

Master Module
Proteinbiochemistry and Bioinformatics
November 2025

Session: Protein interaction networks

4. Graph-theoretical aspects of protein interaction networks

How can I use protein interaction data in biological research?

What is the function of my gene of interest?



Is the protein of my interest part of a protein complex?

Can I find new protein complexes?



I found 20 genes in my screen that rescued phenotype X:

- do these genes work in the same biological process?
- are these genes part of the same protein complex?
- > do these proteins (tend to) interact with each other?



My protein has many interaction partners,
does it mean that it is of functional importance?



How can I use protein interaction data in biological research?

Resources for protein interactions

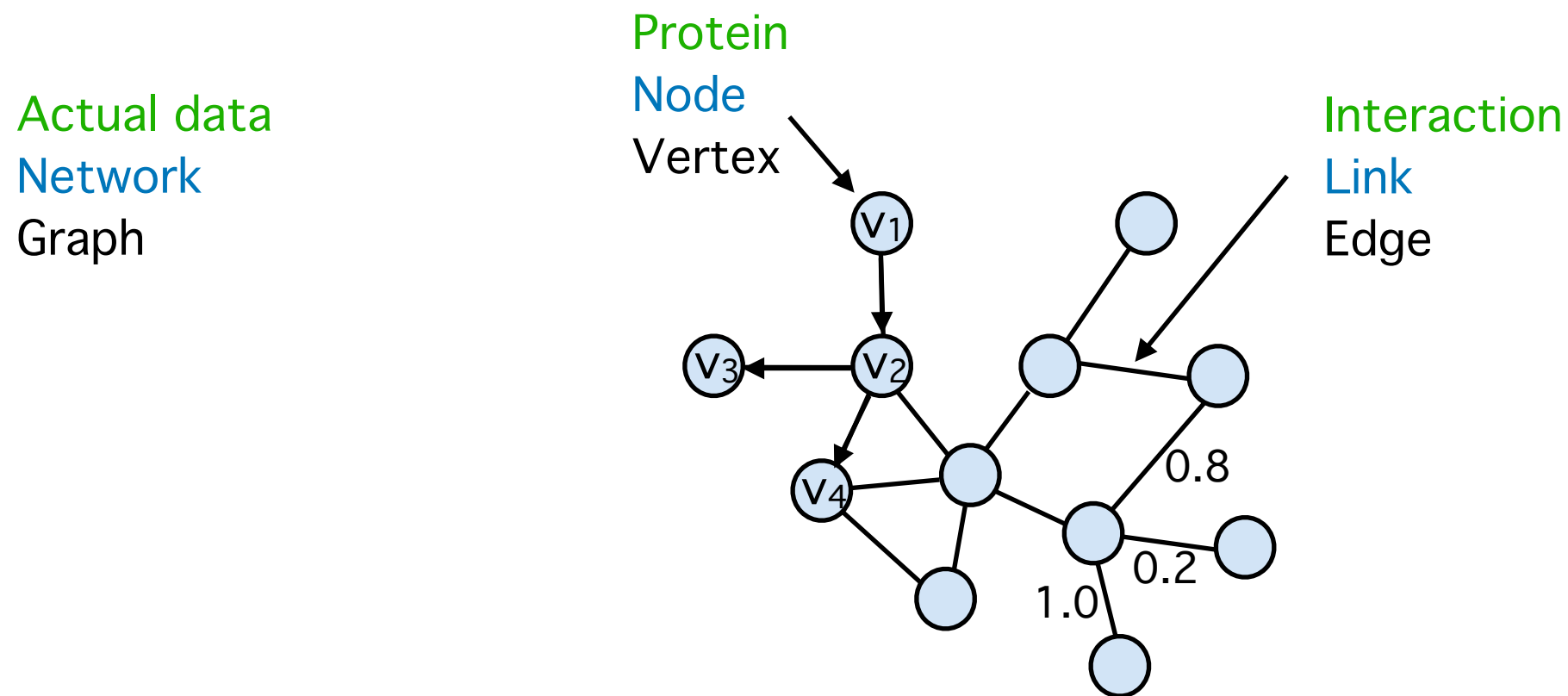


Methods to analyze protein interaction data



Graph theory

Protein interaction data as a graph



$$V = \{v_1, v_2, v_3, v_4, \dots\}$$

$$E = \{(v_1, v_2), (v_2, v_3), (v_2, v_4), \dots\}$$

- undirected vs directed graph
- weighted vs unweighted graph

Degree, average degree, and degree distribution

- Degree: number of edges of a vertex
(i.e. number of interactions of a protein)

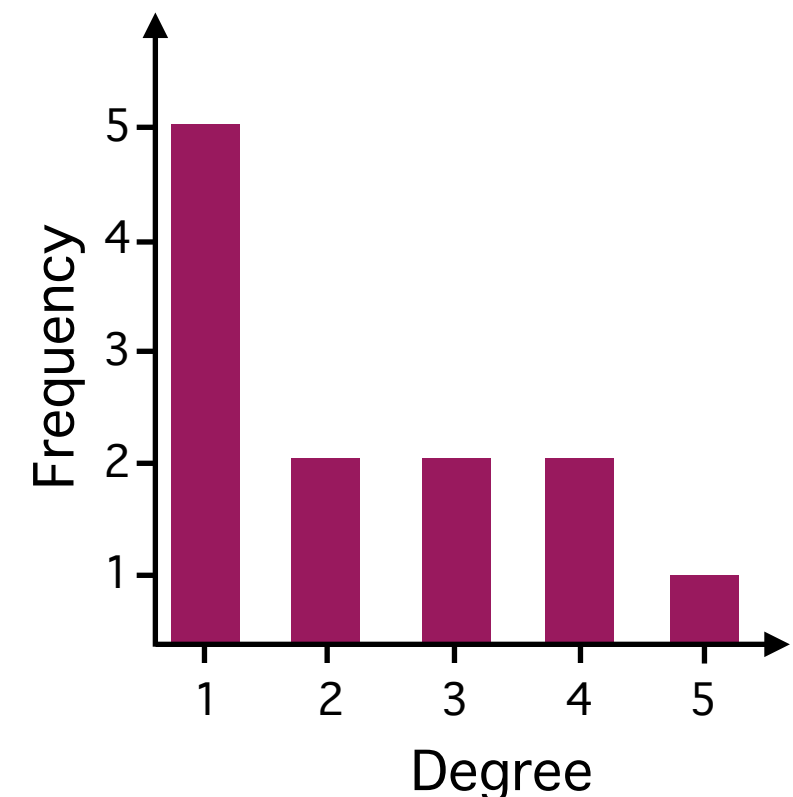
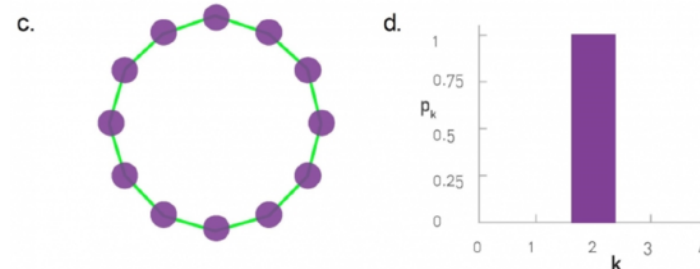
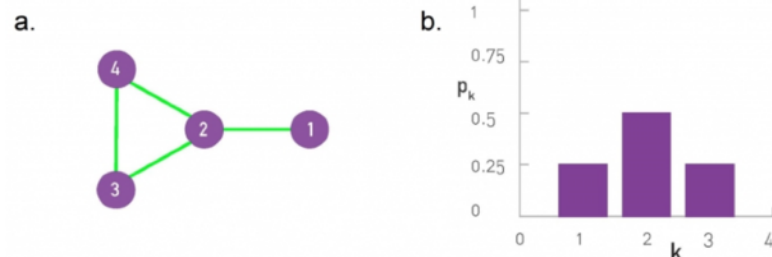
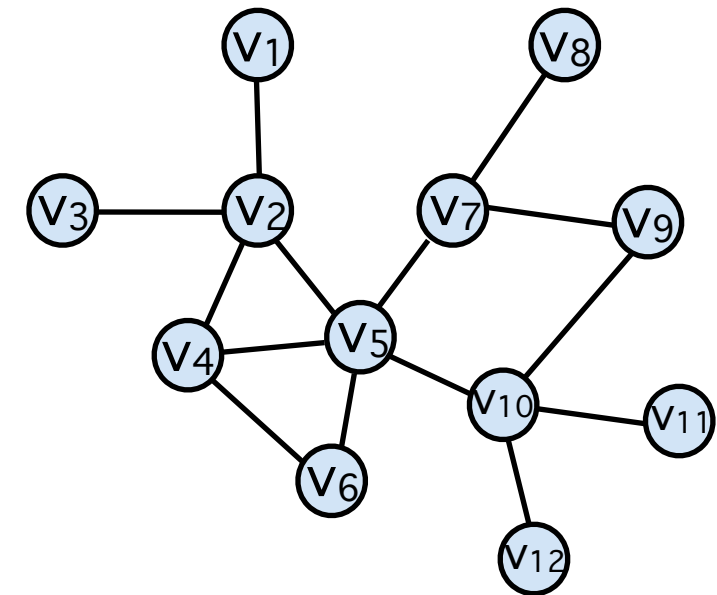
$$k_1 = 1, k_2 = 4, k_4 = 3$$

-> k_i is the degree of vertex v_i

- Average degree
-> network property

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i \quad N = \text{number of vertices in graph}$$

- Degree distribution
-> network property, informs about the topology of the network

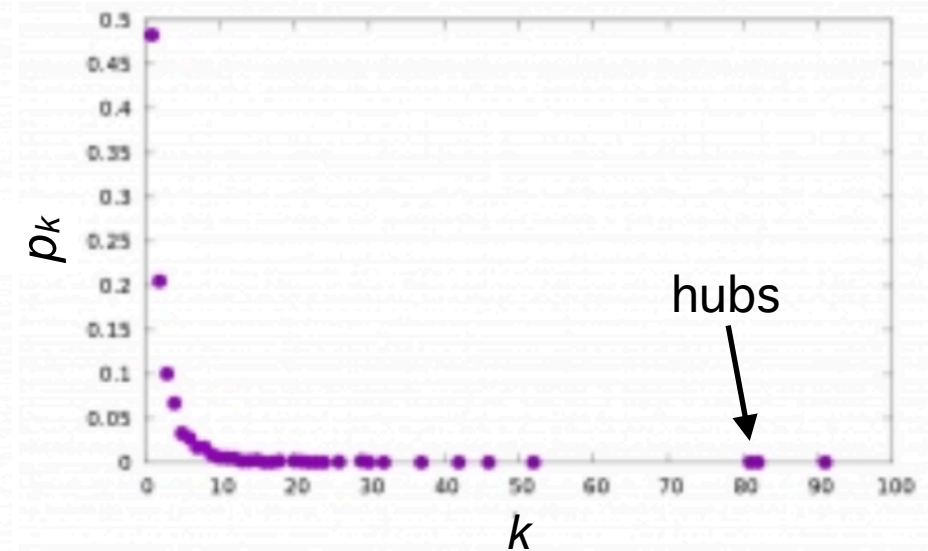


Degree distributions of biological networks

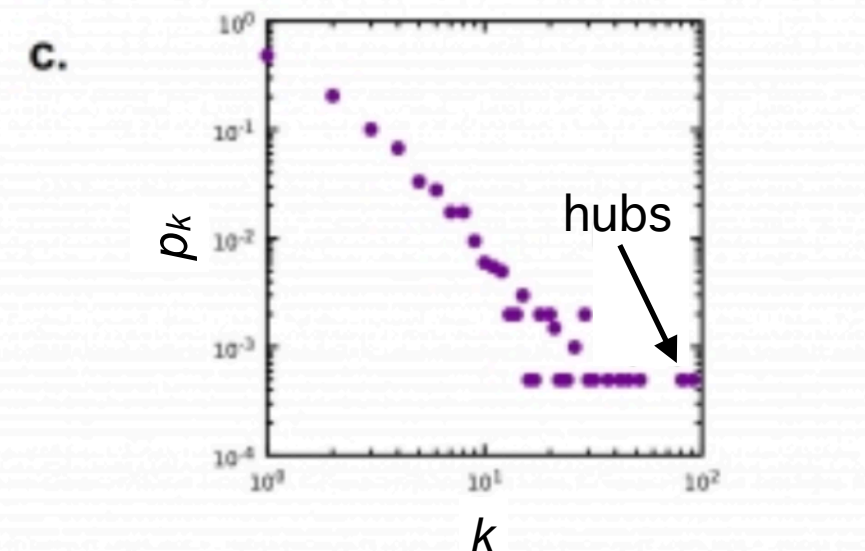
Yeast protein
interaction
network



Degree distribution - normal scale

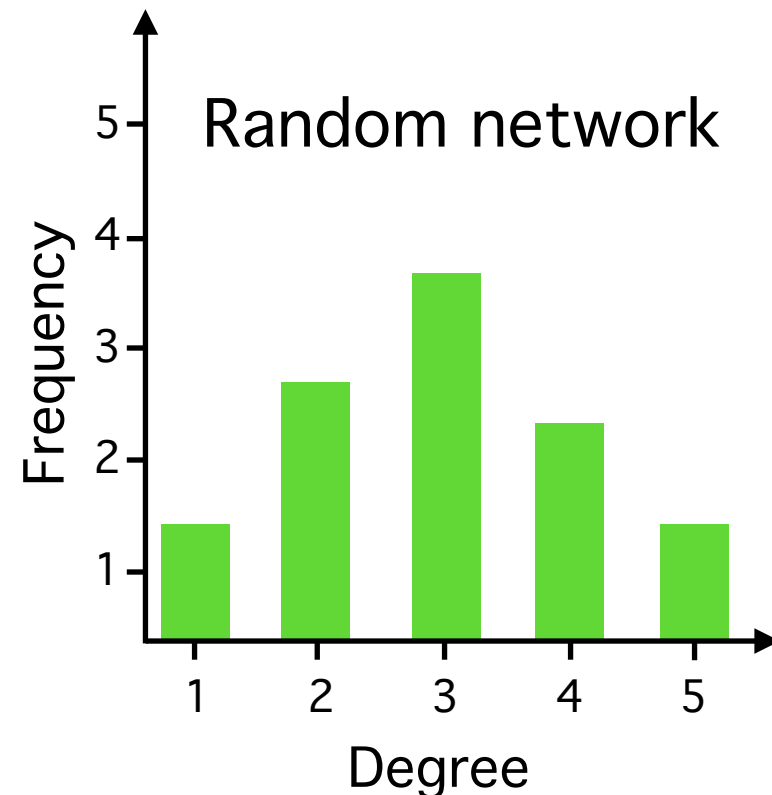
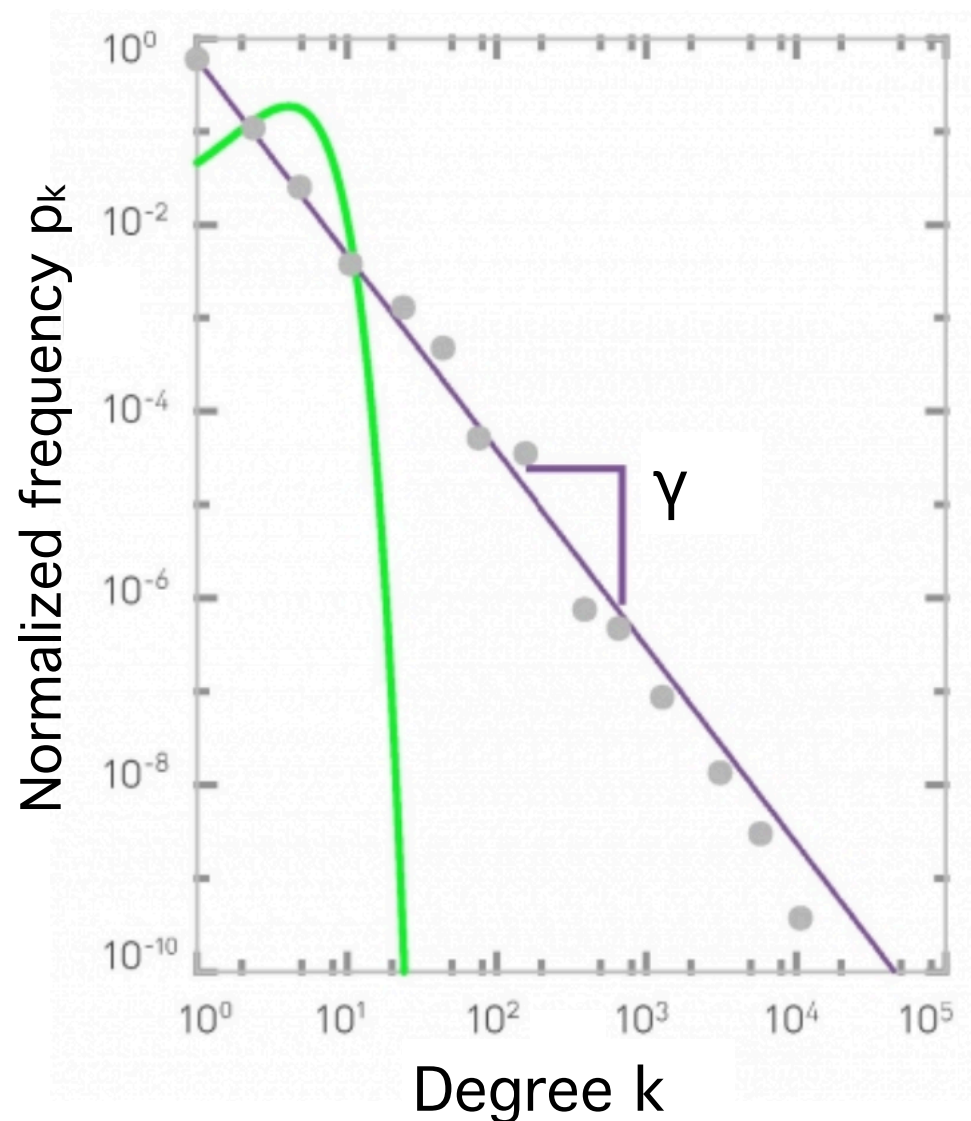


Degree distribution - log-log scale



Degree distributions of biological networks

Degree distributions of many real world networks follow a power law distribution in log-log scale



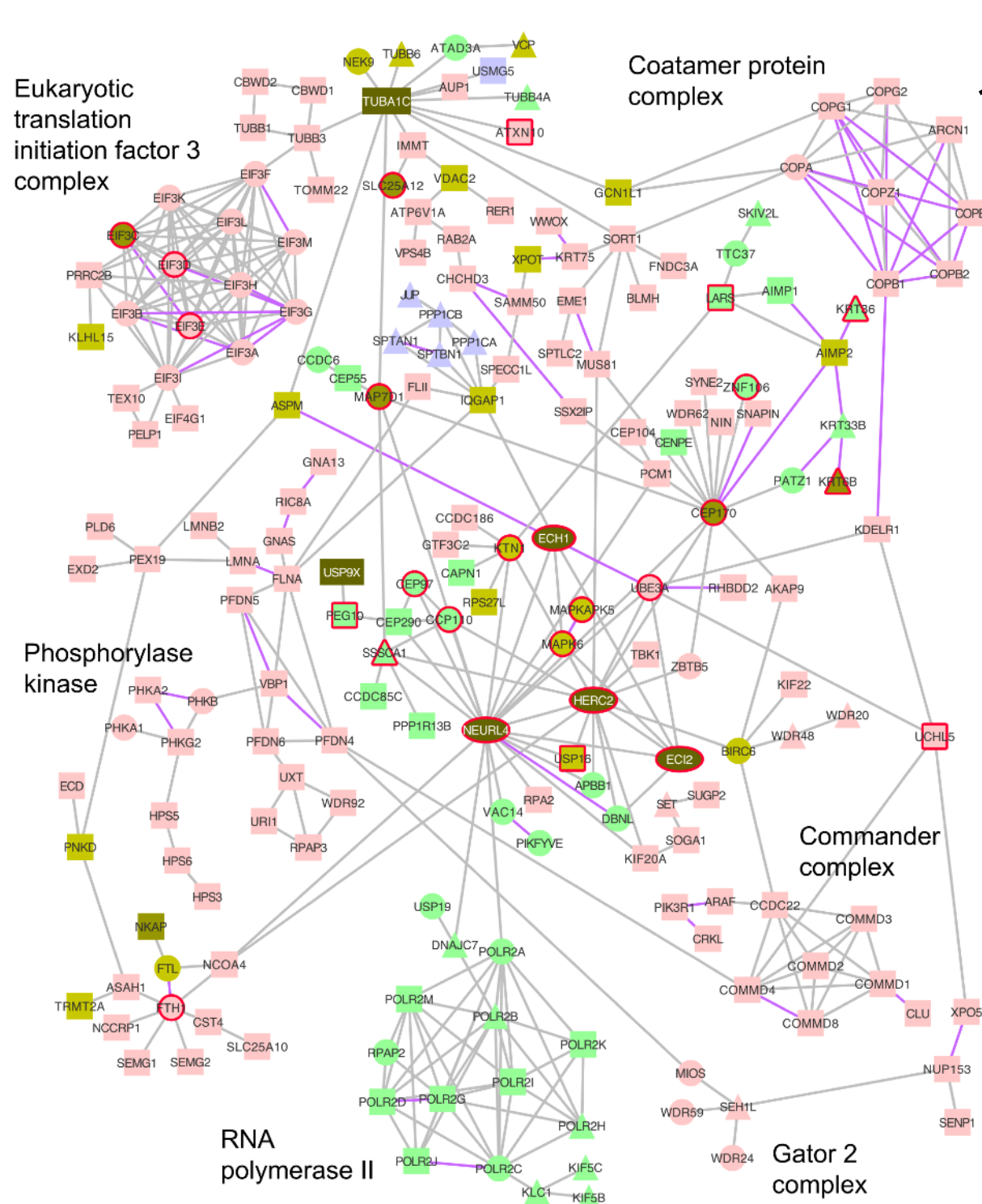
Power law distribution

$$p_k \sim k^{-\gamma}$$

$$\log p_k \sim -\gamma \log k$$

Networks whose degree distribution follows a power law, are called scale-free.
Most biological networks are scale-free.

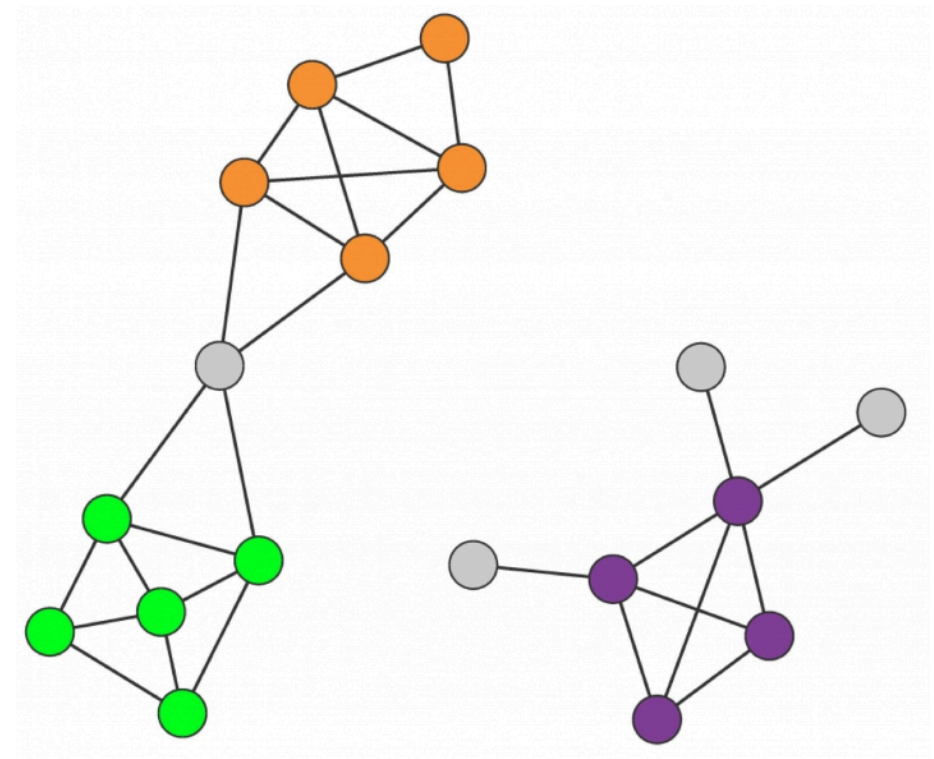
Finding communities in graphs



Protein complexes show as clusters in a network

Communities are locally dense connected subgraphs in a network

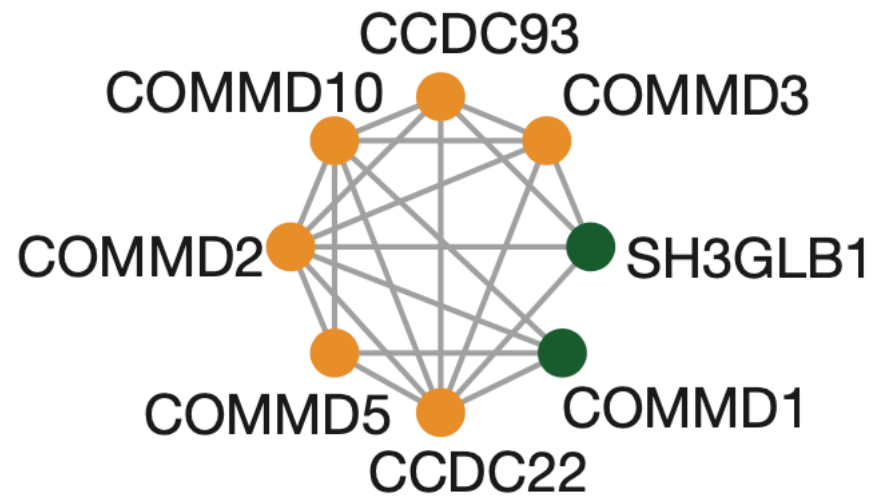
Vertex of a community is more linked to other vertices of that community than to vertices outside



Numerous algorithms exist to find communities in a graph

Can I find new protein complexes or complex members?

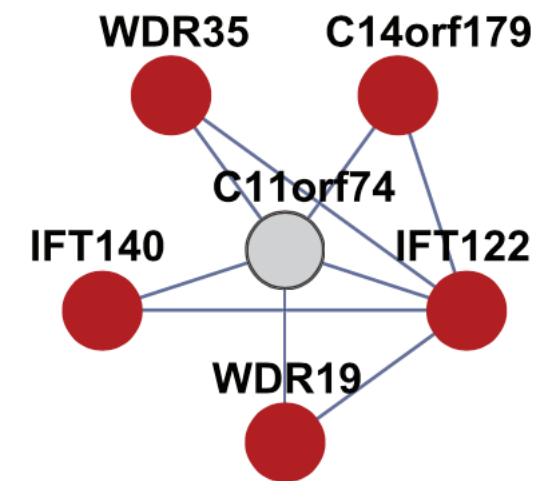
Identification of Commander complex



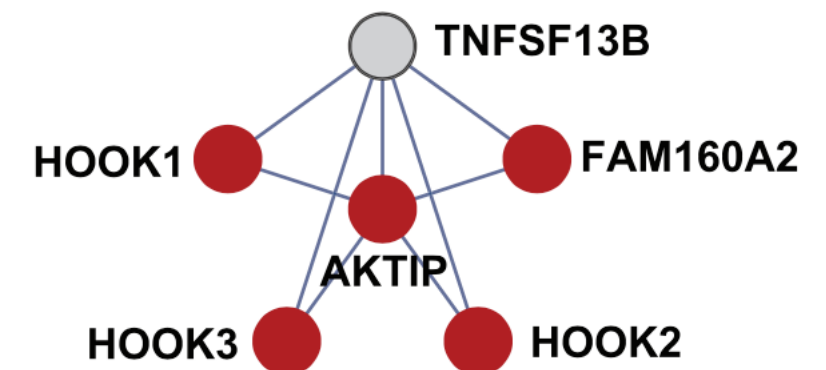
↓
Role in embryonic development

Identification of new complex members

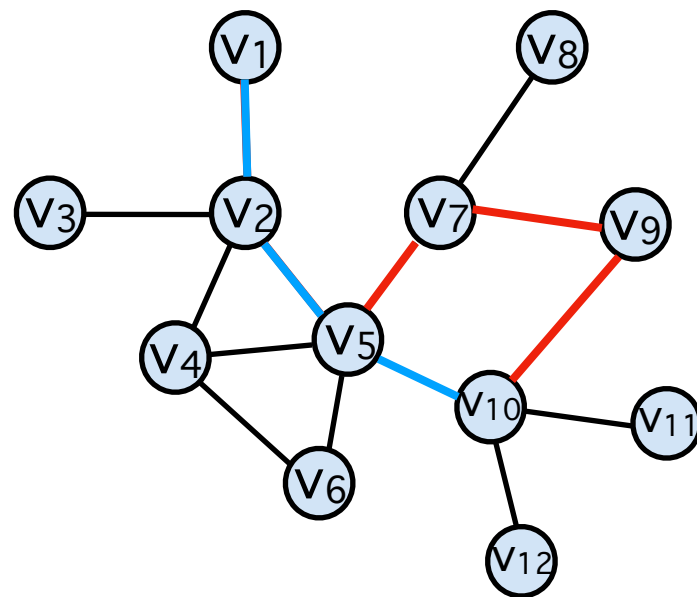
Intraciliary Transport Particle A



FHF complex



Shortest paths in graphs and betweenness centrality



A path between two vertices is formed by the edges that lead from one vertex to the other.

A path from v_1 to v_{10}

Shortest path d from v_1 to v_{10}

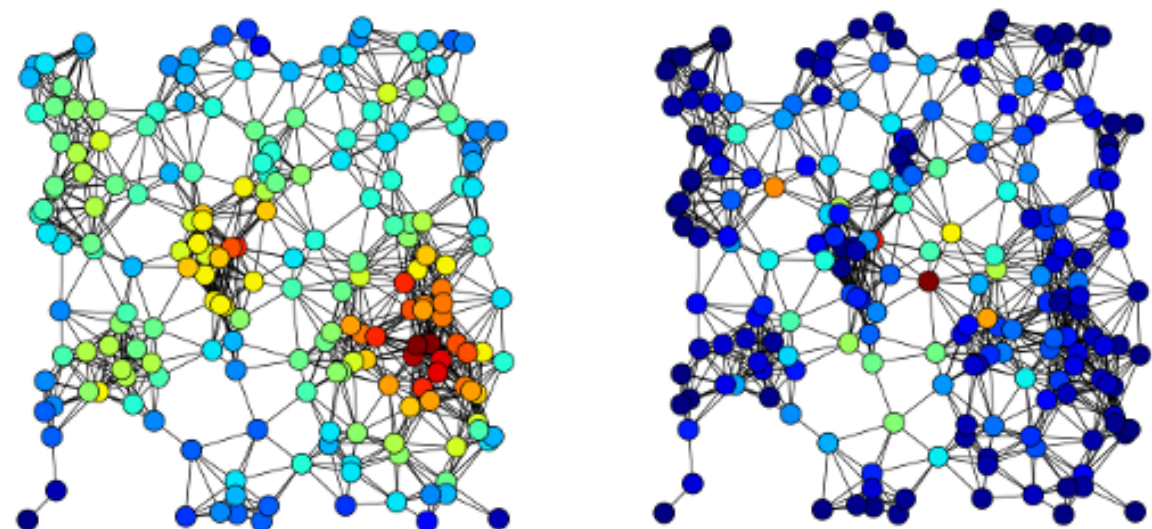
-> a path can represent information flow in a graph

How many shortest paths cross a vertex? $\longrightarrow$ Node betweenness

How many shortest paths go over an edge? $\longrightarrow$ Edge betweenness

High degree $\neq$ high betweenness

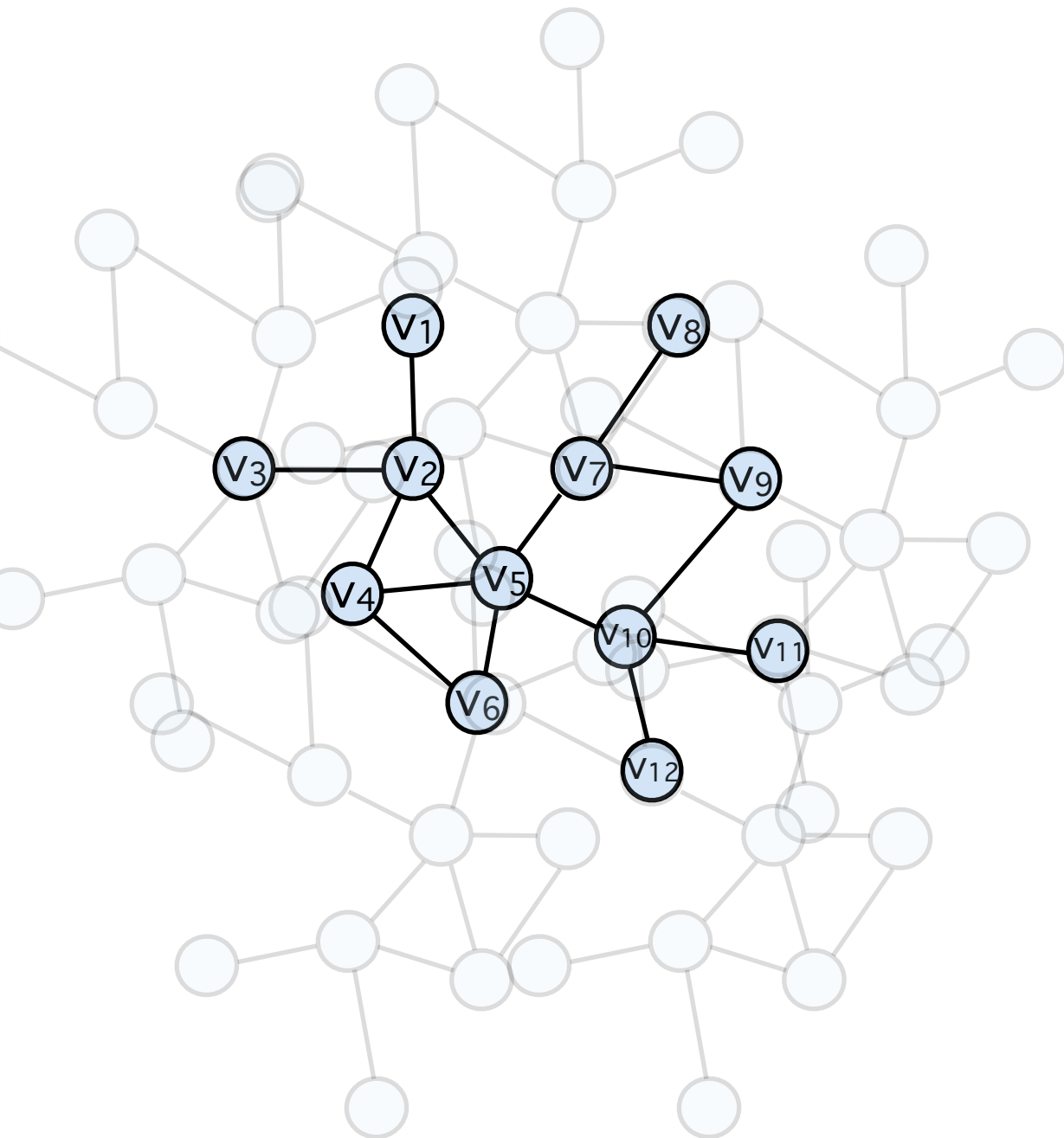
High betweenness
 $\downarrow$
Important for system



Measuring closeness in networks

Do candidate proteins from my screen tend to interact with each other?

-> count number of edges between vertices that are candidate proteins
or
calculate average shortest path between them:



How close are all the vertices v_1 to v_{12} to each other?



Calculate the average shortest path:

$$L_G = \frac{1}{N \cdot (N-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^N d_{i,j} \quad N = 12$$

Randomizing graphs to compute significances

Do candidate proteins **tend** to interact with each other?

Number of edges: 14
Average shortest path: 2.17

How close would be **12 randomly selected proteins** in the network?

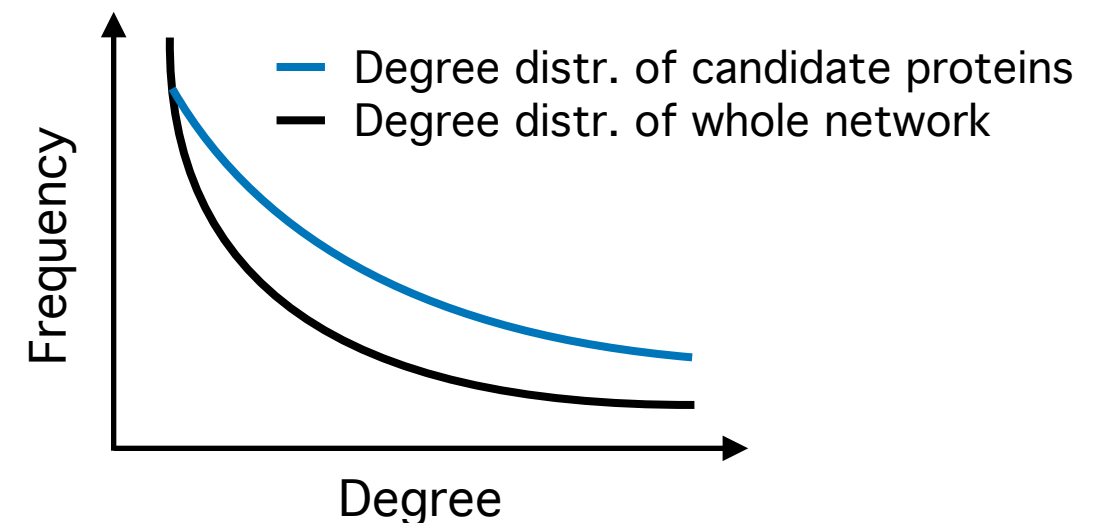
Can I randomly choose any 12 proteins in the network?

Need to randomly choose 12 proteins with the same degree distribution like candidate proteins

Hard



Possible scenario



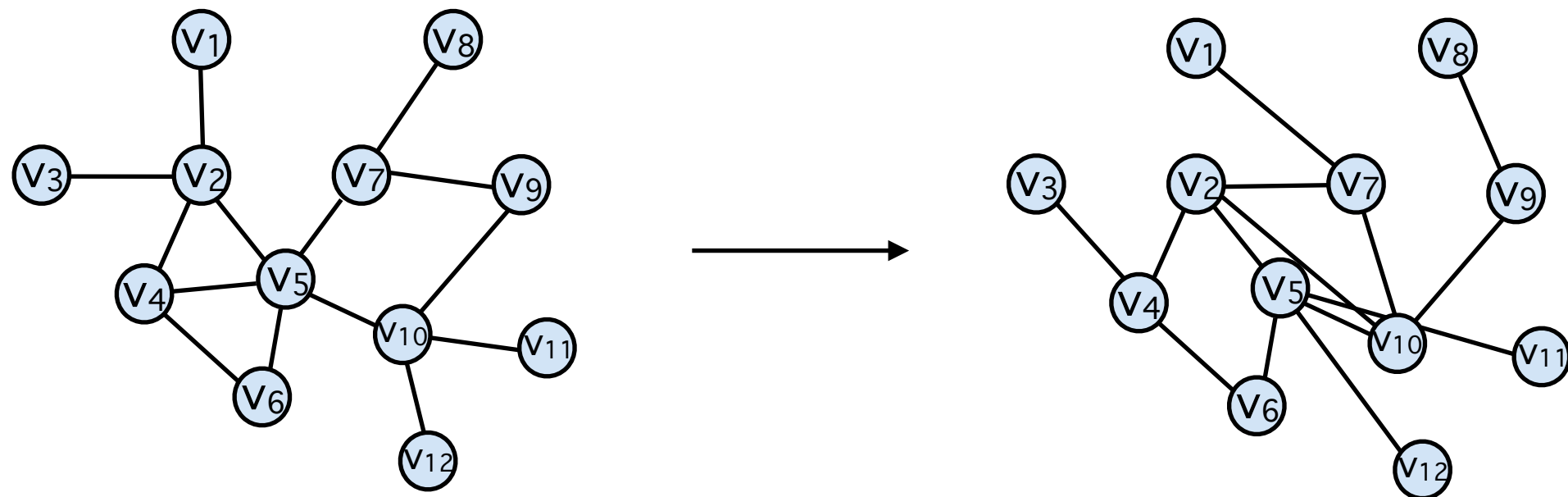
Randomizing graphs to compute significances

Need to randomly choose 12 proteins with the same degree distribution like candidate proteins

Hard



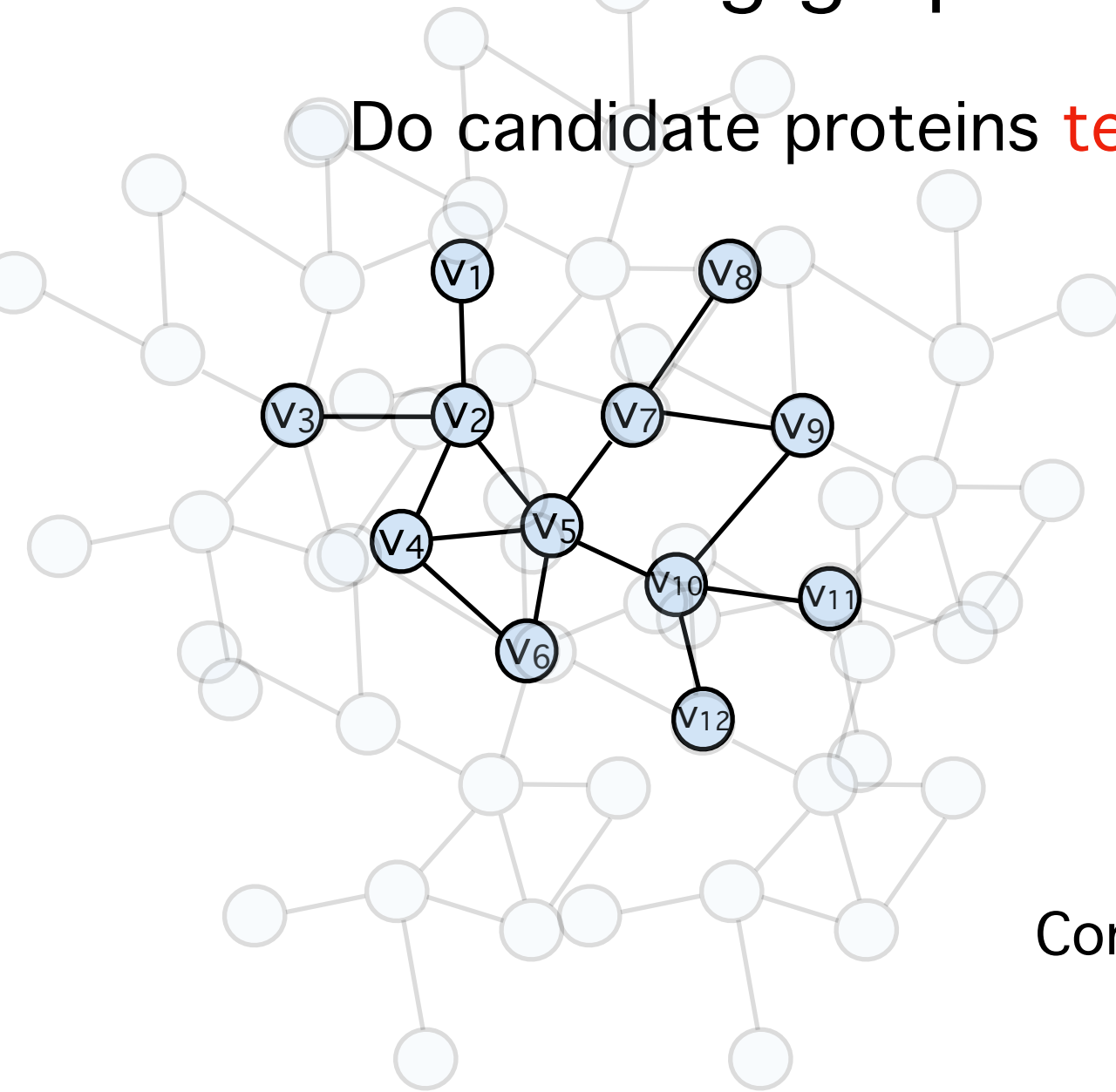
Solution: Randomize network instead - in a degree-controlled way



Edges are shuffled such that every vertex maintains its degree

Randomizing graphs to compute significances

Do candidate proteins **tend** to interact with each other?



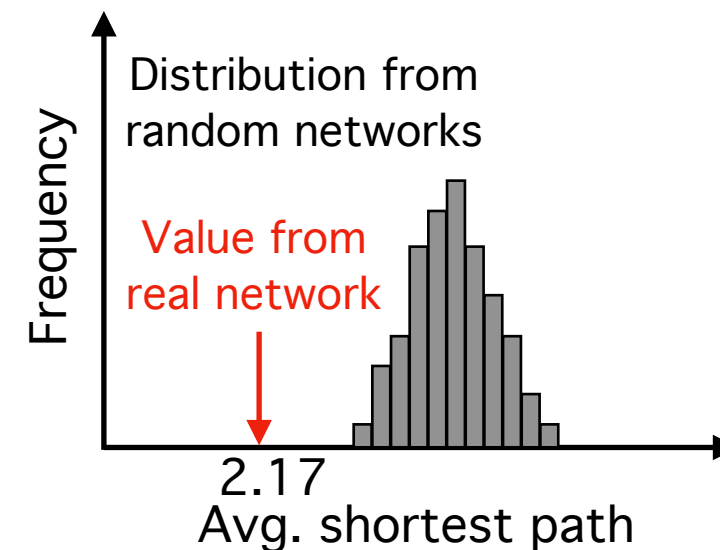
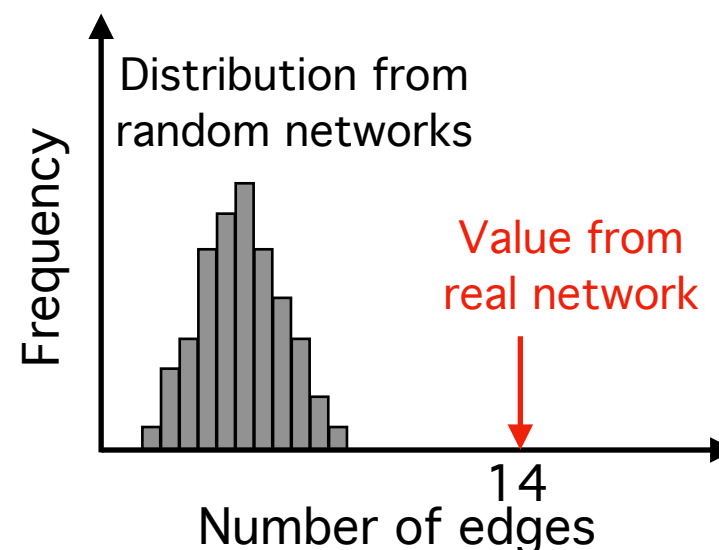
Number of edges: 14
Average shortest path: 2.17



Generate a high number of
degree-controlled randomized
networks

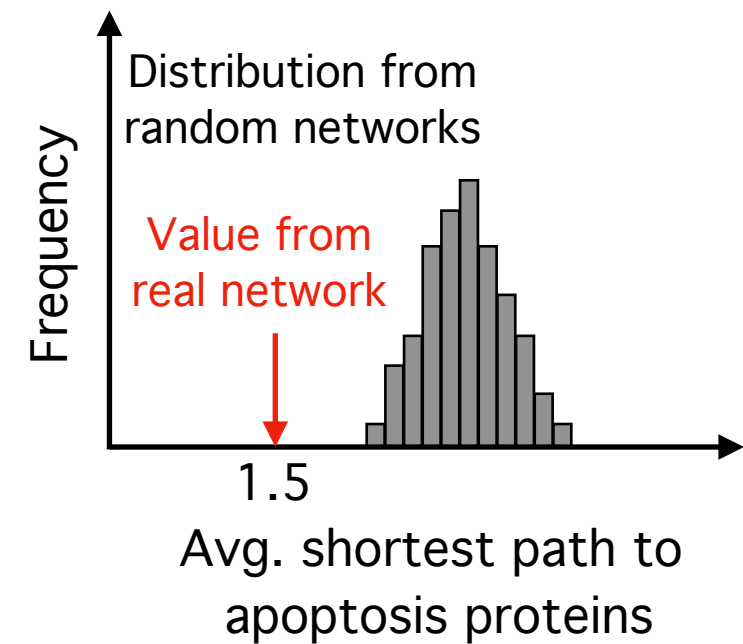
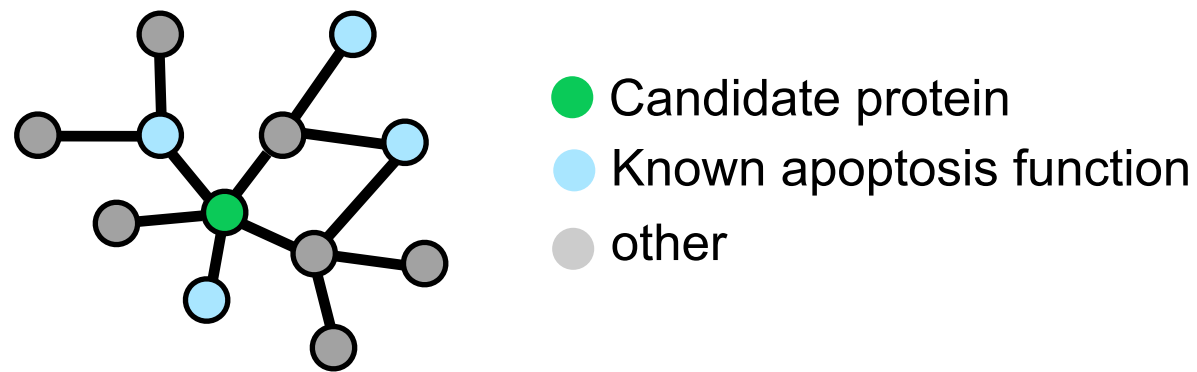


Compute closeness of candidate proteins
in each of them

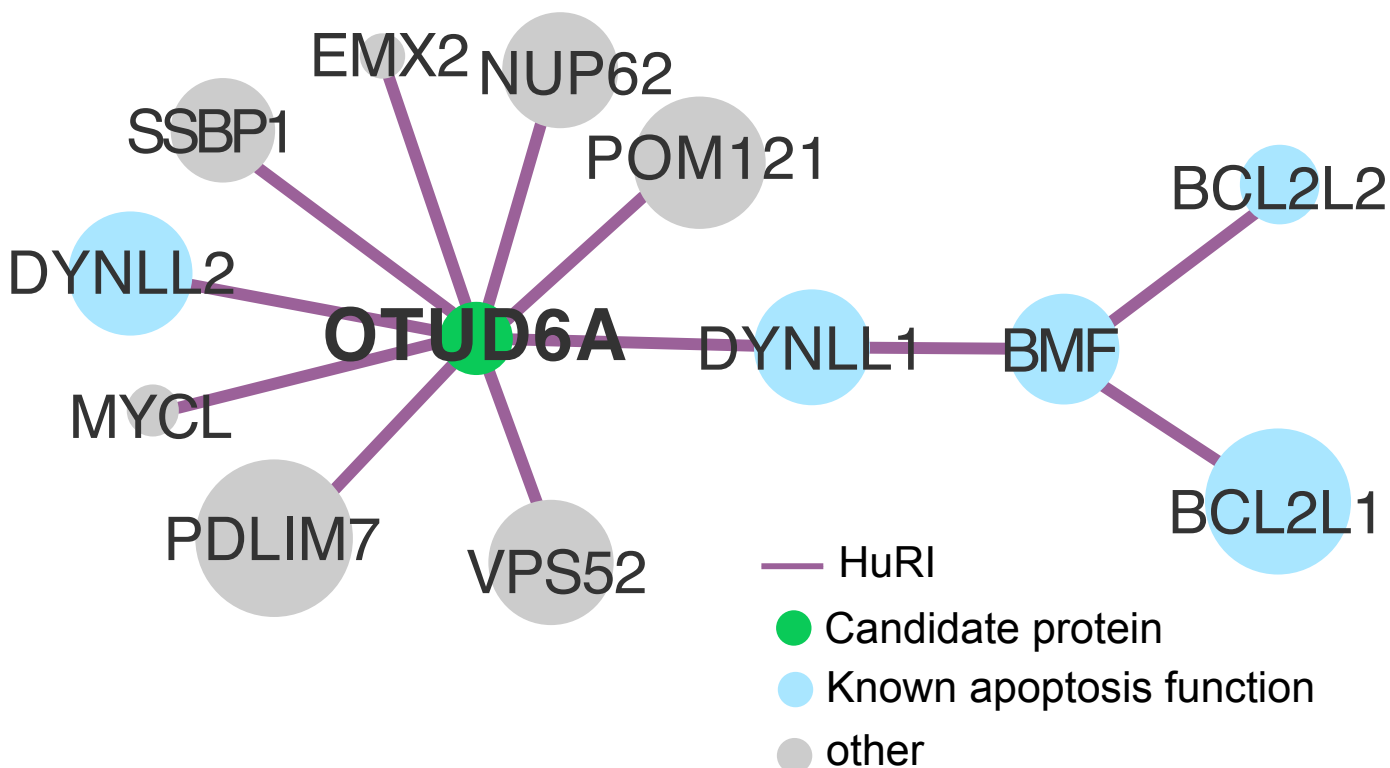


What is the function of my gene of interest?

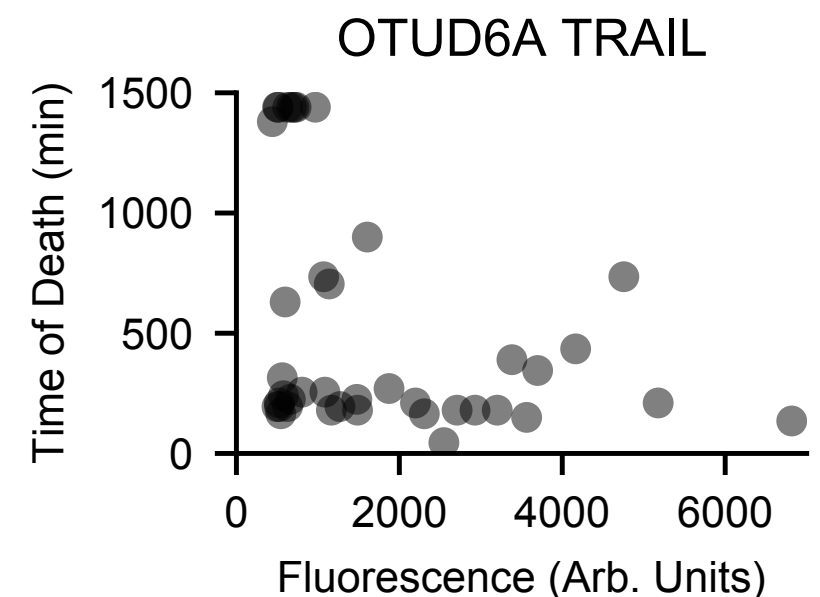
Guilt-by-association



OTU deubiquitinase 6A



OTUD6A expression results in earlier cell death



Summary

- Molecular interaction data can be represented as graphs
- Graph properties can indicate biological properties of proteins and interactions
- Biological networks are scale-free
- Use degree-controlled randomized networks to look for trends
- Guilt-by-association is a method to predict functions of proteins using interaction data