network on gradual node removal. Random removal of nodes (‘failures’) is represented by the green line, attacks against all hubs by the brown line, attacks against party hubs by the blue line, and attacks against date hubs by the red line. The 'breakdown point' is the threshold after which the main component of the network starts disintegrating.

**Date Hub Subnetworks**

The main component of the FYI network splits into small subnetworks after the removal of date hubs, whereas it stays almost intact after the removal of party hubs.

Each subnetwork has a tendency to be homogeneous in function

Subnetworks $\longleftrightarrow$ biological modules

We can assign a ‘most likely’ function for each subnetwork by examining functional annotation

**Genetic Interactions**

- Organized modularity model predicts that genetic perturbations of party hubs should differ from those of date hubs
Genetic Perturbation

- Date hubs and party hubs are comparable in terms of functional essentiality.
- Date hubs have more genetic interactions than party hubs.

b. Date and party hubs are both more likely to be essential than non-hubs, but their single knockout affects cellular viability to the same extent. c. Date hubs participate in more genetic interactions than party hubs or non-hubs, as measured by genetic interaction density (GID) based on genetic interactions gathered at MIPS14.

Summary

In silico investigation and genetic interaction analysis both describe a protein interaction model where:

i. There is an organized modularity
ii. Date hubs act as module connectors
iii. Party hubs function at a lower level within modules.

We can use this prediction method to classify and organize other interactomes into a modular network. Identification of party and date hubs may provide insight into potential drug targets.