Master Module
Proteinbiochemistry and Bioinformatics
March 2022

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 ✓

How can I use protein interaction data in biological research?

Resources for protein interactions ✓

Methods to analyze protein interaction data ✗

How can I use protein interaction data in biological research?

Resources for protein interactions ✓

Methods to analyze protein interaction data ✗

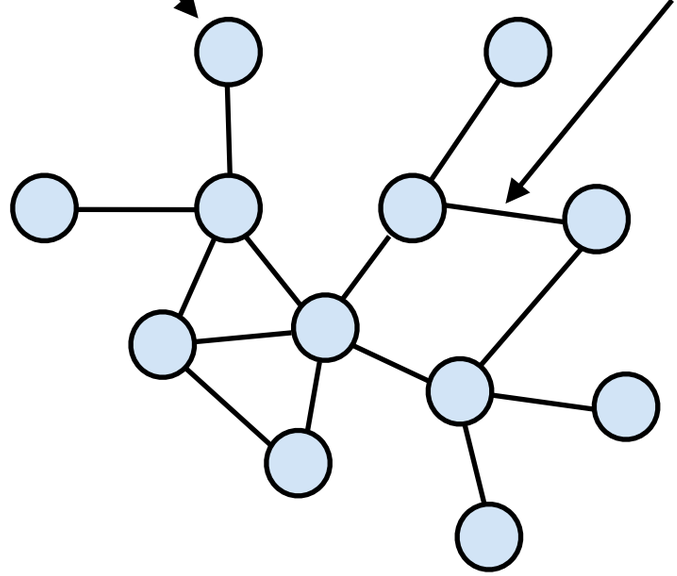
↓

Graph theory

Protein interaction data as a graph

Actual data
Network
Graph

Protein
Node
Vertex



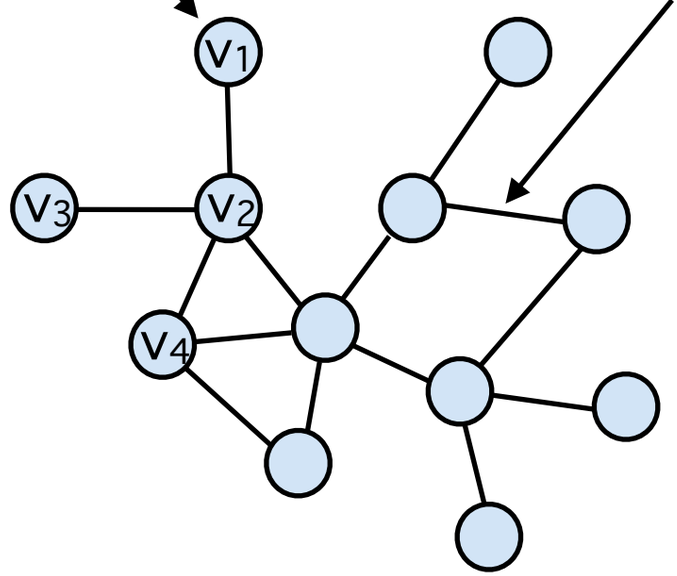
Interaction
Link
Edge

Protein interaction data as a graph

Actual data
Network
Graph

Protein
Node
Vertex

Interaction
Link
Edge



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

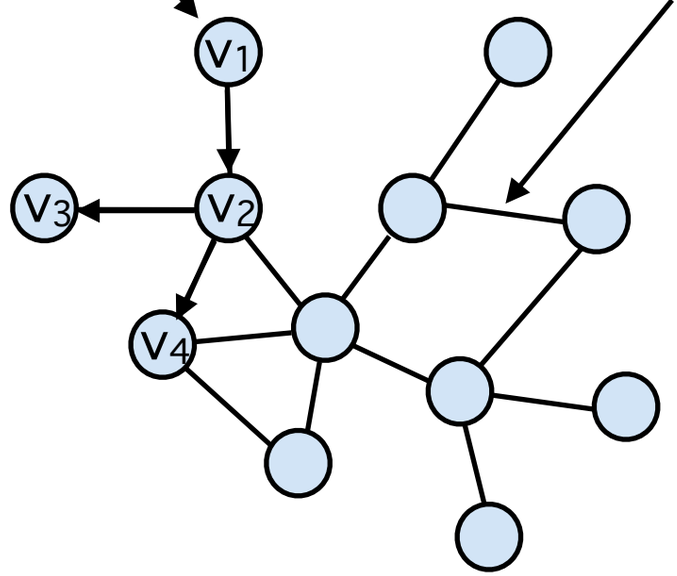
$E = \{(v_1, v_2), (v_2, v_3), (v_2, v_4), \ldots\}$

Protein interaction data as a graph

Actual data
Network
Graph

Protein
Node
Vertex

Interaction
Link
Edge



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

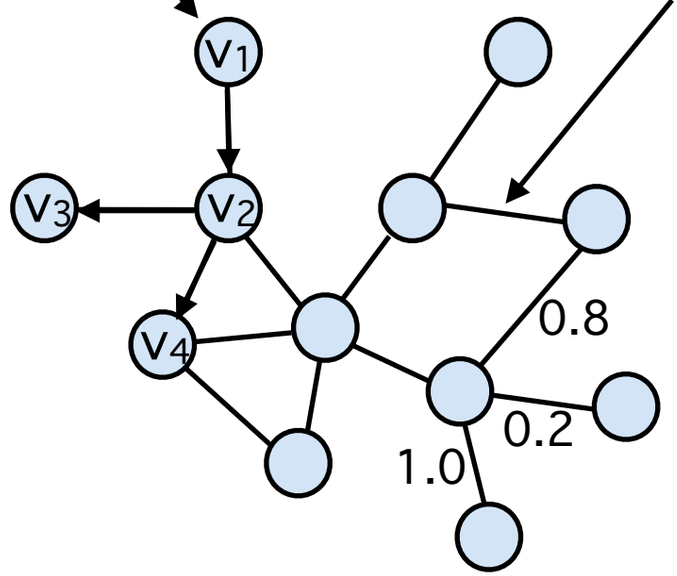
- undirected vs directed graph

Protein interaction data as a graph

Actual data
Network
Graph

Protein
Node
Vertex

Interaction
Link
Edge



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

- undirected vs directed graph
- weighted vs unweighted graph

Degree, average degree, and degree distribution

Degree, average degree, and degree distribution

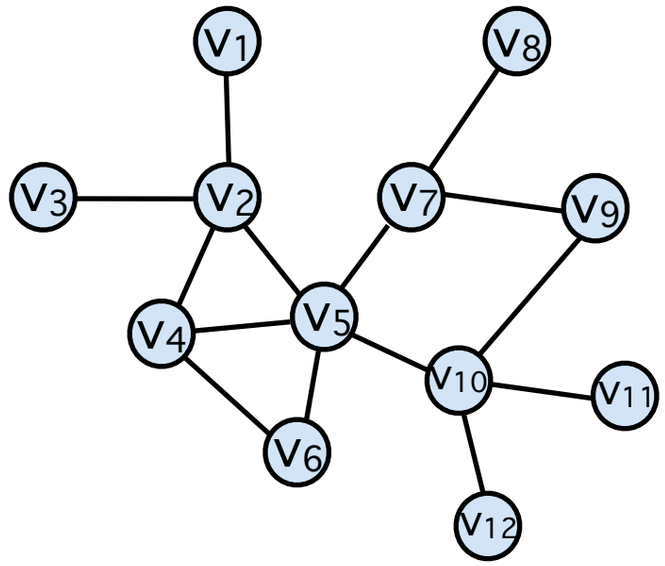
- Degree: number of edges of a vertex
  (i.e. number of interactions of a protein)

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

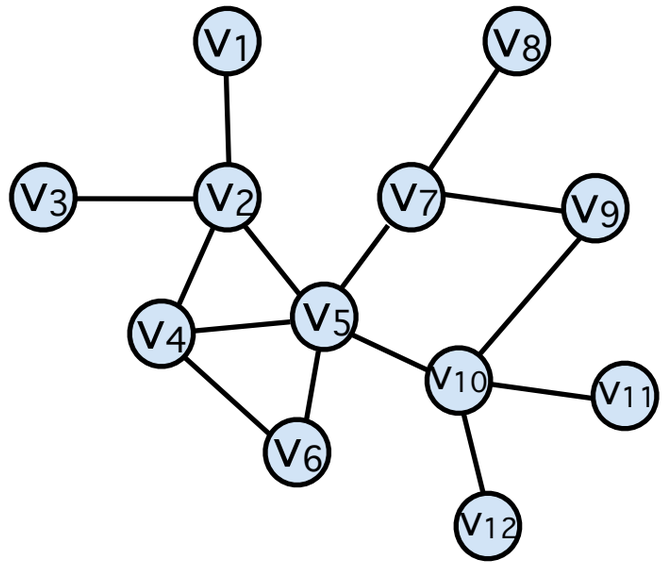


Degree, average degree, and degree distribution

- 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

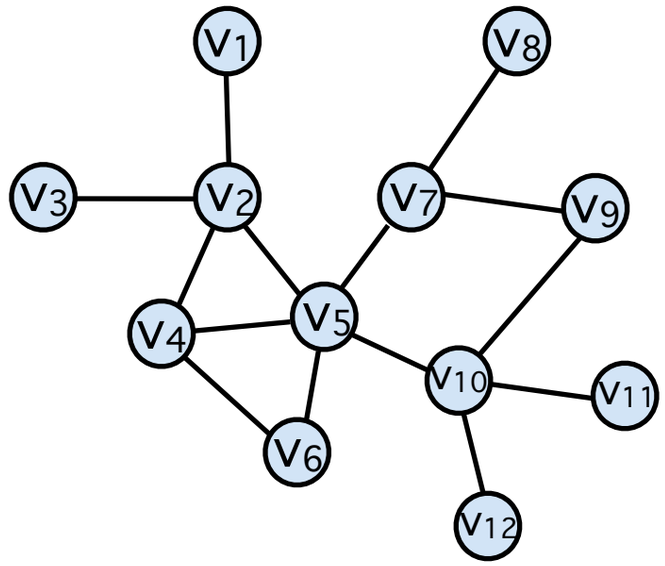
$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \qquad N = \text{number of vertices in graph}$$

Degree, average degree, and degree distribution

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

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

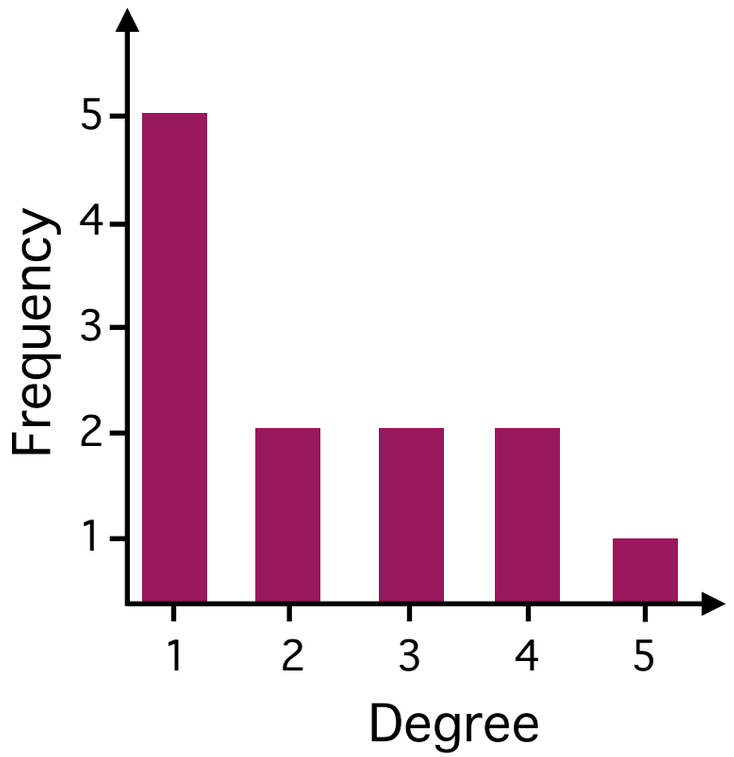
  -> k_i is the degree of vertex v_i



- Average degree
  -> network property

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

- Degree distribution

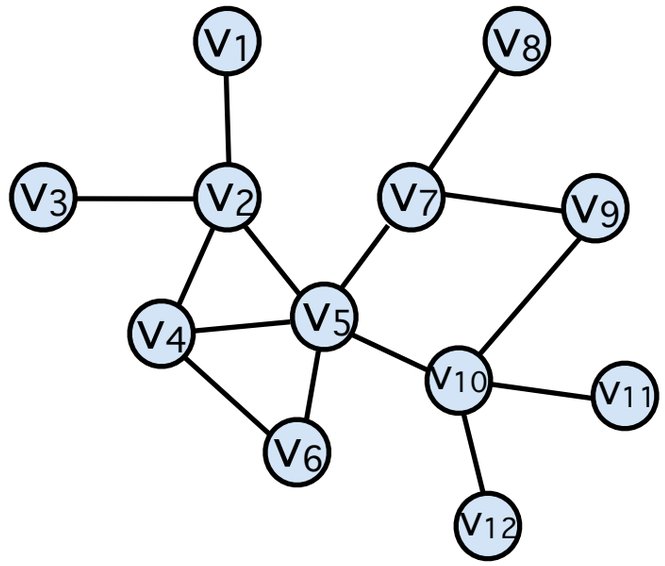


Degree, average degree, and degree distribution

- 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

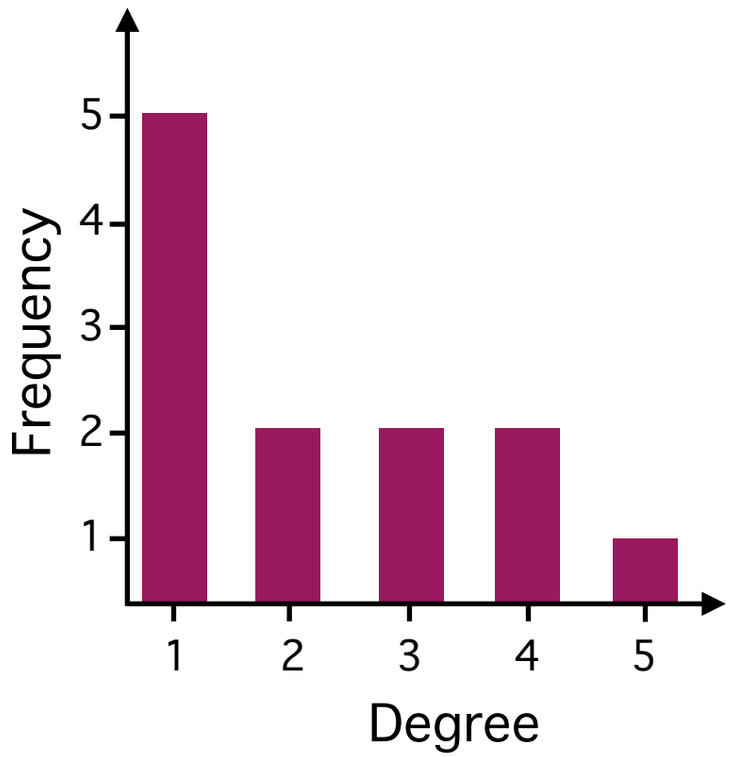


- Average degree
  -> network property

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

- Degree distribution

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

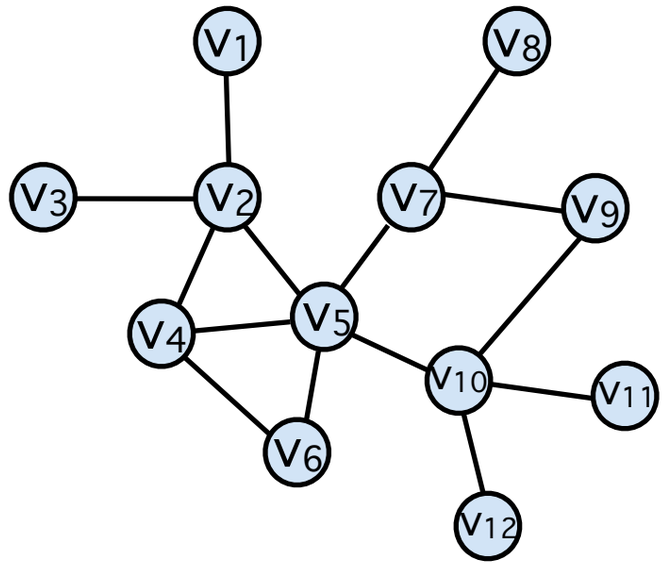


Degree, average degree, and degree distribution

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

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

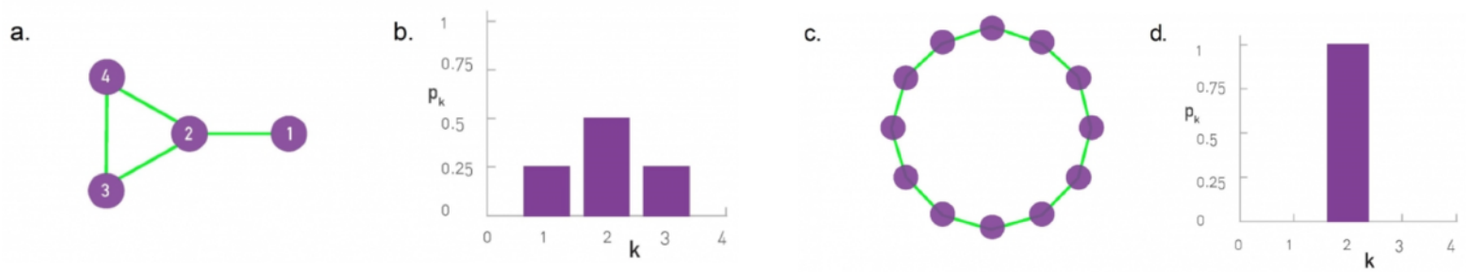
  -> k_i is the degree of vertex v_i

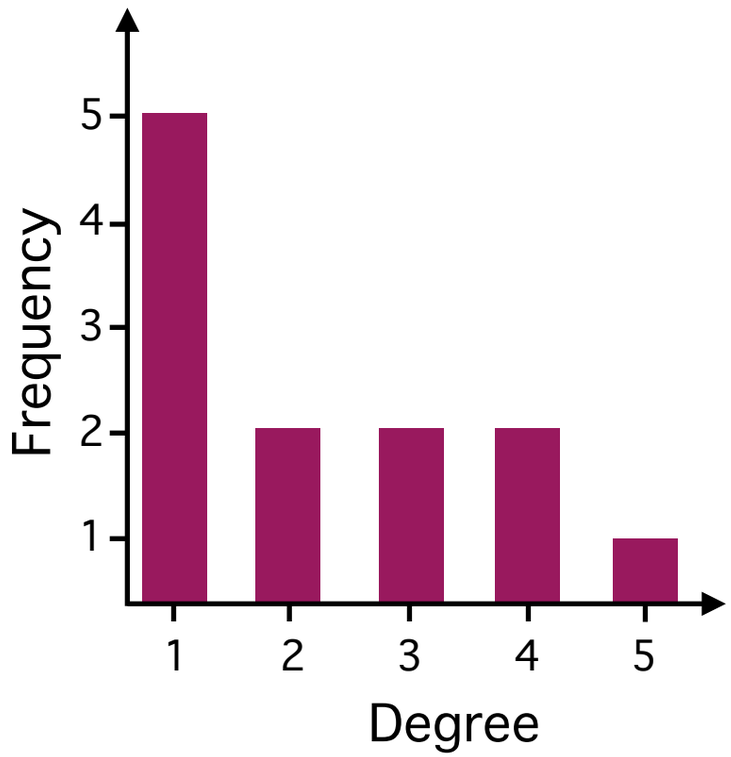


- Average degree
  -> network property

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

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





networksciencebook.com

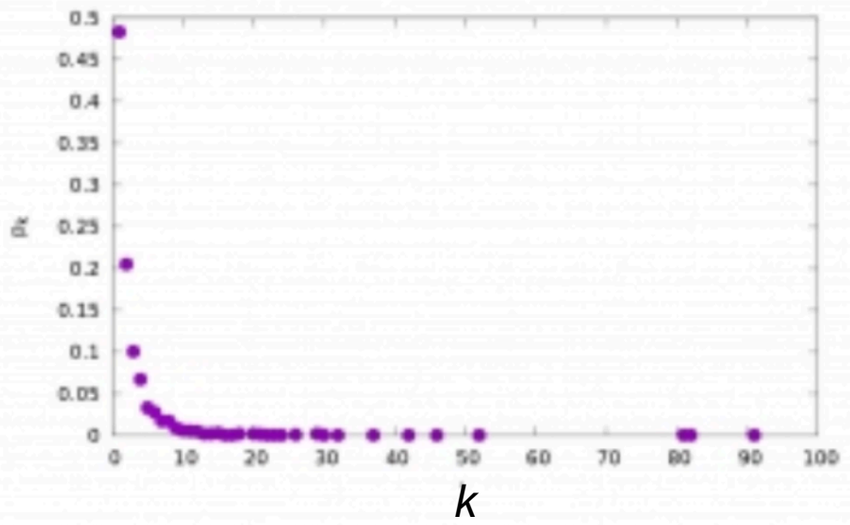
Protein interaction networks are scale-free

Yeast protein interaction network

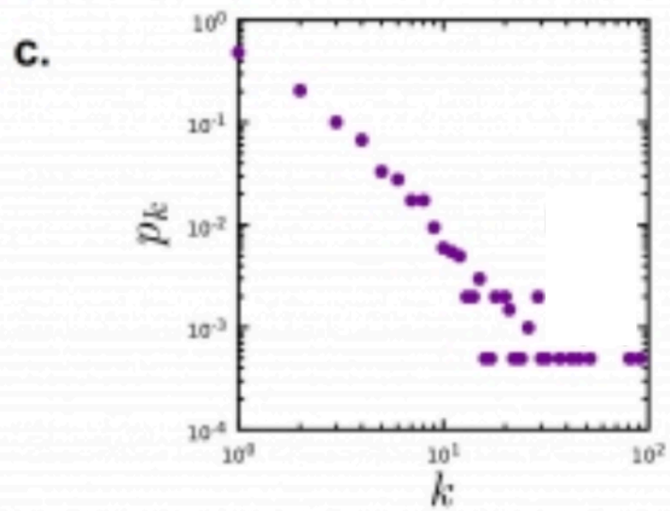
Protein interaction networks are scale-free

Yeast protein interaction network

Degree distribution - normal scale

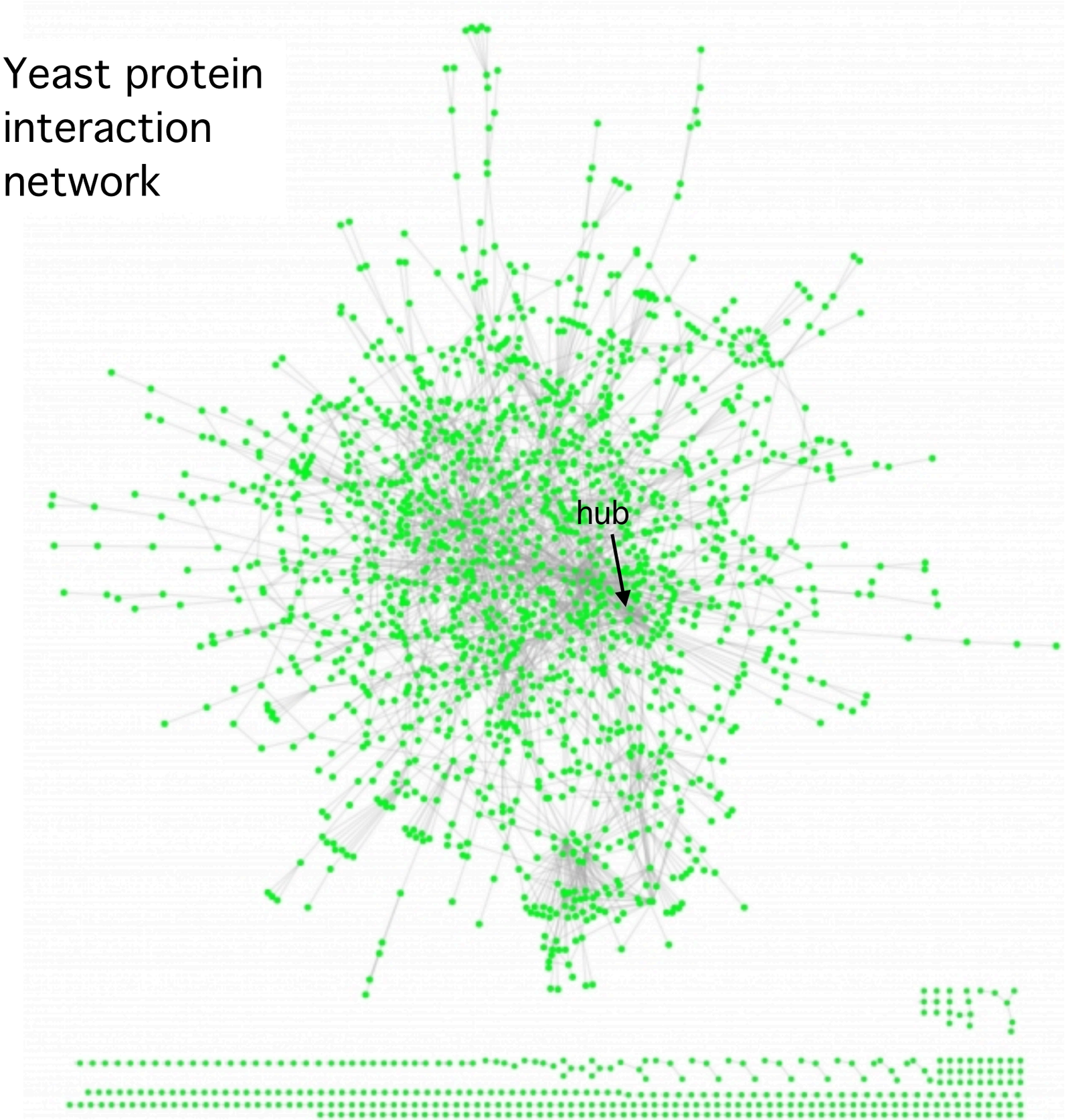


Degree distribution - log-log scale

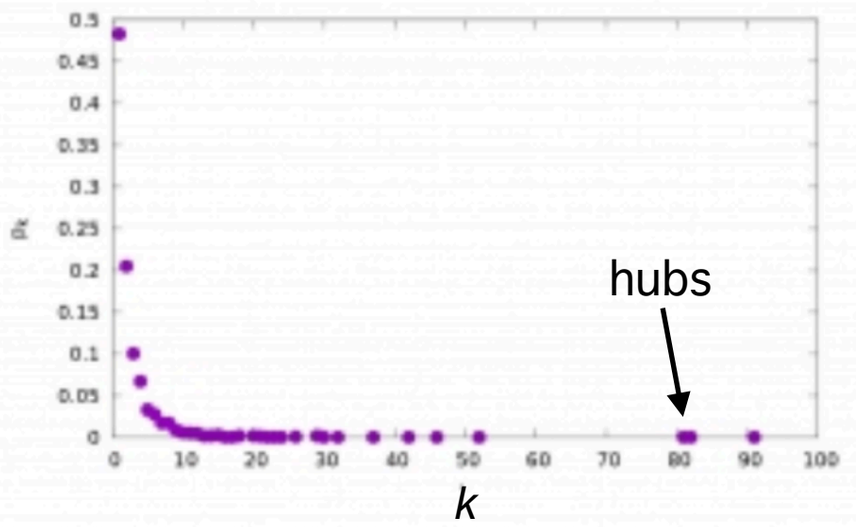


Protein interaction networks are scale-free

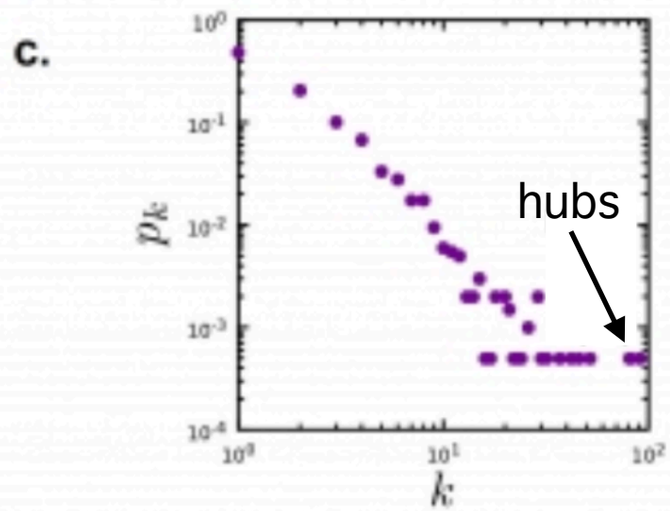
Yeast protein interaction network



Degree distribution - normal scale

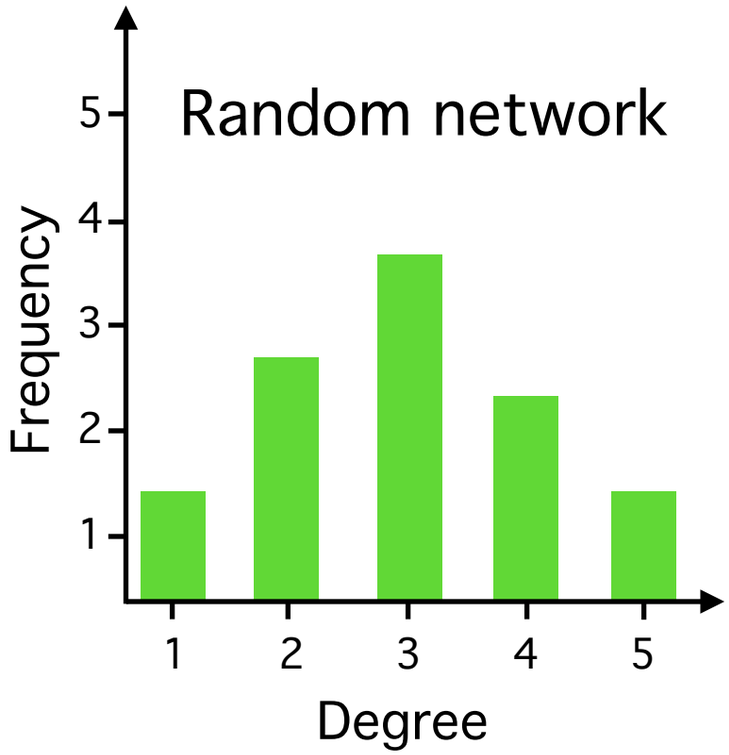


Degree distribution - log-log scale

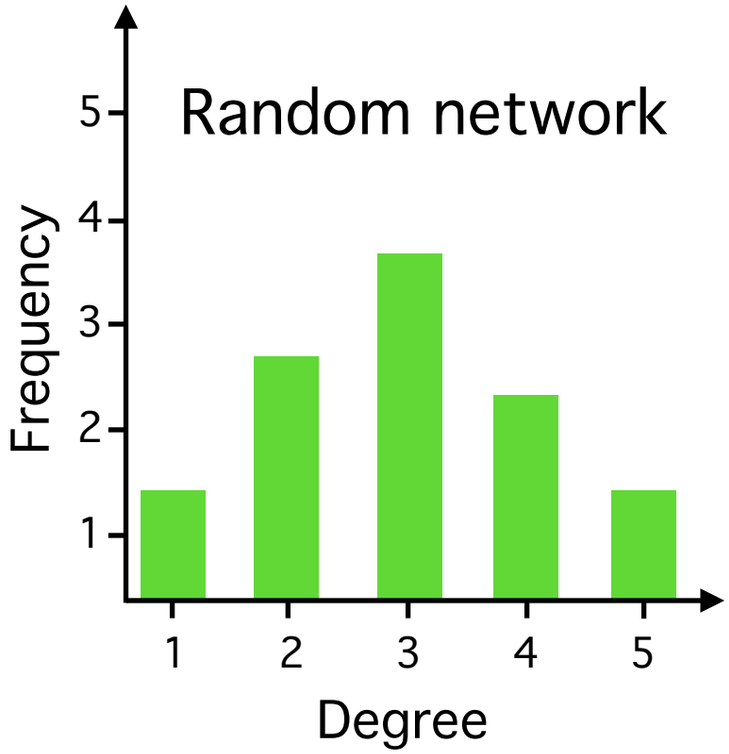


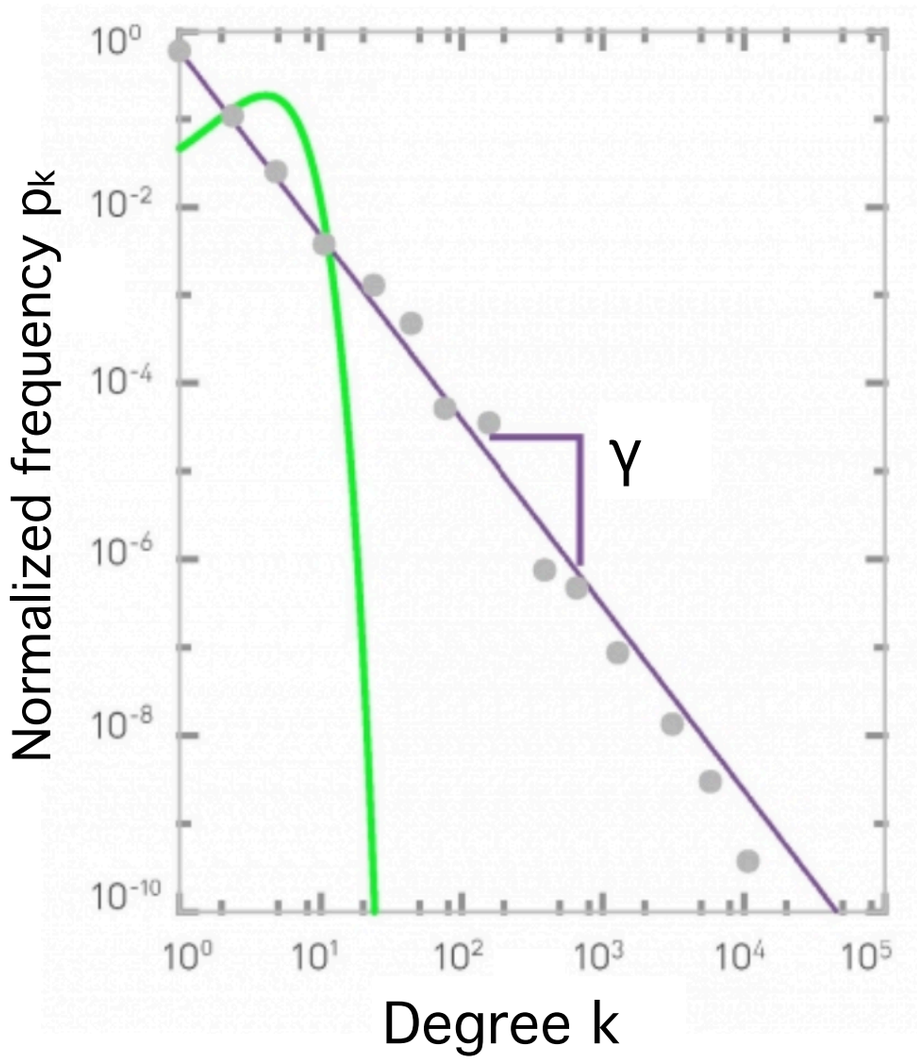
Degree distribution and scale-free networks

Degree distribution and scale-free networks



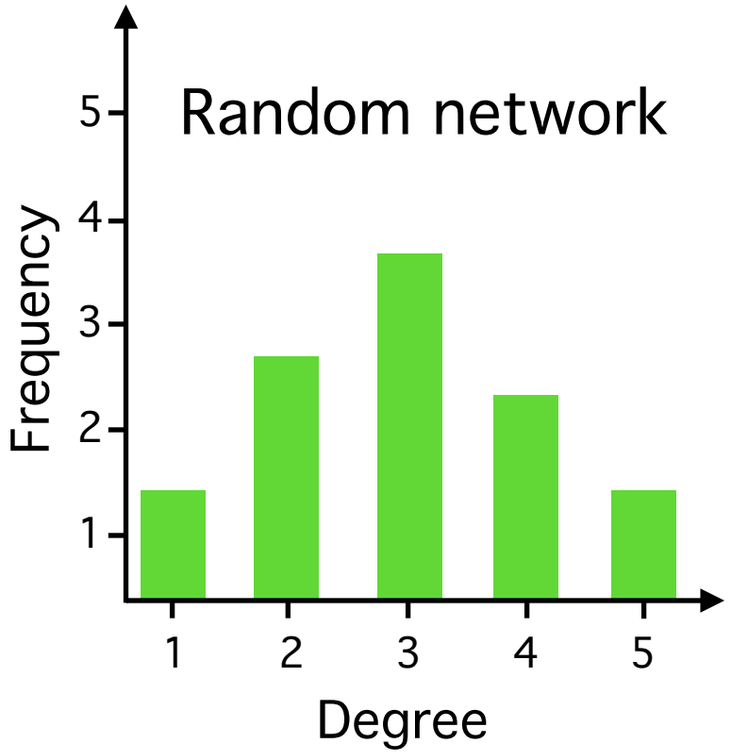
Degree distribution and scale-free networks

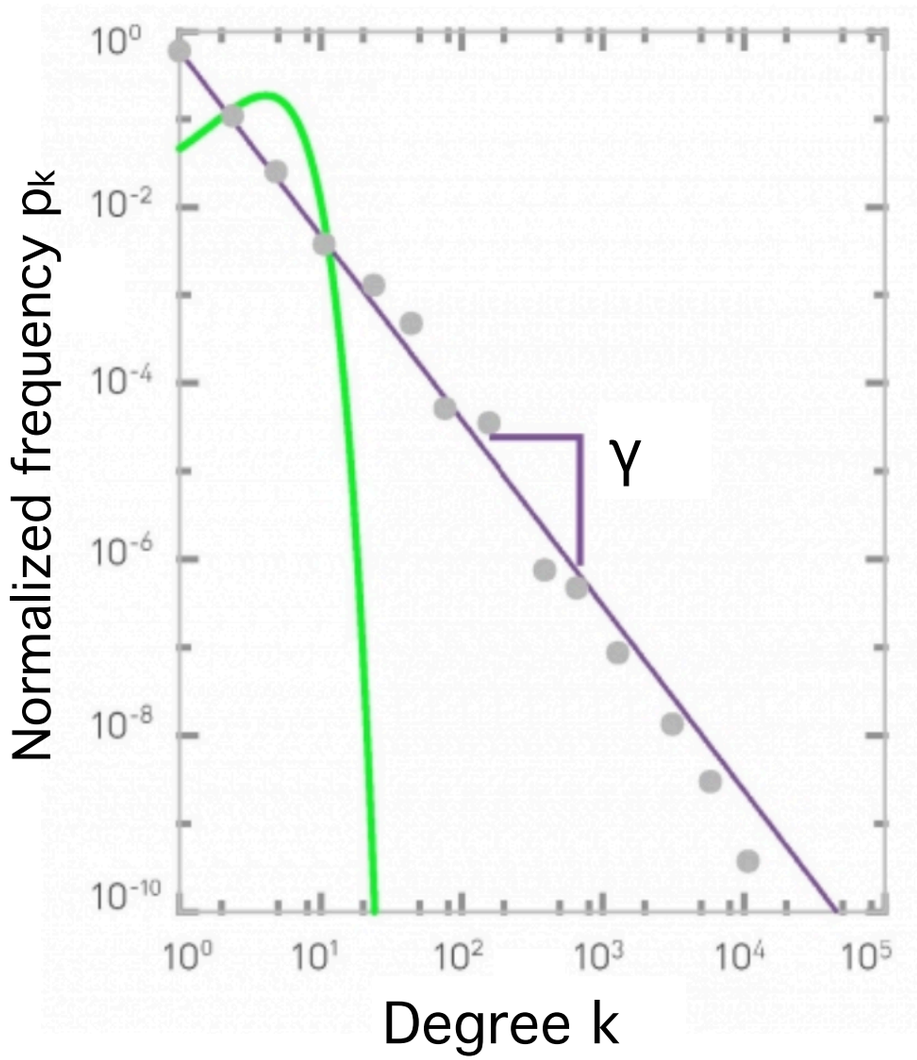




Degree distribution and scale-free networks

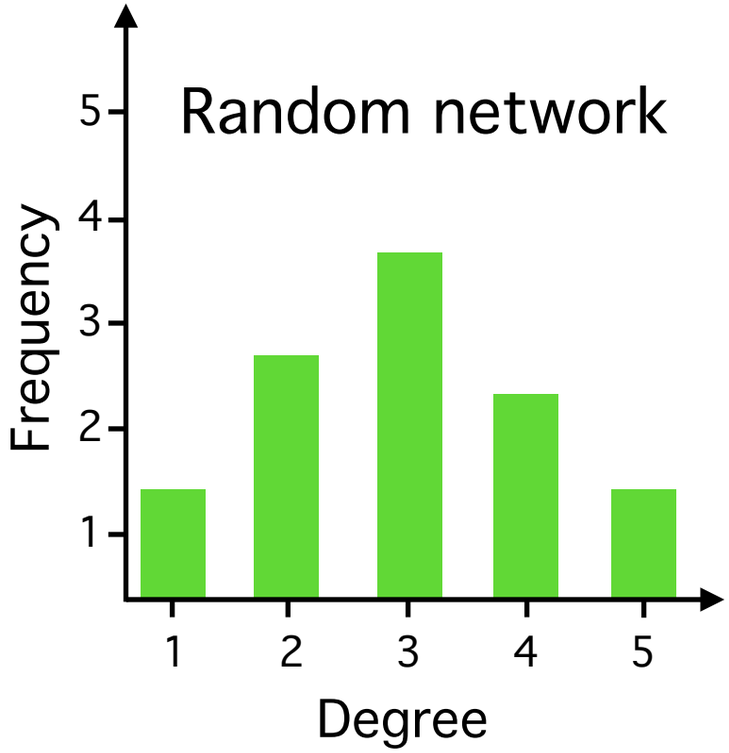
Degree distributions of many real world networks follow a power law distribution in log-log scale

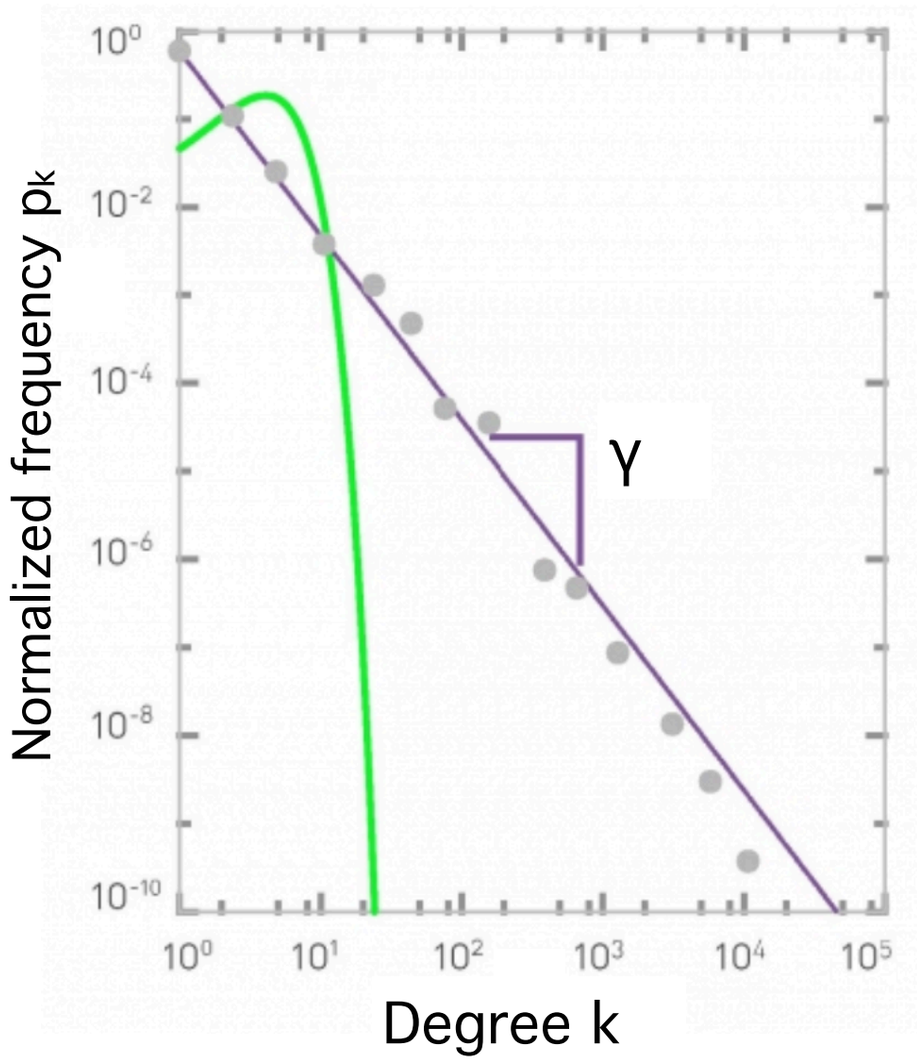




Degree distribution and scale-free networks

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





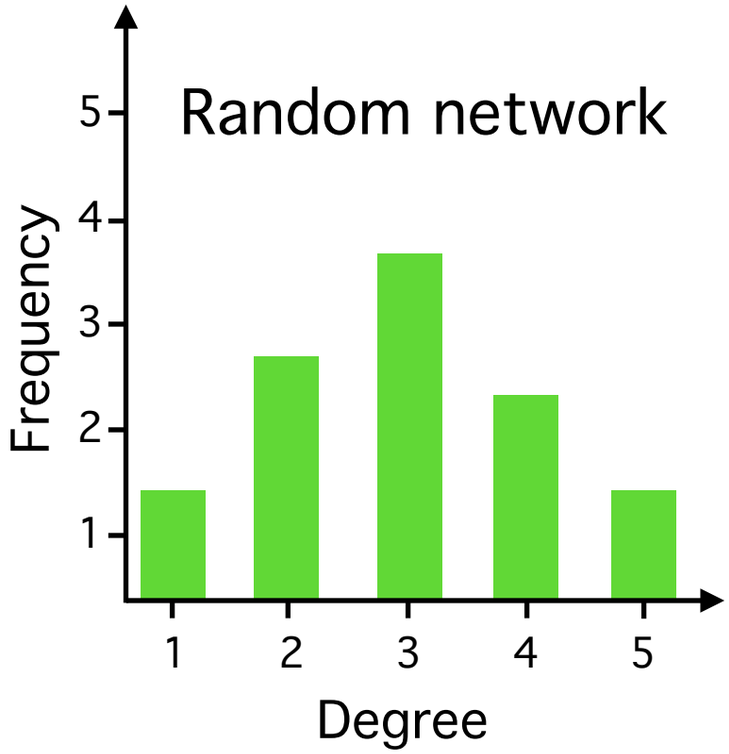
Power law distribution

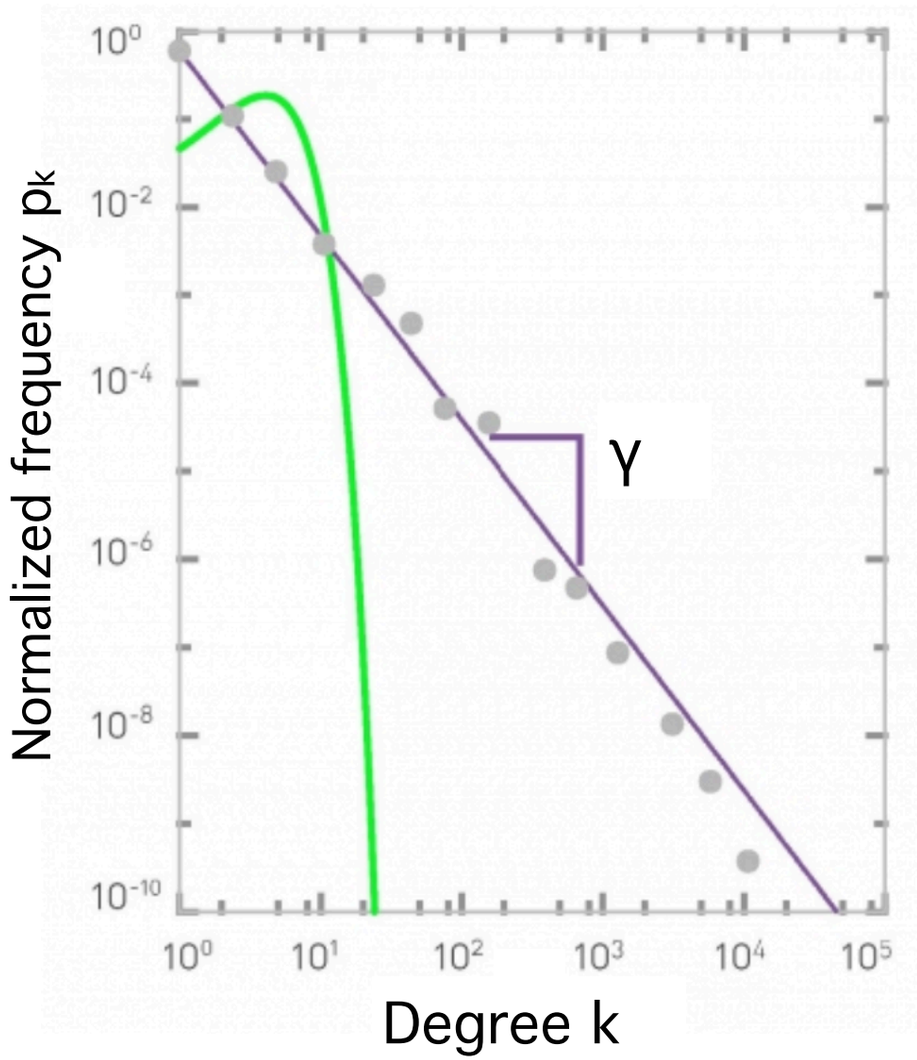
$p_k \sim k^{-\gamma}$

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

Degree distribution and scale-free 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.

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

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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into its large-scale characteristics. Our first goal was therefore to identify the architecture of this network, determining whether it is best described by an inherently uniform exponential topology, with proteins on average possessing the same number of links, or by a highly heterogeneous

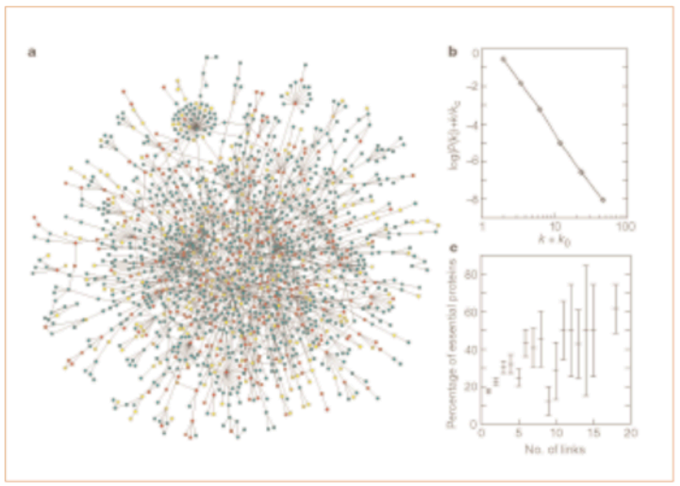


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-lethal; orange, slow growth; yellow, unknown). **b**, Connectivity distribution $P(k)$ of interacting yeast proteins, giving the probability that a given protein interacts with k other proteins. The exponential cut-off[6] indicates that the number of proteins with more than 20 interactions is slightly less than expected for pure scale-free networks. In the absence of data on the link directions, all interactions have been considered as bidirectional. The parameter controlling the short-length scale correction has value $k_0=1$. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database[13]. Detailed statistical analysis, including $r=0.75$ for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

Jeong et al *Nature* 2001

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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into
Our first g
architecture
whether it i
uniform exponential topology, with proteins on average possessing the same number of links, or by a highly heterogeneous

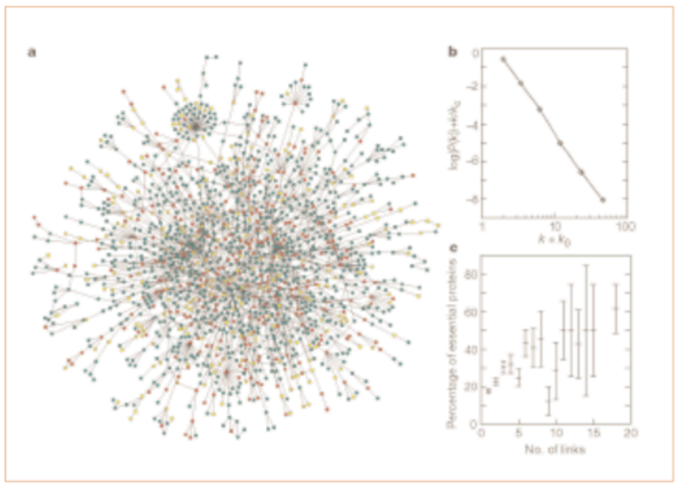


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-

Cited 3,266 times!

distribution $P(k)$ of interacting yeast proteins, giving the probability that a
-off[6] indicates that the number of proteins with more than 20 interactions
absence of data on the link directions, all interactions have been consid-
h scale correction has value $k_0=1$. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database[13]. Detailed statistical analysis, including $r=0.75$ for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

Jeong et al *Nature* 2001

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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into

Our first g
architecture
whether it i
uniform exponential topology, with proteins on average possessing the same number of links, or by a highly heterogeneous

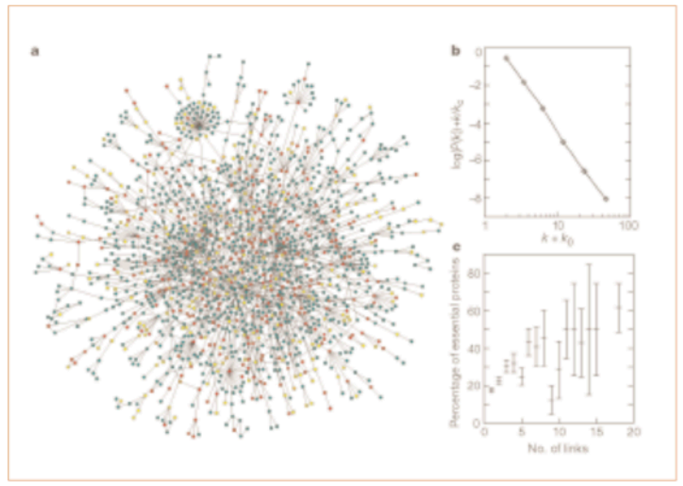


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-

istribution $P(k)$ of interacting yeast proteins, giving the probability that a

-off[6] indicates that the number of proteins with more than 20 interactions

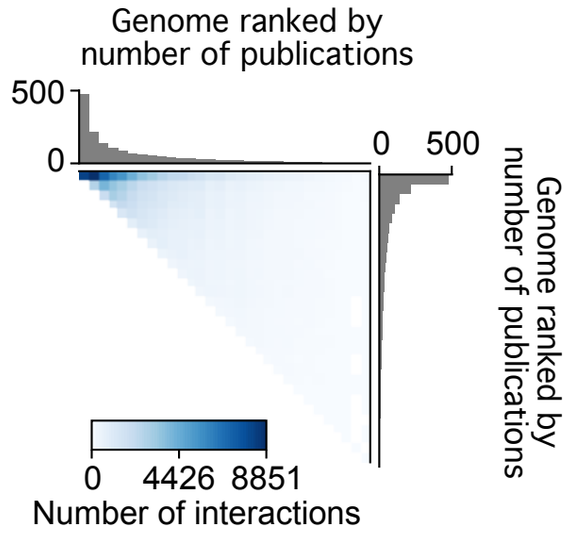
absence of data on the link directions, all interactions have been consid-

h scale correction has value k_0=1. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database[13]. Detailed statistical analysis, including r = 0.75 for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

Cited 3,266 times!

Jeong et al *Nature* 2001

Are essential genes more highly studied?



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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[5]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into

Our first g
architecture
whether it i
uniform exponential topology, with pro-

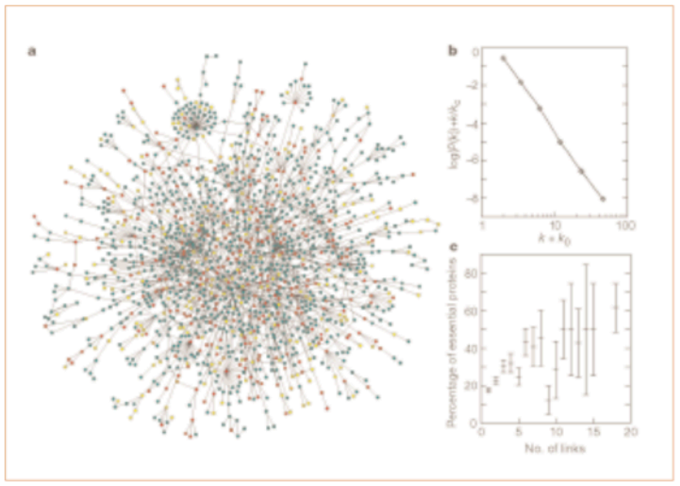


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-

listribution $P(k)$ of interacting yeast proteins, giving the probability that a

-off[6] indicates that the number of proteins with more than 20 interactions

absence of data on the link directions, all interactions have been consid-

h scale correction has value k_0=1. **c**, The fraction of essential proteins

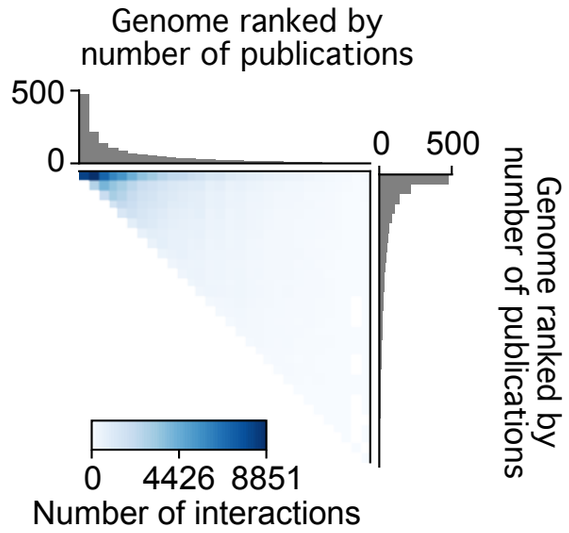
with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was

obtained from the Proteome database[13]. Detailed statistical analysis, including r=0.75 for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

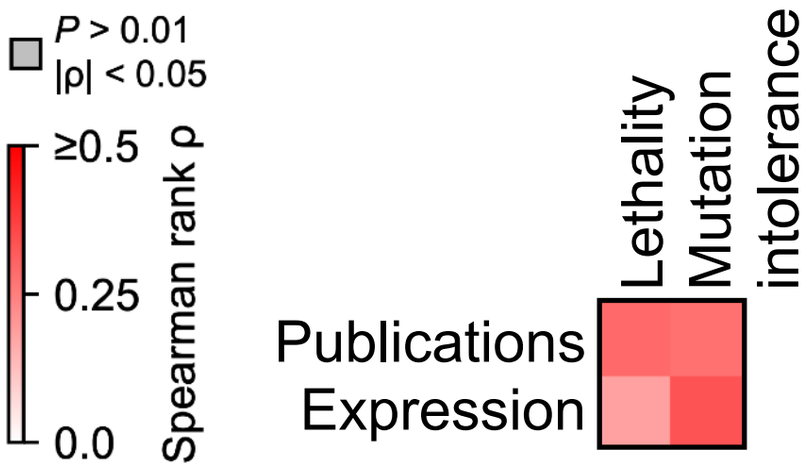
Cited 3,266 times!

Jeong et al *Nature* 2001

Are essential genes more highly studied?



Degree distributions are influenced by technical assay biases



Luck et al *Nature* 2020

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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into

Our first g
architecture
whether it i
uniform exponential topology, with pro-

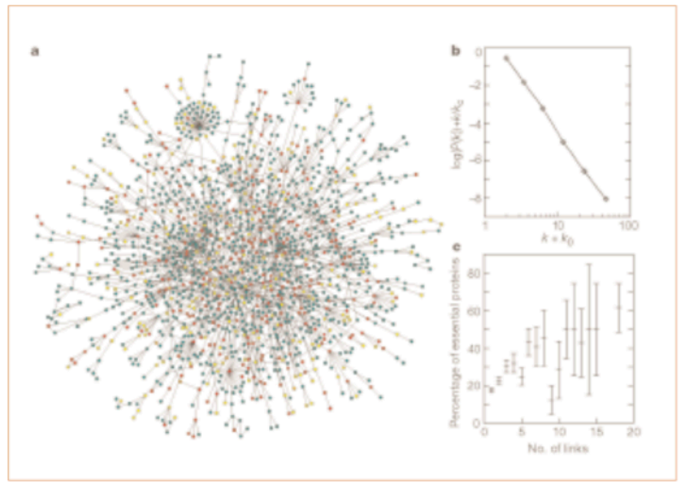


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-

istribution $P(k)$ of interacting yeast proteins, giving the probability that a

-off[6] indicates that the number of proteins with more than 20 interactions

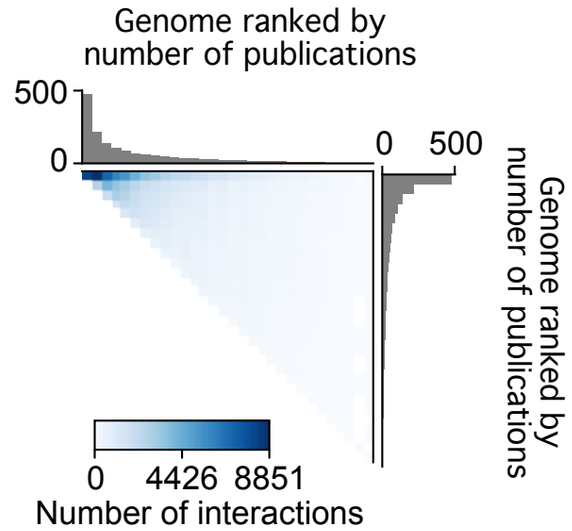
absence of data on the link directions, all interactions have been consid-

h scale correction has value k_0=1. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database[13]. Detailed statistical analysis, including r=0.75 for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

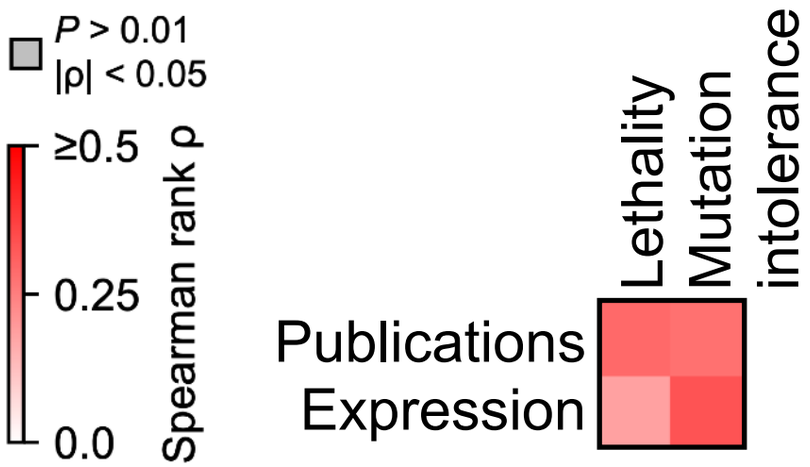
Cited 3,266 times!

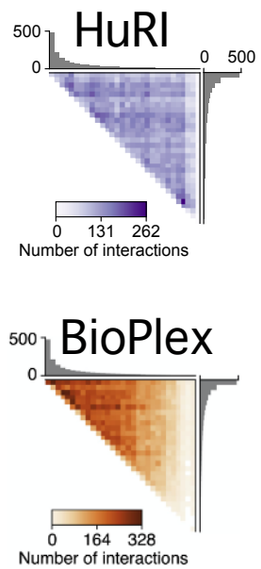
Jeong et al *Nature* 2001

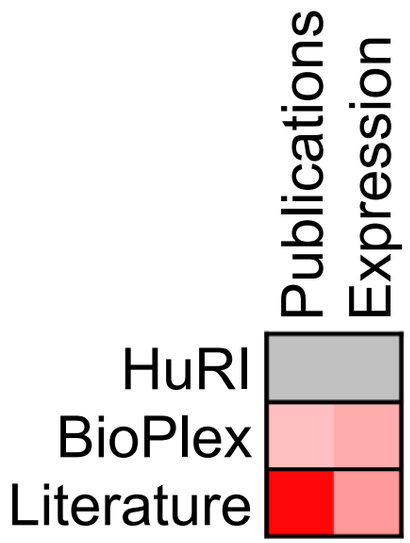
Are essential genes more highly studied?



Degree distributions are influenced by technical assay biases







Luck et al *Nature* 2020

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

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein interaction network we investigate has 1,870 proteins as nodes, connected by 2,240 identified direct physical interactions, and is derived from combined, non-overlapping data[3,4], obtained mostly by systematic two-hybrid analyses[3]. Owing to its size, a complete map of the network (Fig. 1a), although informative, in itself offers little insight into

Our first g
architecture
whether it i
uniform exponential topology, with pro-

teins on average possessing the same number of links, or by a highly heterogeneous

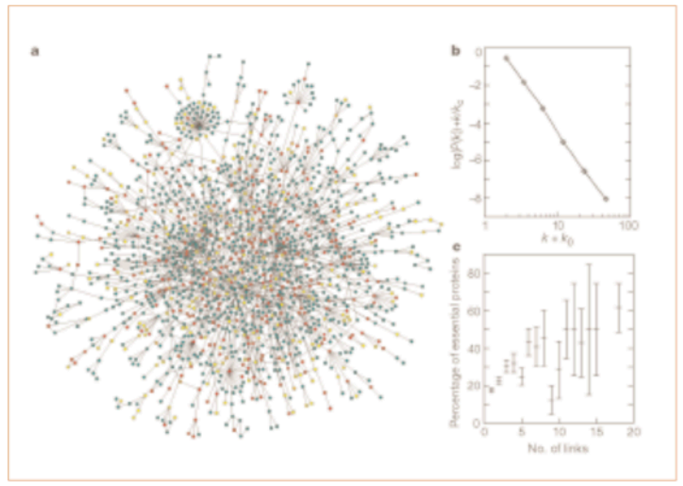


Figure 1 Characteristics of the yeast proteome. **a**, Map of protein–protein interactions. The largest cluster, which contains ~78% of all proteins, is shown. The colour of a node signifies the phenotypic effect of removing the corresponding protein (red, lethal; green, non-

istribution $P(k)$ of interacting yeast proteins, giving the probability that a

-off[6] indicates that the number of proteins with more than 20 interactions

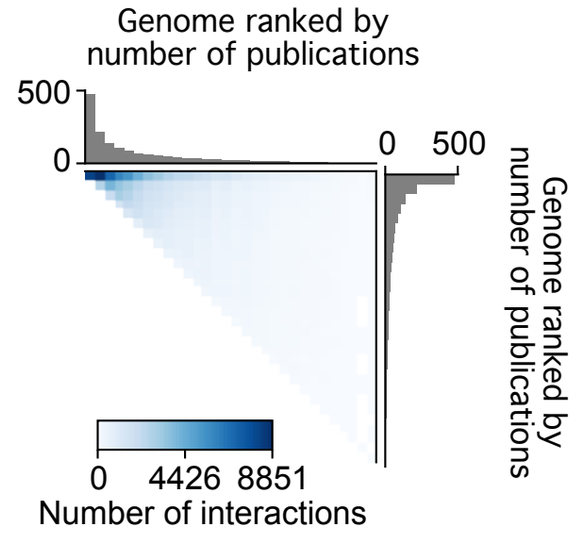
absence of data on the link directions, all interactions have been consid-

h scale correction has value $k_0 \approx 1$. **c**, The fraction of essential proteins with exactly k links versus their connectivity, k, in the yeast proteome. The list of 1,572 mutants with known phenotypic profile was obtained from the Proteome database[13]. Detailed statistical analysis, including $r = 0.75$ for Pearson's linear correlation coefficient, demonstrates a positive correlation between lethality and connectivity. For additional details, see http://www.nd.edu/~networks/cell.

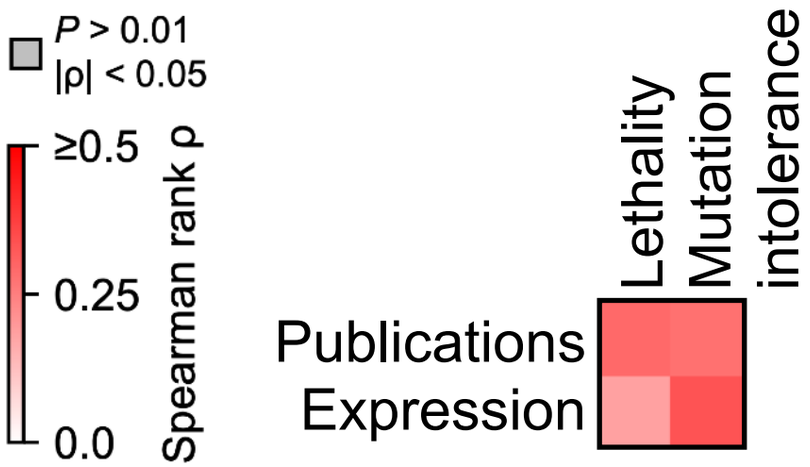
Cited 3,266 times!

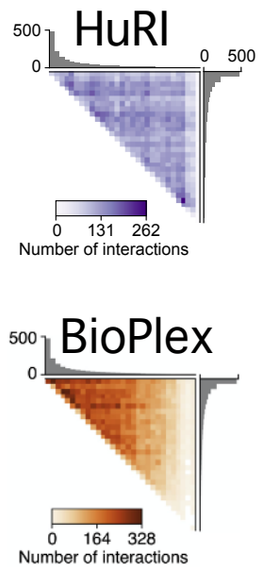
Jeong et al *Nature* 2001

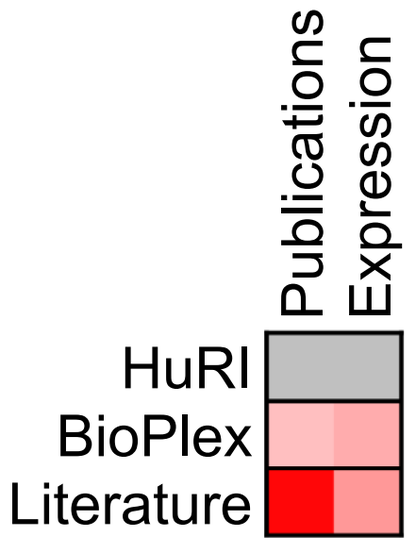
Are essential genes more highly studied?

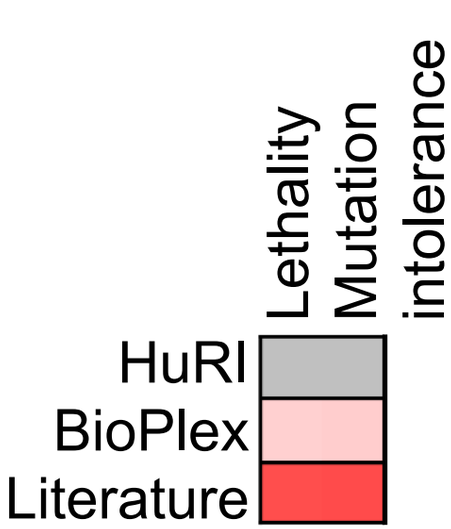


Degree distributions are influenced by technical assay biases



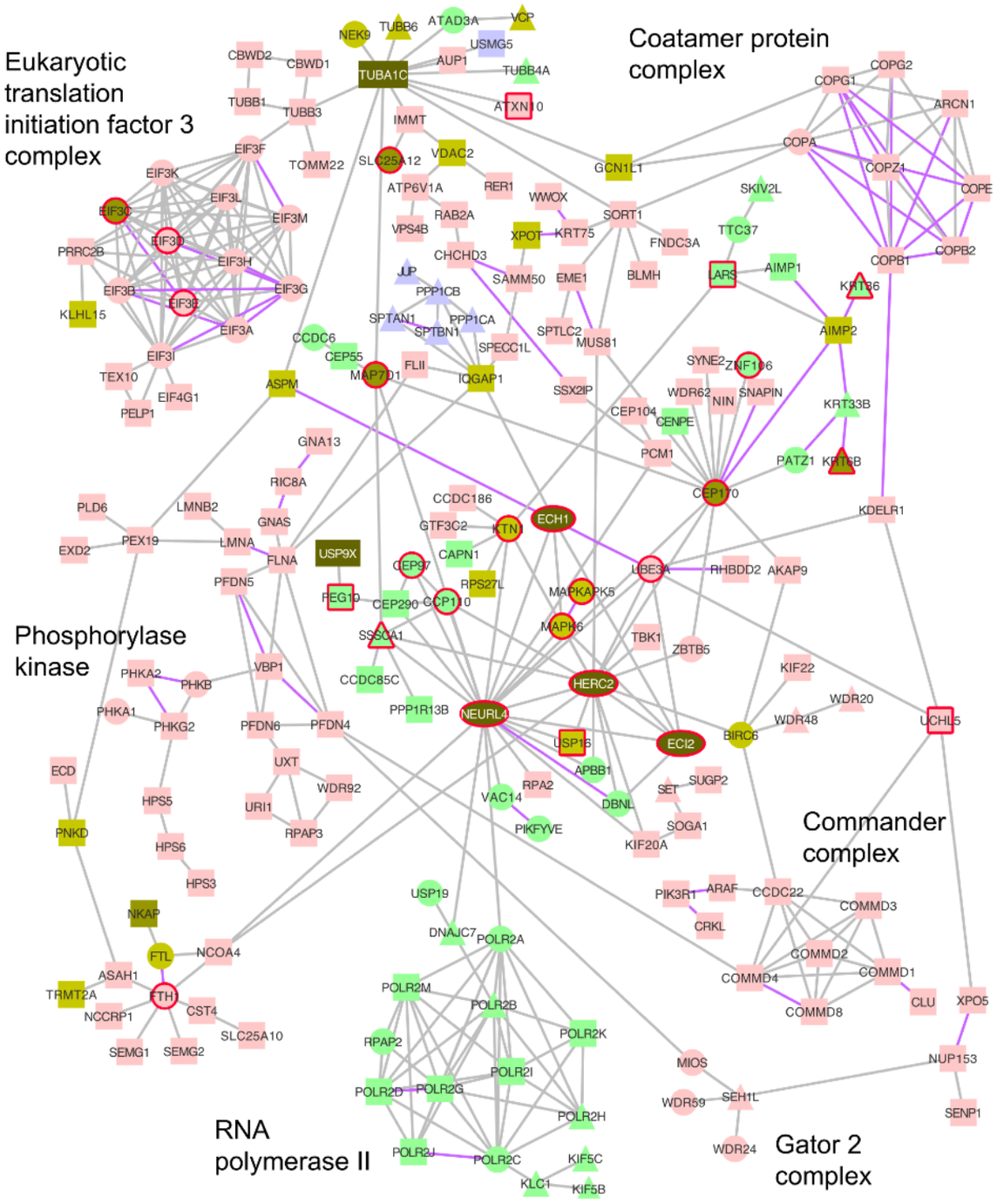






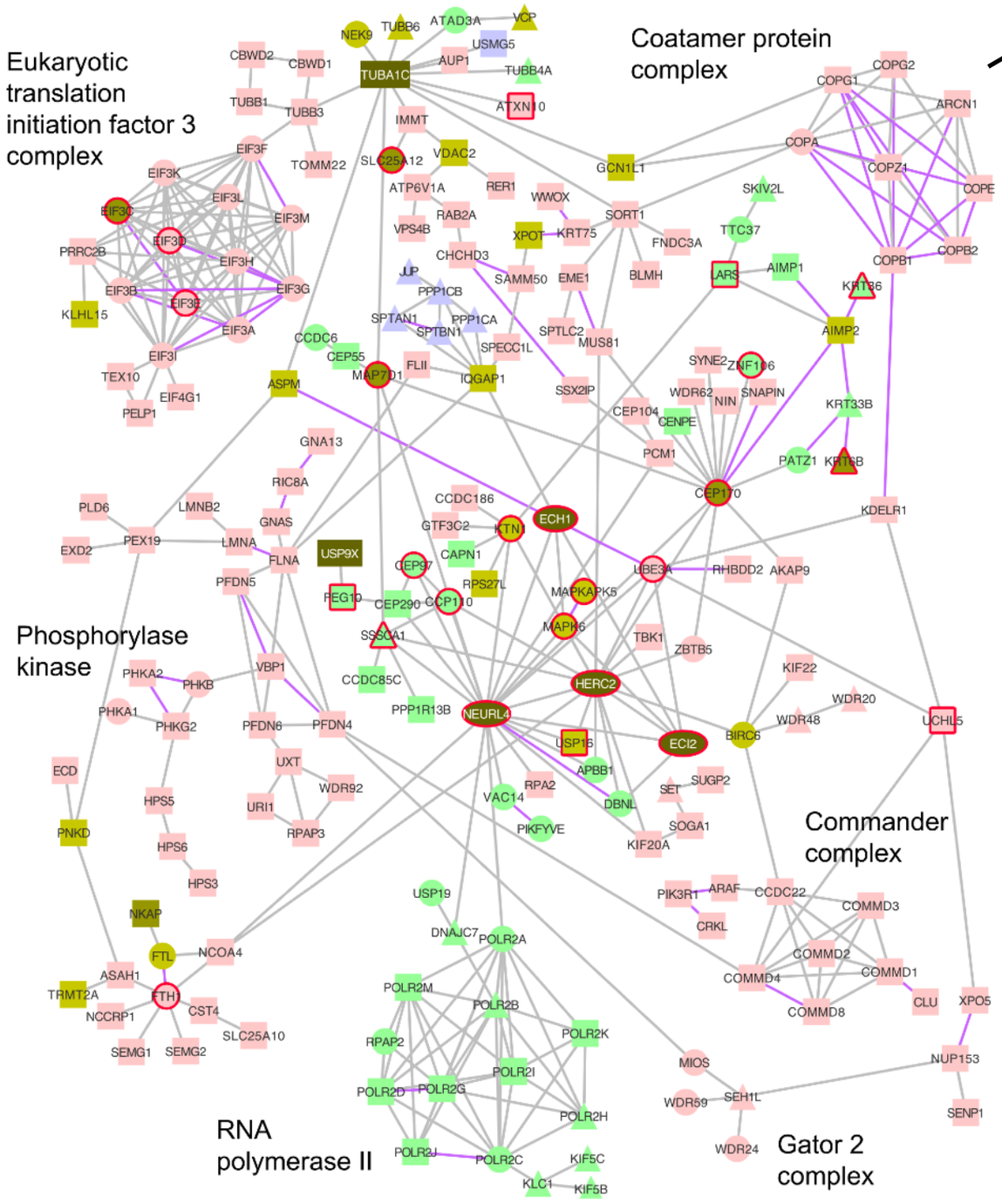
Luck et al *Nature* 2020

Finding communities in graphs



Martinez-Noel et al *JMB* 2018, networksciencebook.com

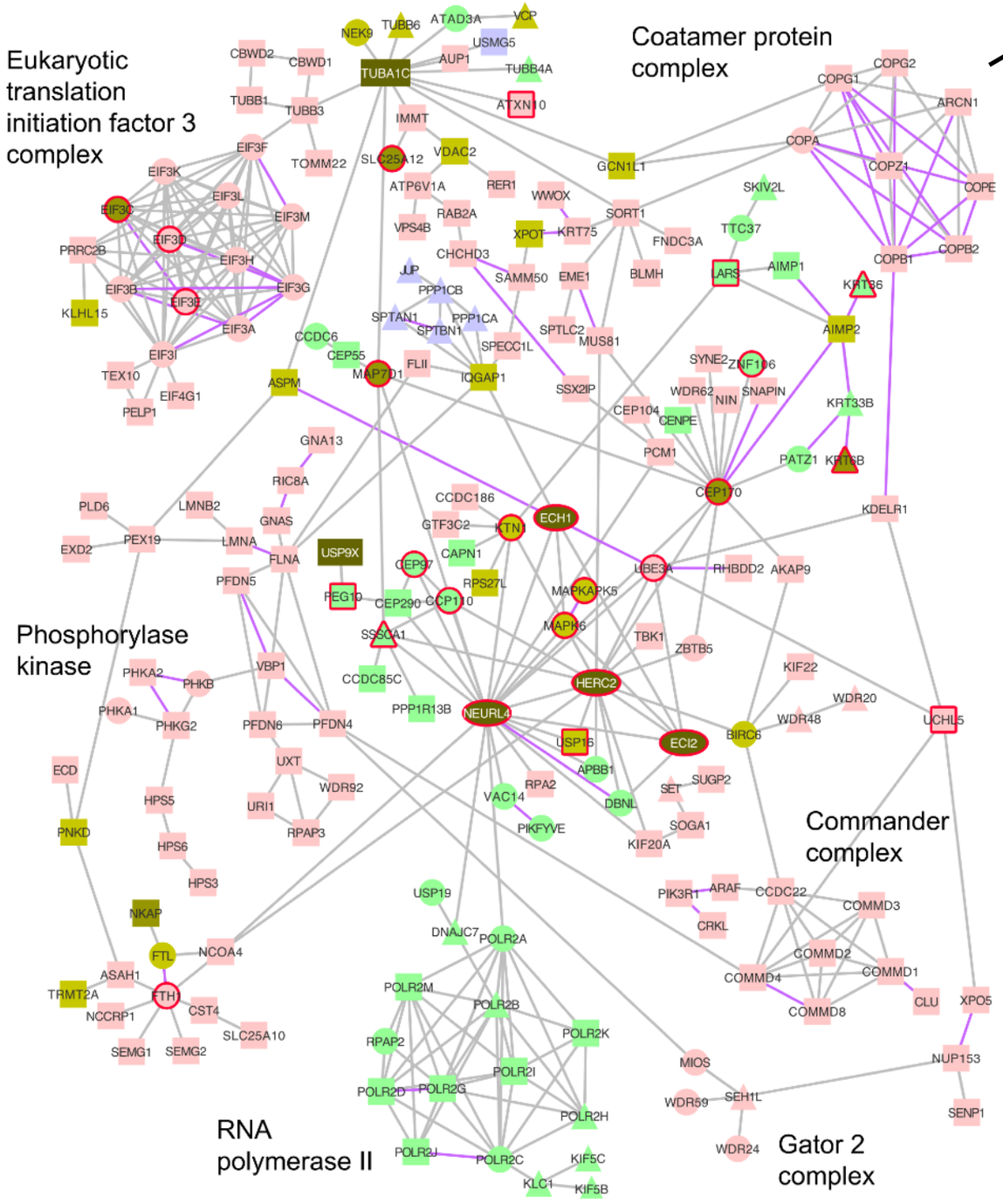
Finding communities in graphs



Protein complexes show as clusters in a network

Martinez-Noel et al *JMB* 2018, networksciencebook.com

Finding communities in graphs



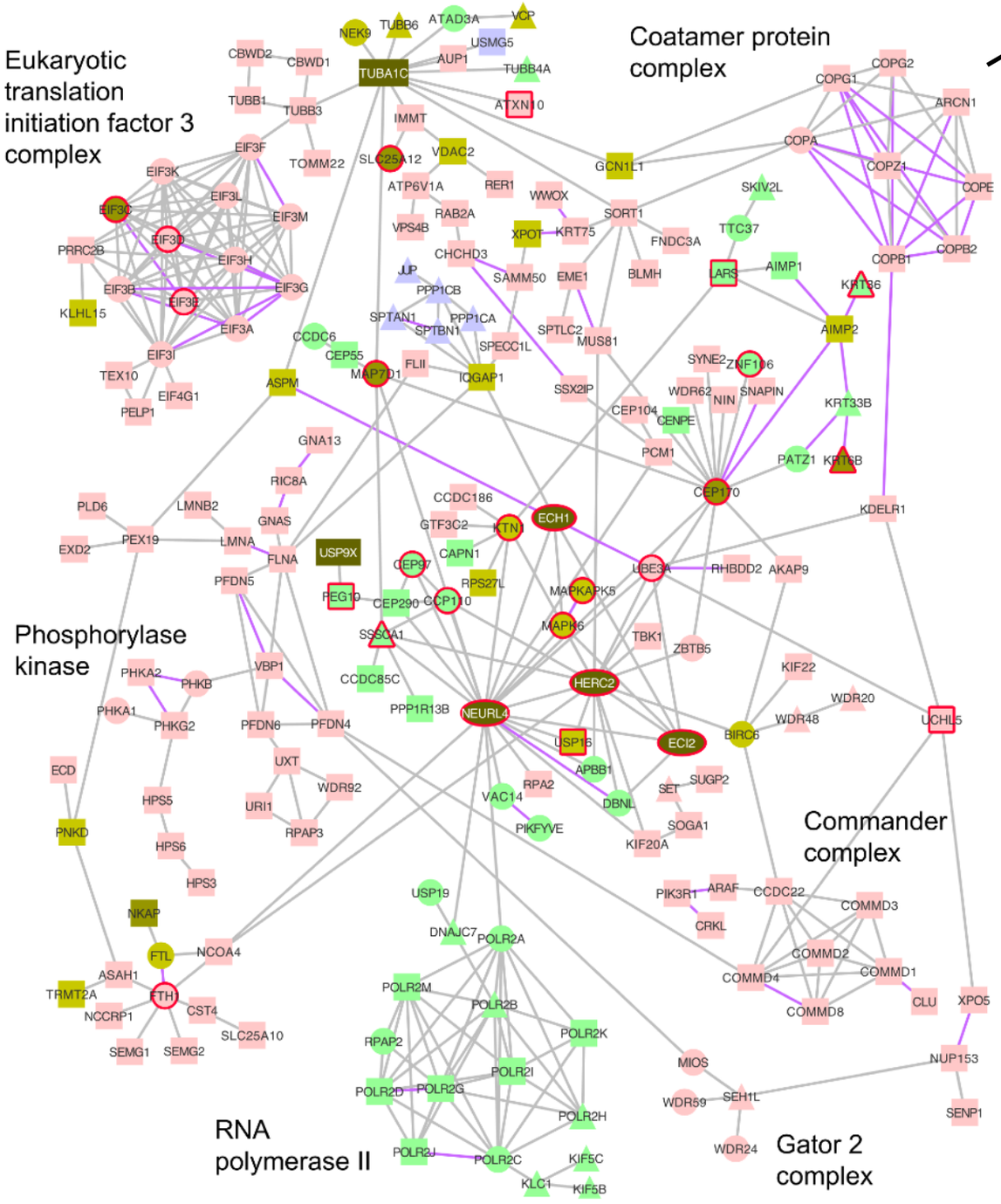
Protein complexes show as clusters in a network

↓

Communities are locally dense connected subgraphs in a network

Martinez-Noel et al *JMB* 2018, networksciencebook.com

Finding communities in graphs



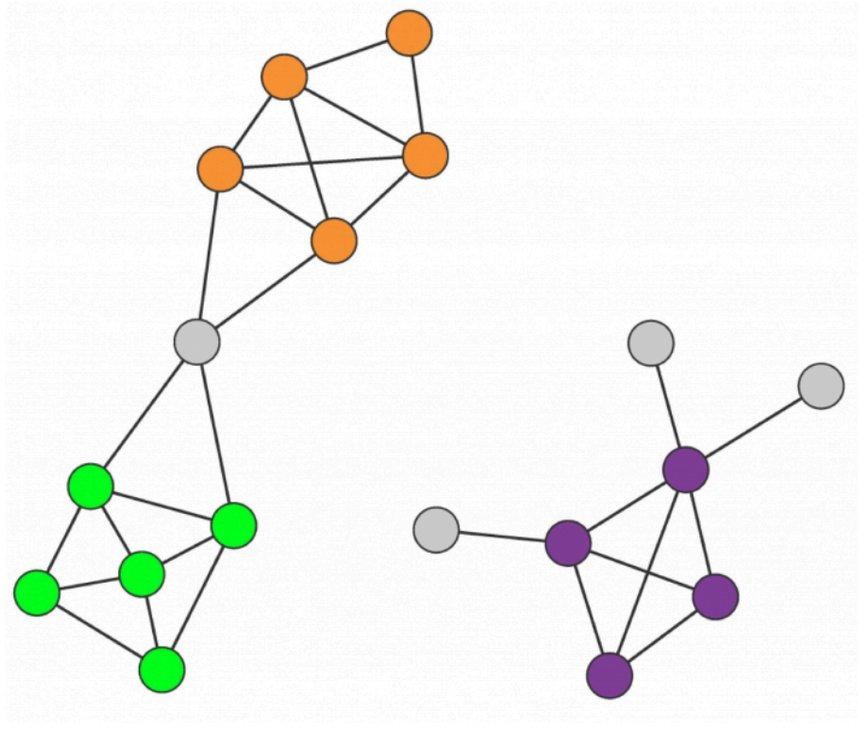
Protein complexes show as clusters in a network

↓

Communities are locally dense connected subgraphs in a network

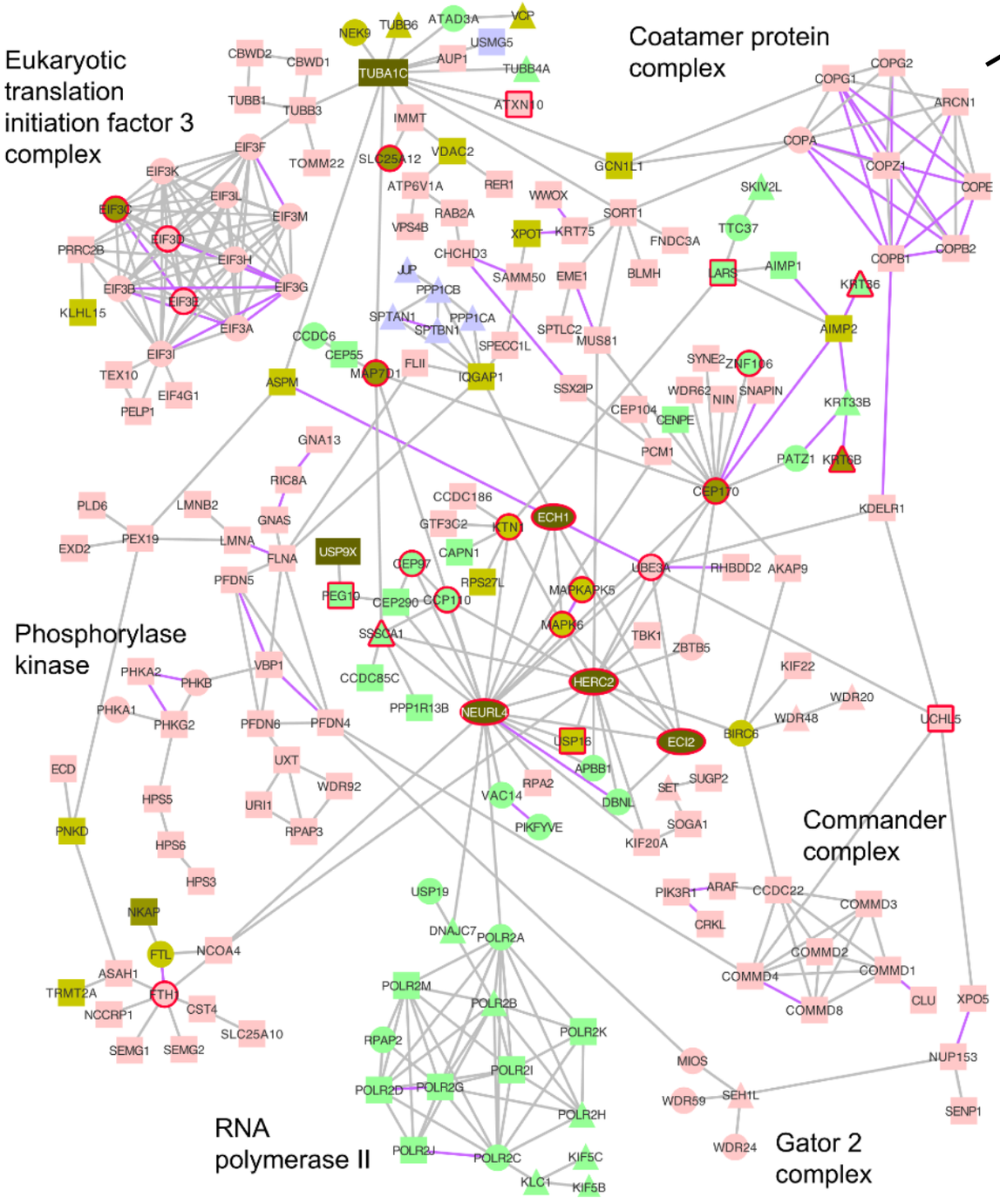
↓

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



Martinez-Noel et al *JMB* 2018, networksciencebook.com

Finding communities in graphs

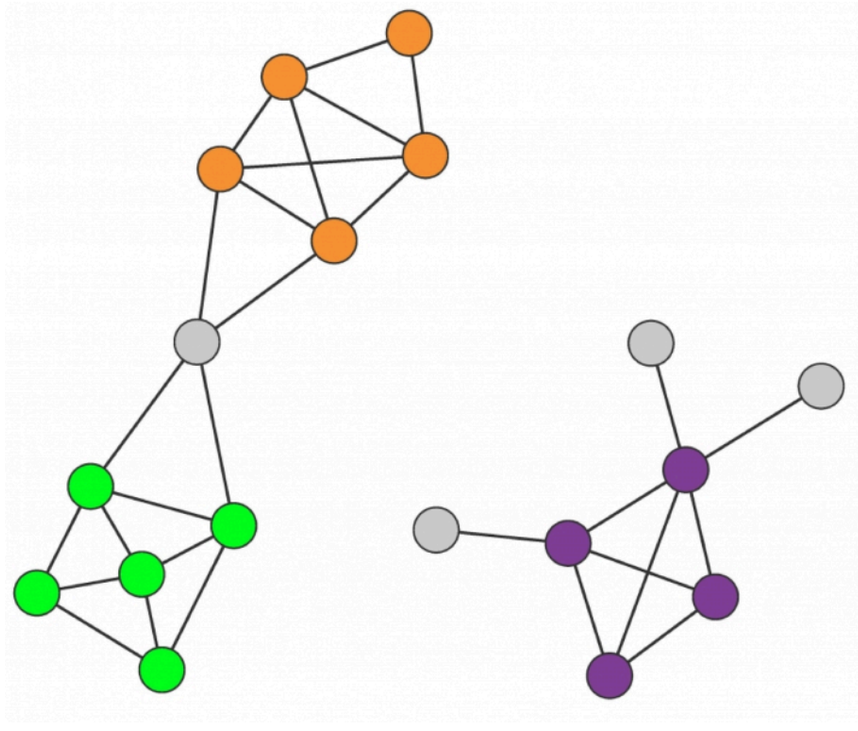


Numerous algorithms exist to find communities in a graph

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

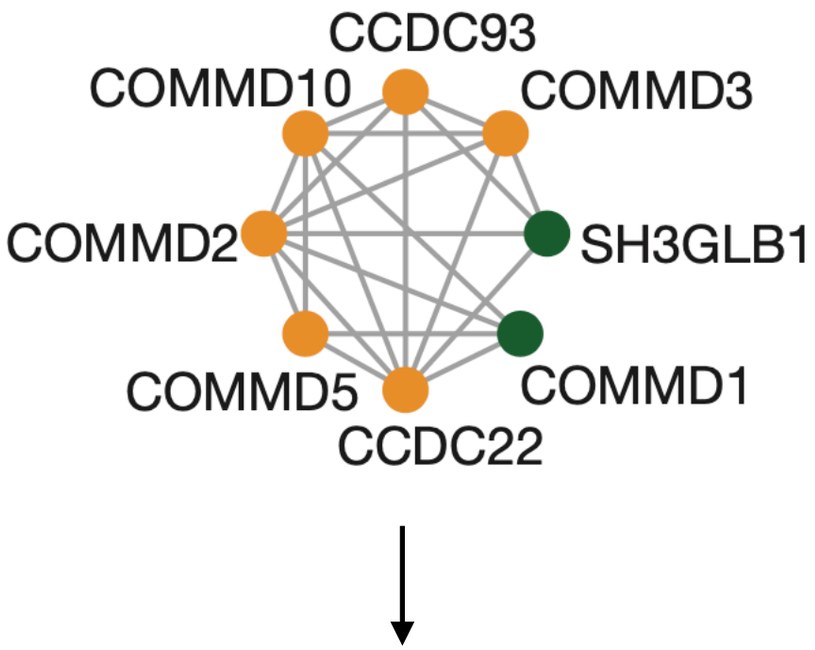


Martinez-Noel et al *JMB* 2018, networksciencebook.com

Can I find new protein complexes or complex members?

Can I find new protein complexes or complex members?

Identification of Commander complex

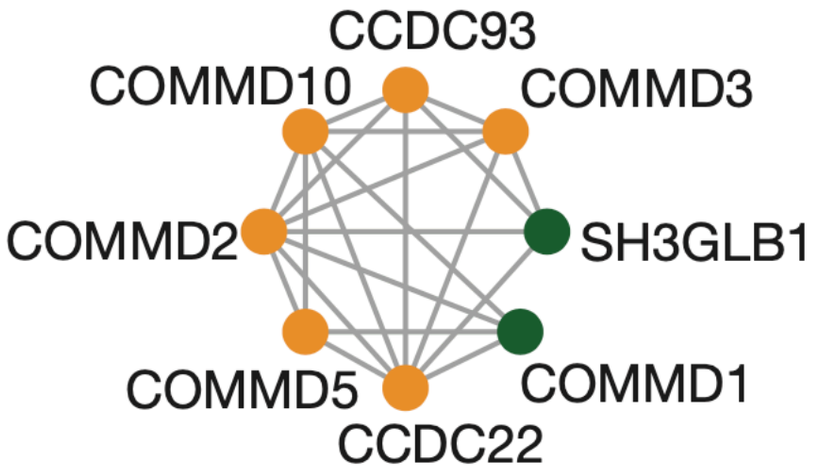


Role in embryonic development

Wan et al *Science* 2015

Can I find new protein complexes or complex members?

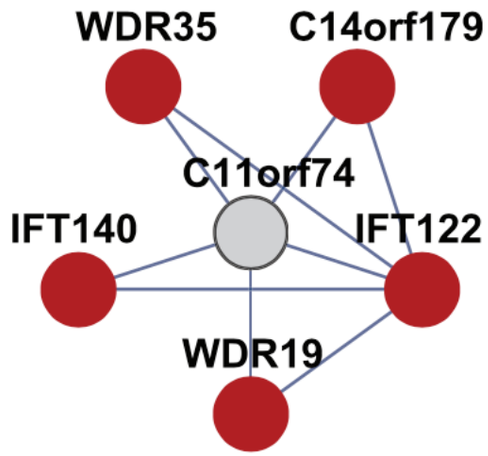
Identification of Commander complex



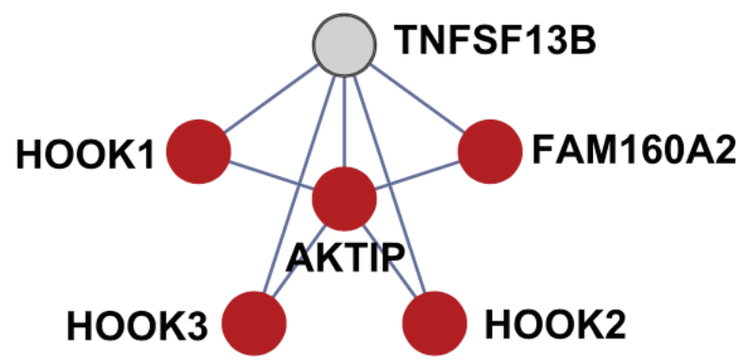
Role in embryonic development

Identification of new complex members

Intraciliary Transport Particle A



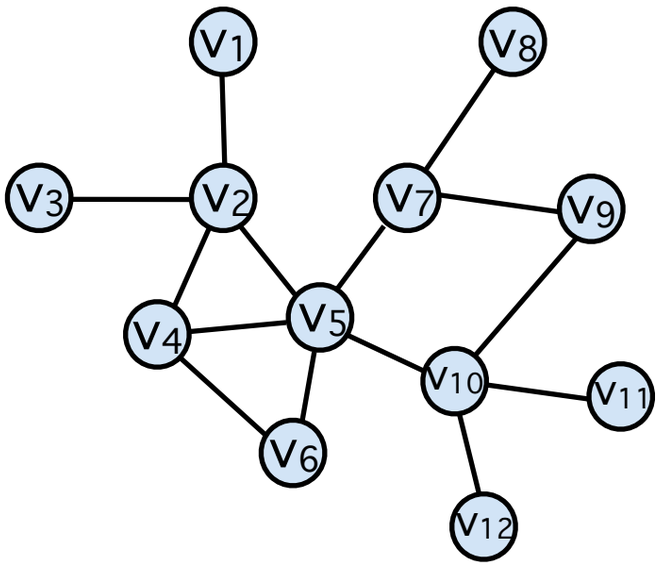
FHF complex



Wan et al *Science* 2015

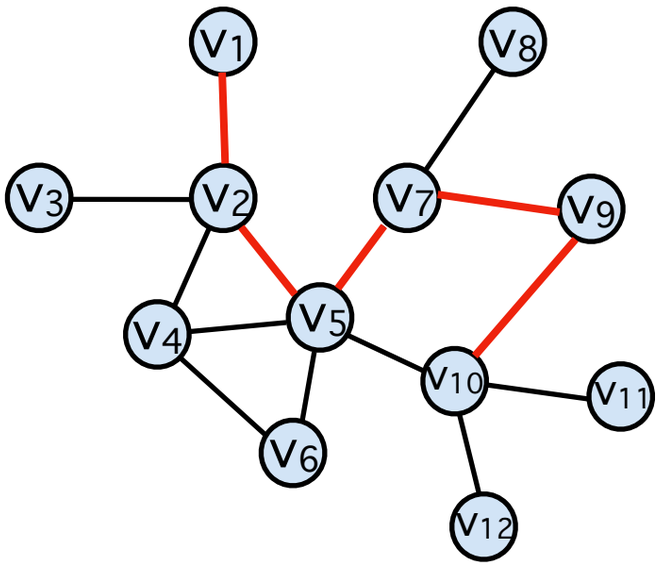
Huttlin et al *Cell* 2015

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.

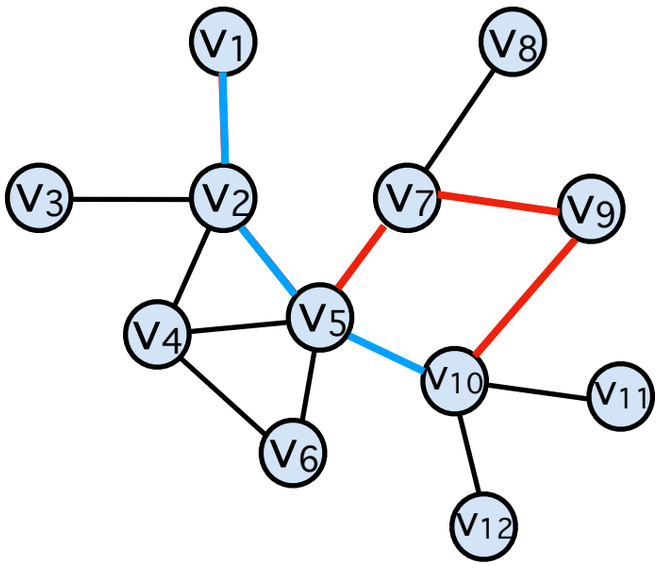
Shortest paths in graphs and betweenness centrality



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 paths in graphs and betweenness centrality

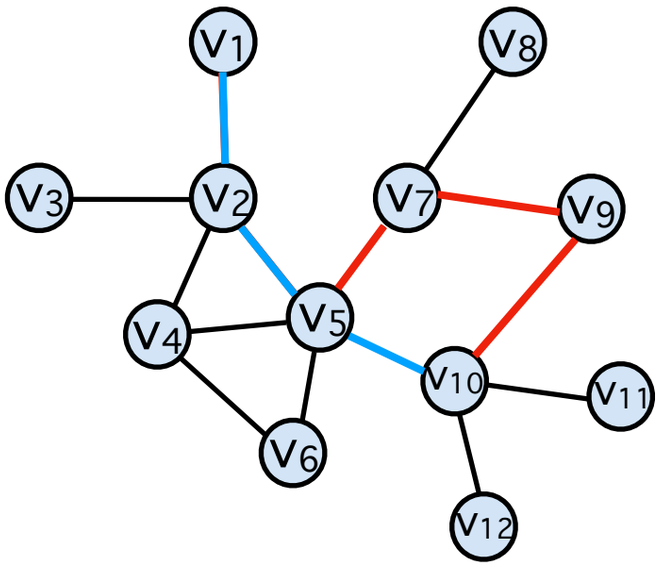


A path between two vertices is formed by the edges that lead from one vertex to the other.

A path from v_1 to v_{10}

Shortest path d from v_1 to v_{10}

Shortest paths in graphs and betweenness centrality



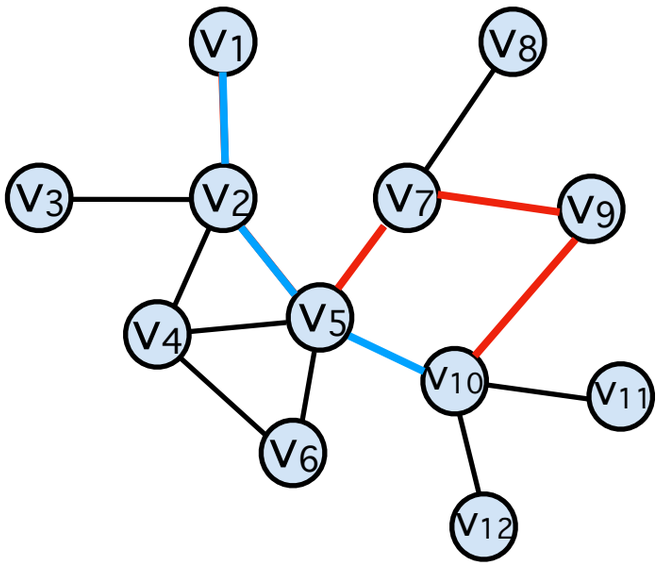
A path between two vertices is formed by
the edges that lead from one vertex to the other.

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

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.

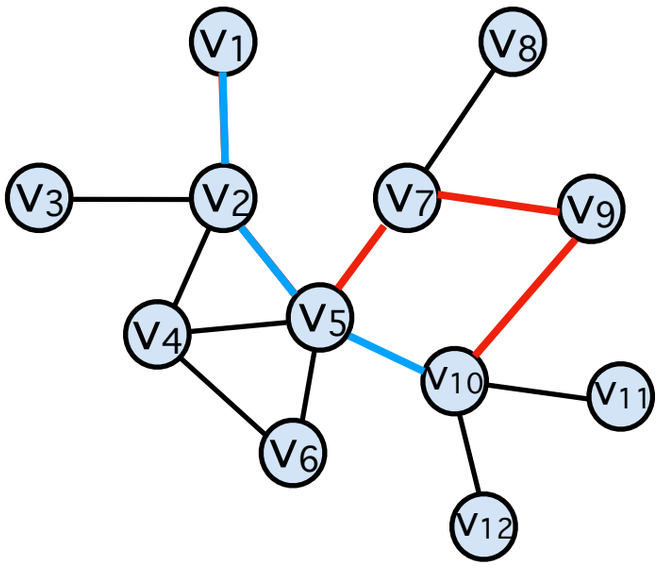
A path from v_1 to v_{10}

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?

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.

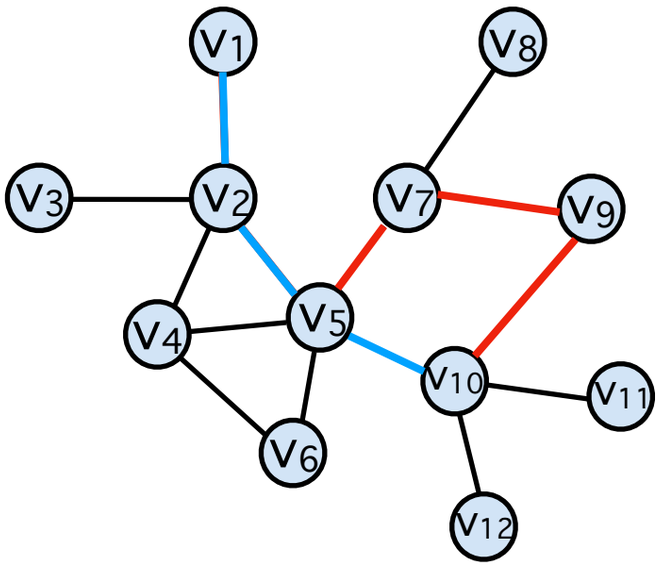
A path from v_1 to v_{10}

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? ⟶ Node betweenness

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}

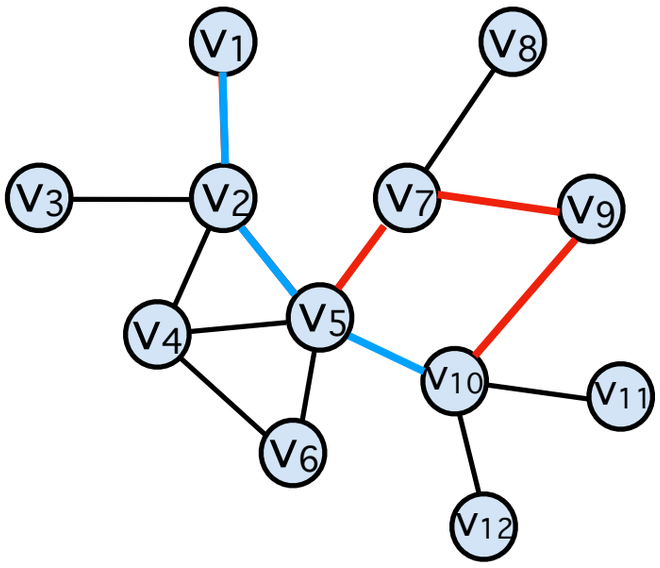
Shortest path d from v_1 to v_{10}

-> a path can represent information flow in a graph

How many shortest paths cross a vertex? ⟶ Node betweenness

How many shortest paths go over an edge?

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}

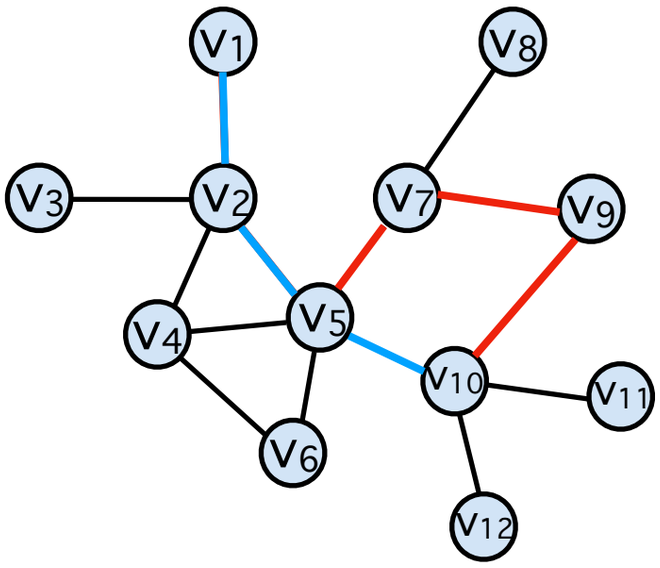
Shortest path d from v_1 to v_{10}

-> a path can represent information flow in a graph

How many shortest paths cross a vertex? ⟶ Node betweenness

How many shortest paths go over an edge? ⟶ Edge betweenness

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? ⟶ Node betweenness

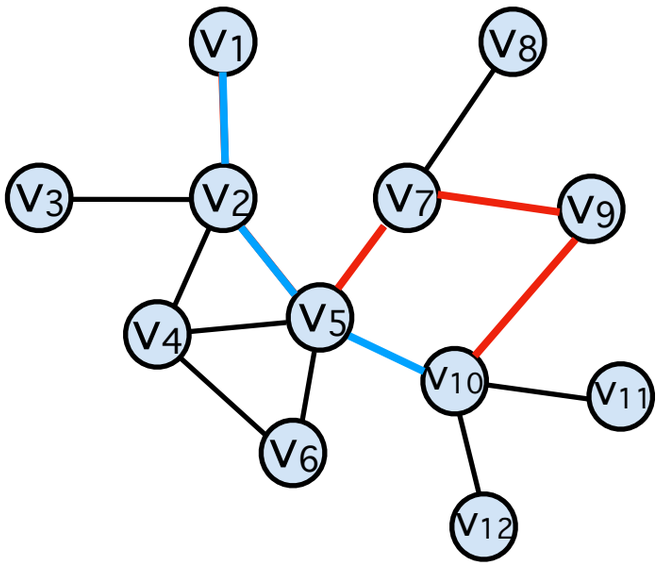
How many shortest paths go over an edge? ⟶ Edge betweenness

High betweenness

↓

Important for system

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? ⟶ Node betweenness

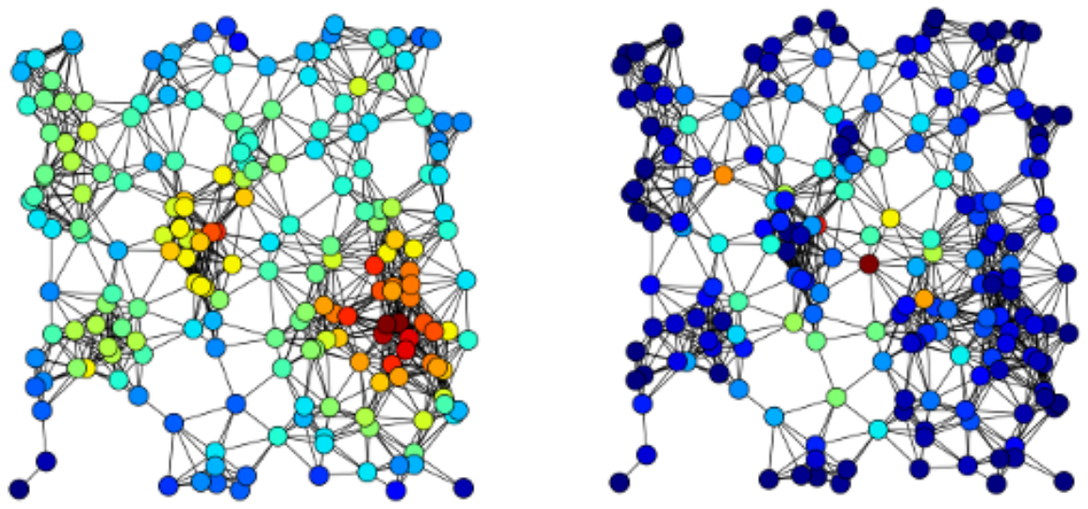
How many shortest paths go over an edge? ⟶ Edge betweenness

High betweenness

↓

Important for system

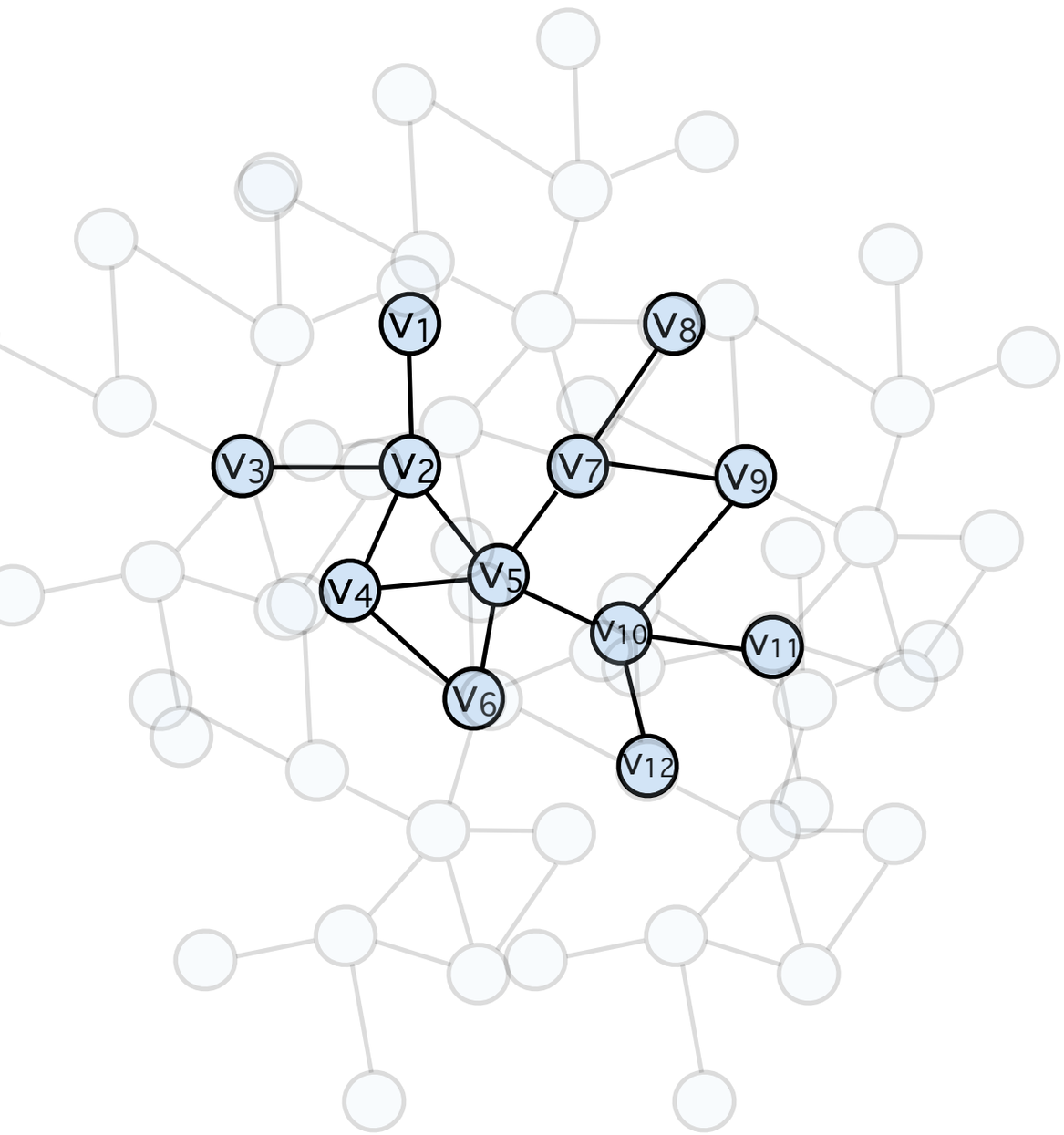
High degree ≠ high betweenness



Tapiocozzo, https://commons.wikimedia.org/w/index.php?curid=39064835

Measuring closeness in networks

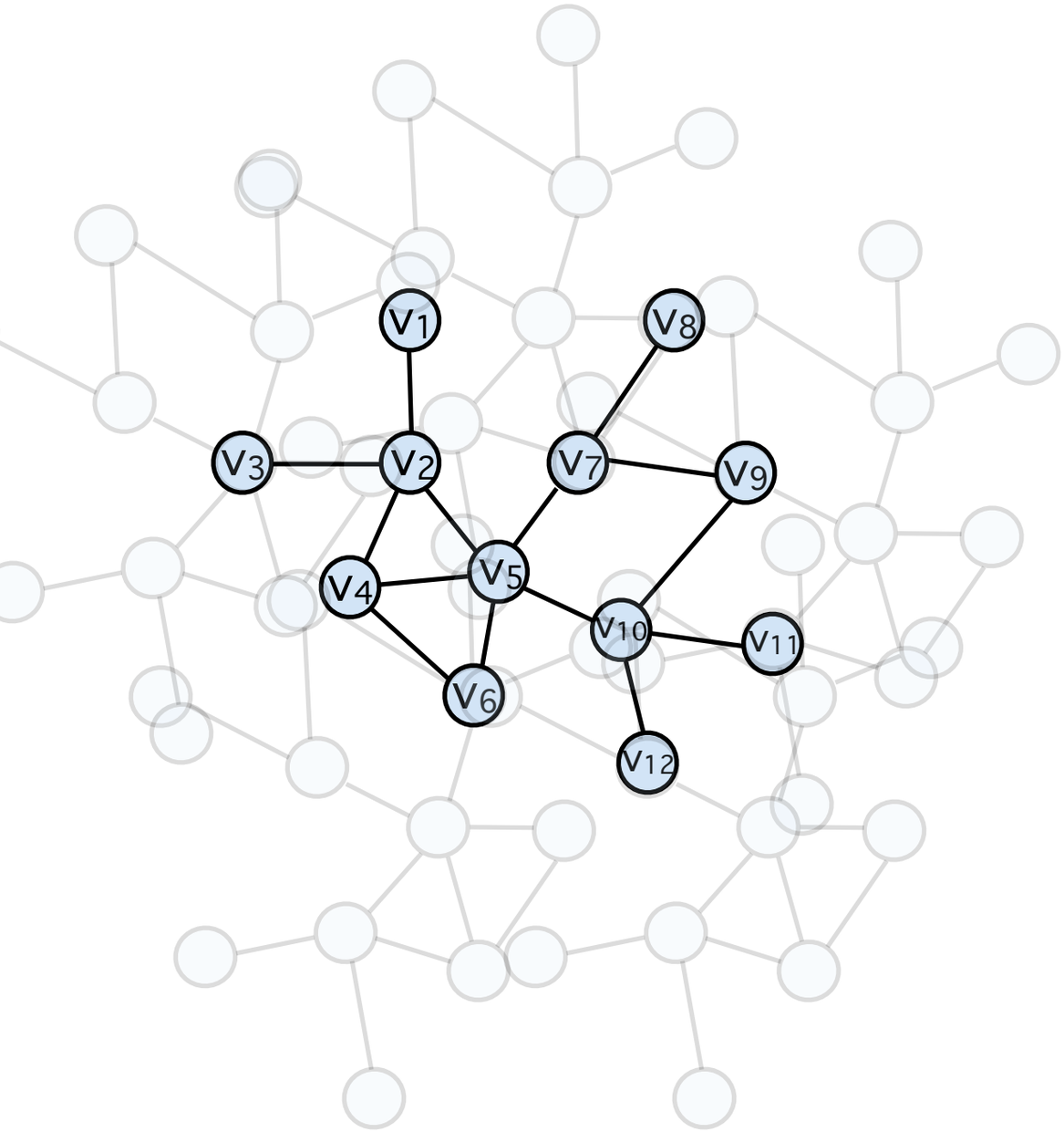
Do candidate proteins from my screen tend to interact with each other?



Measuring closeness in networks

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

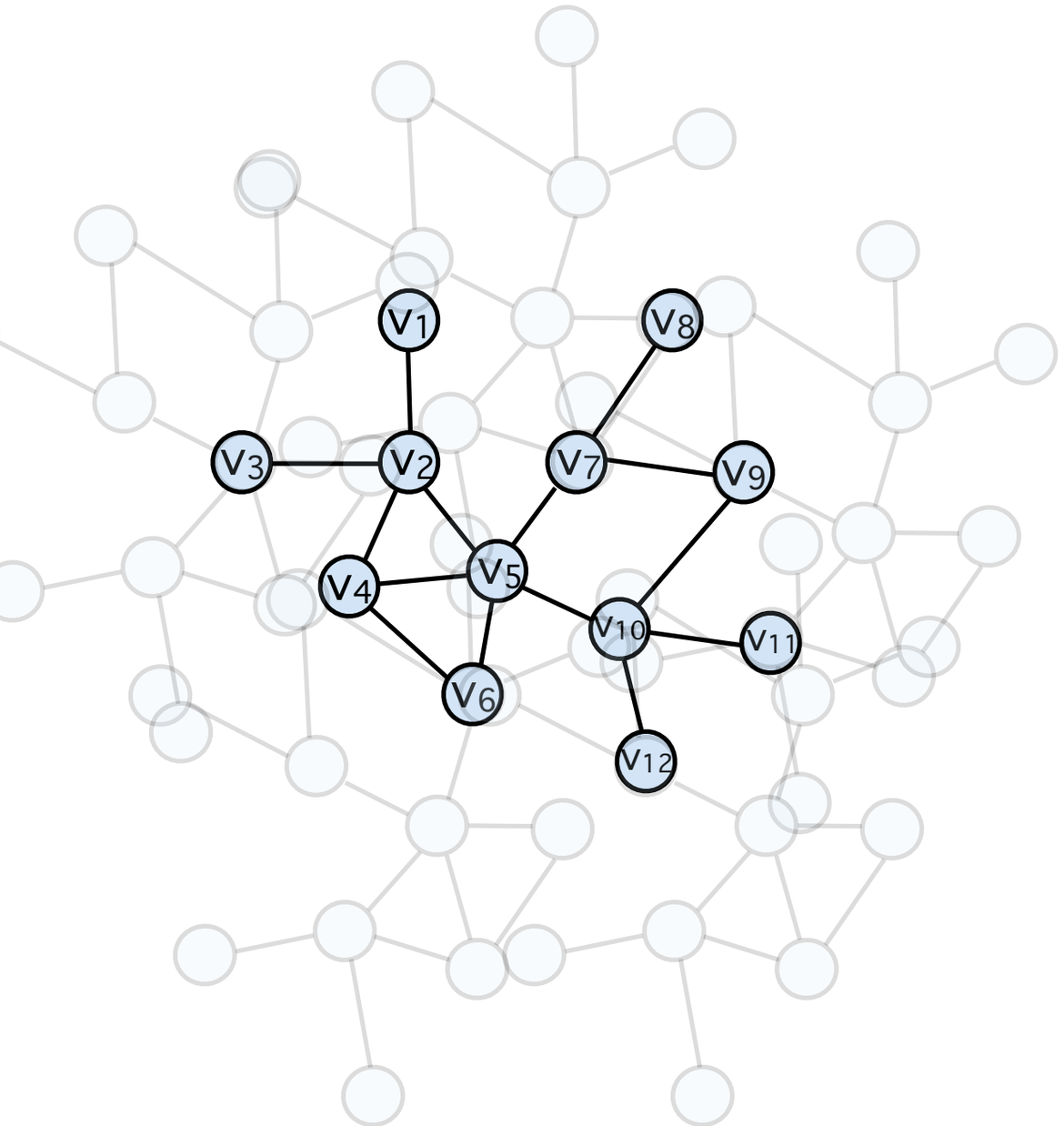
-> count number of edges between vertices that are candidate proteins
or
calculate average shortest path between them:



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?

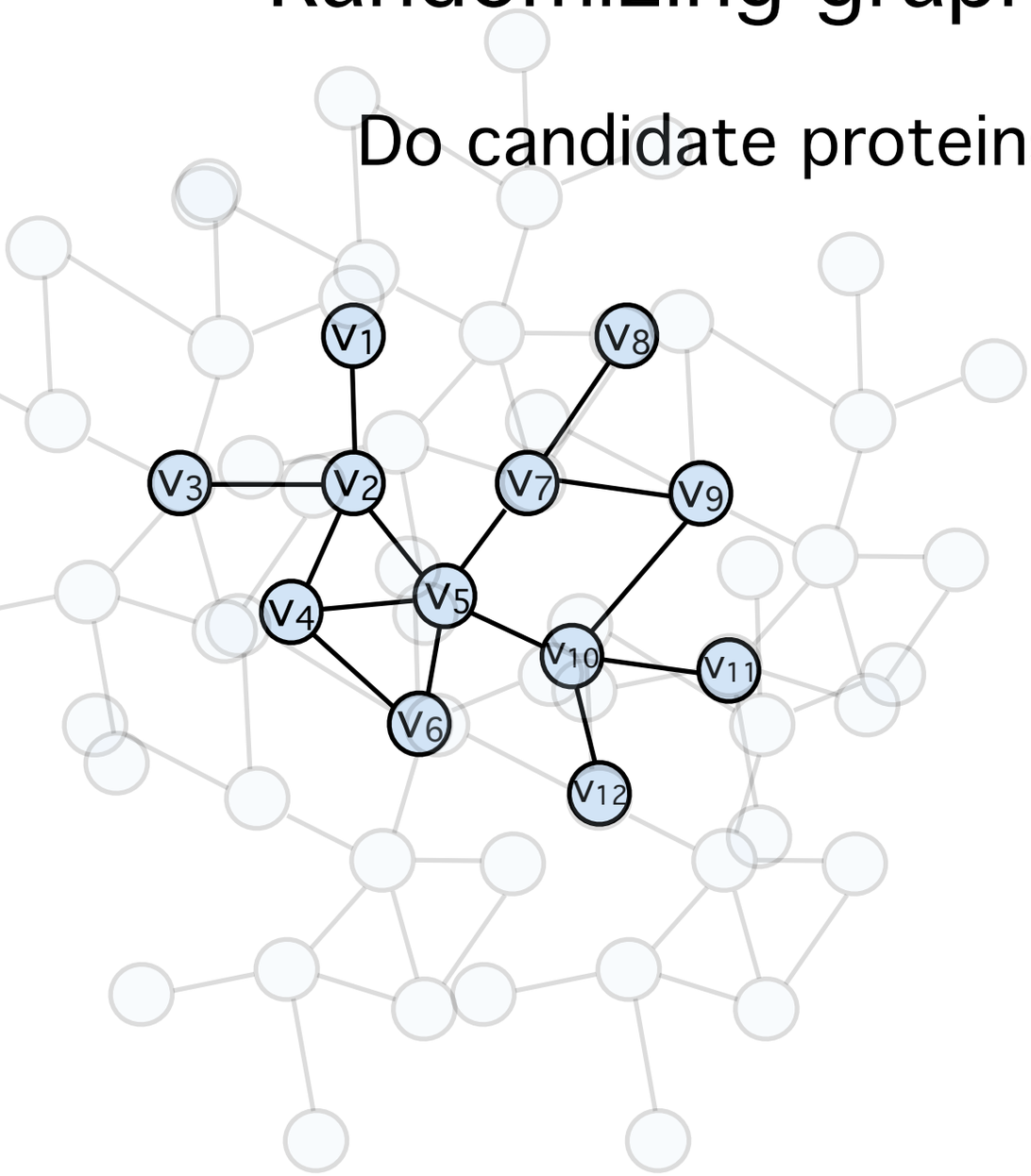
↓

Calculate the average shortest path:

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

Randomizing graphs to compute significances

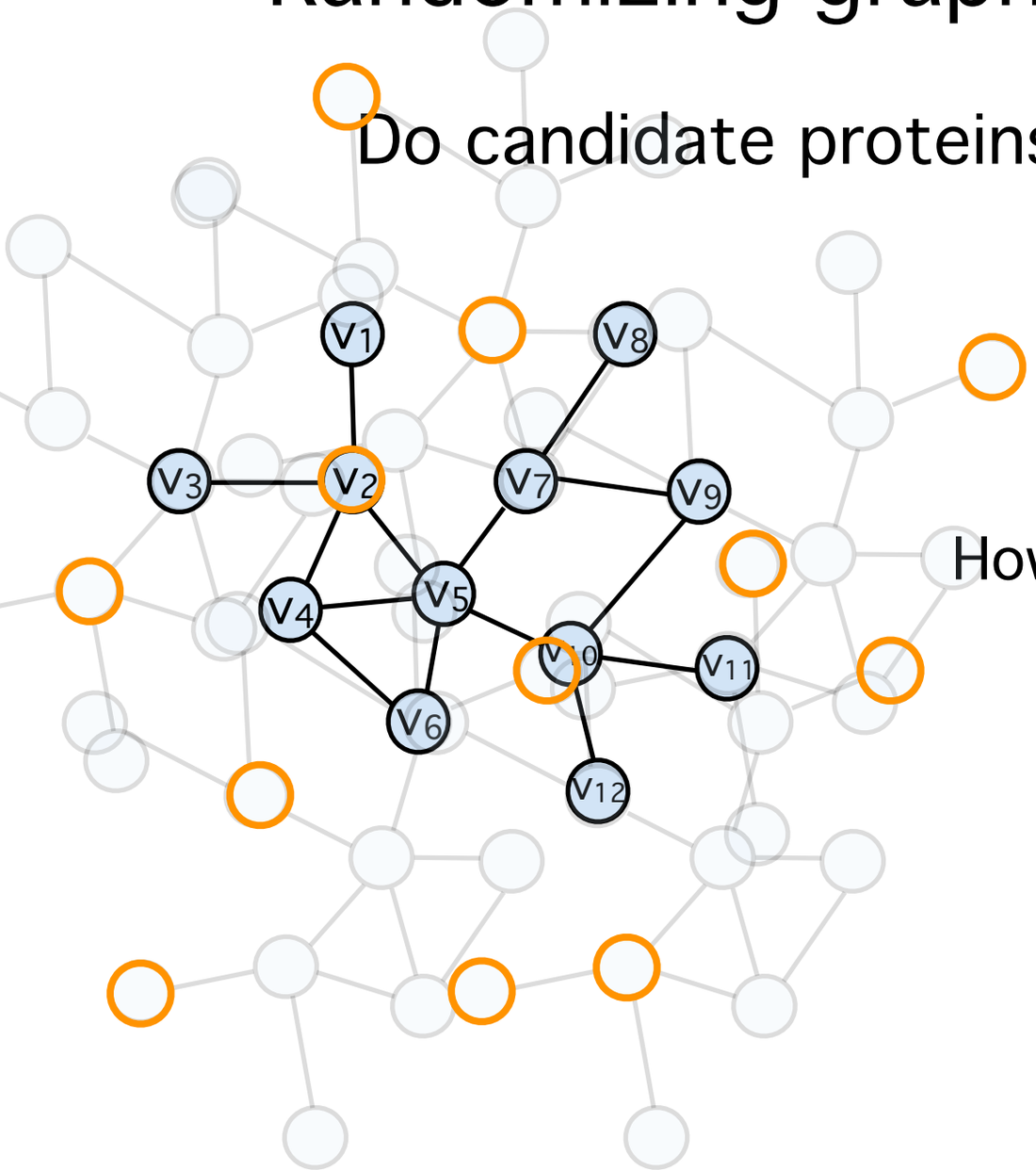
Do candidate proteins tend to interact with each other?



Number of edges: 14
Average shortest path: 2.17

Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



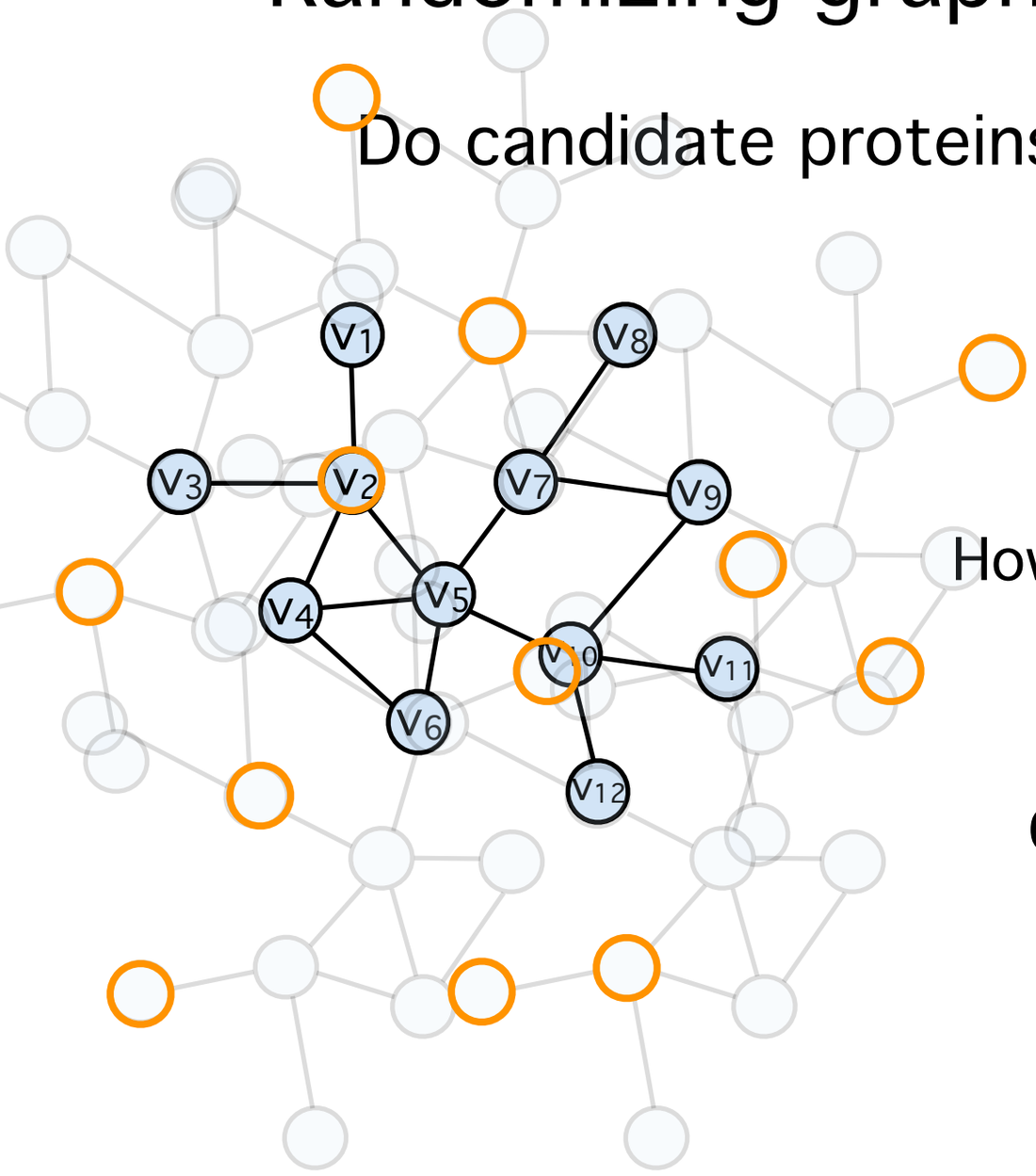
Number of edges: 14
Average shortest path: 2.17

↓

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

Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



Number of edges: 14
Average shortest path: 2.17

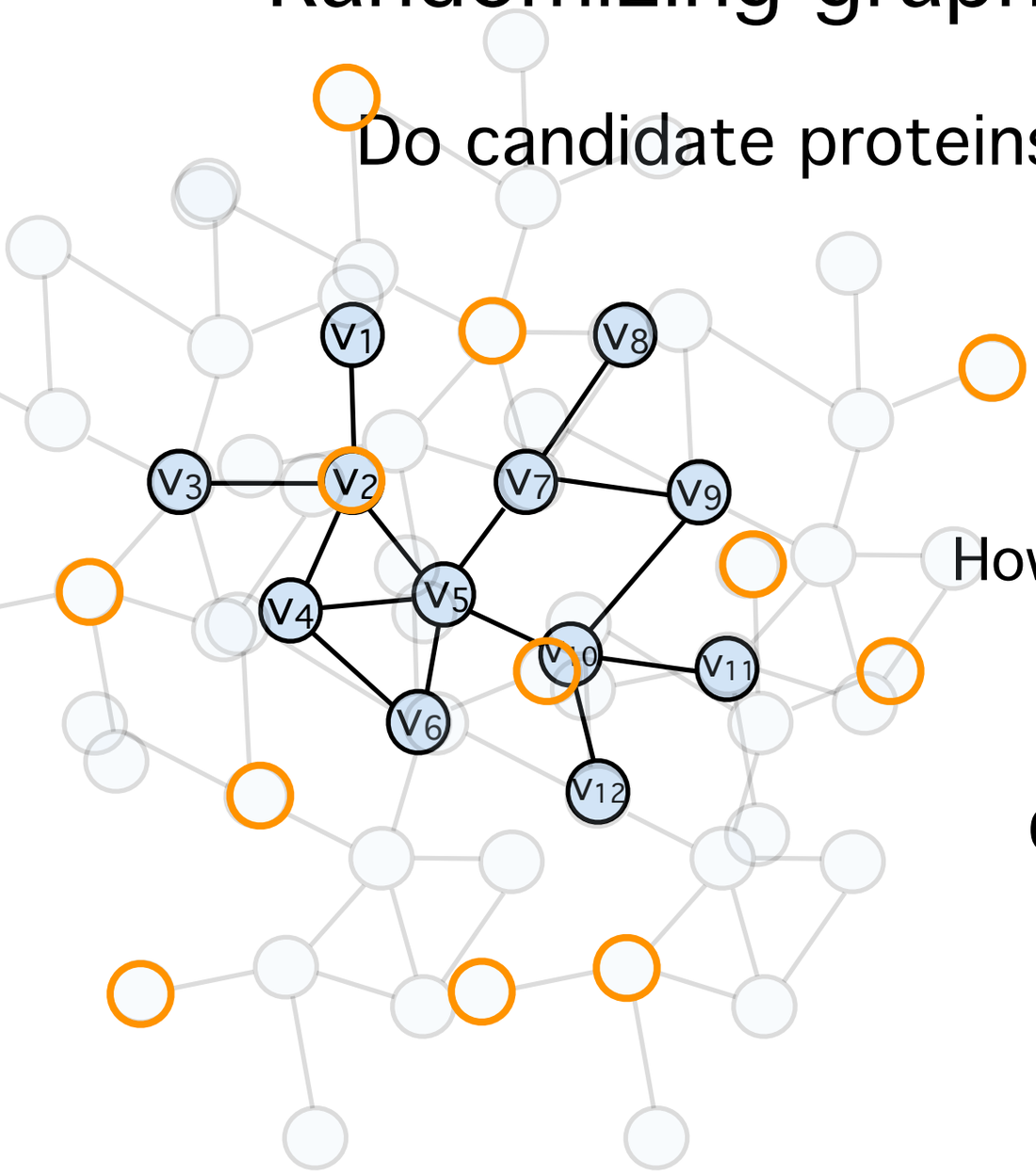
↓

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

↓

Can I randomly choose any 12 proteins in the network?

Randomizing graphs to compute significances



Do candidate proteins tend to interact with each other?

Number of edges: 14
Average shortest path: 2.17

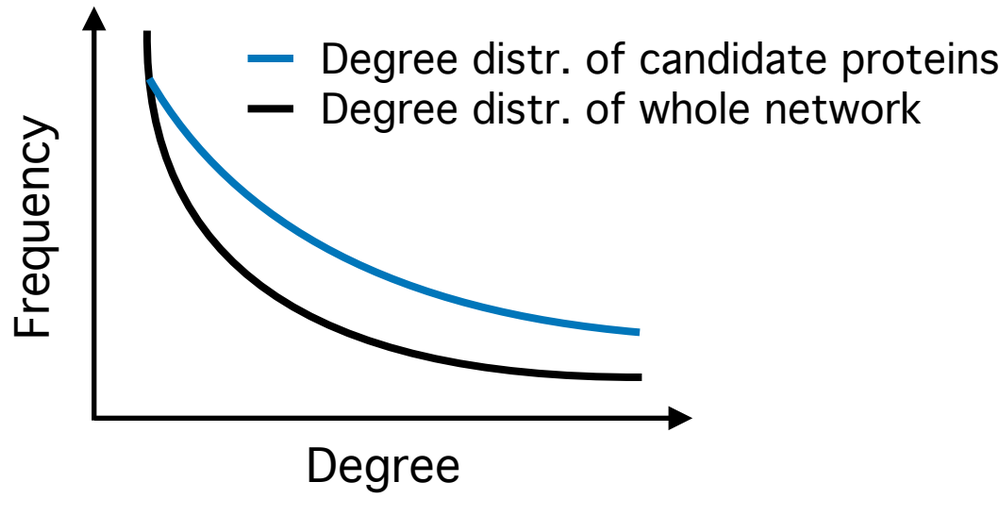
↓

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

↓

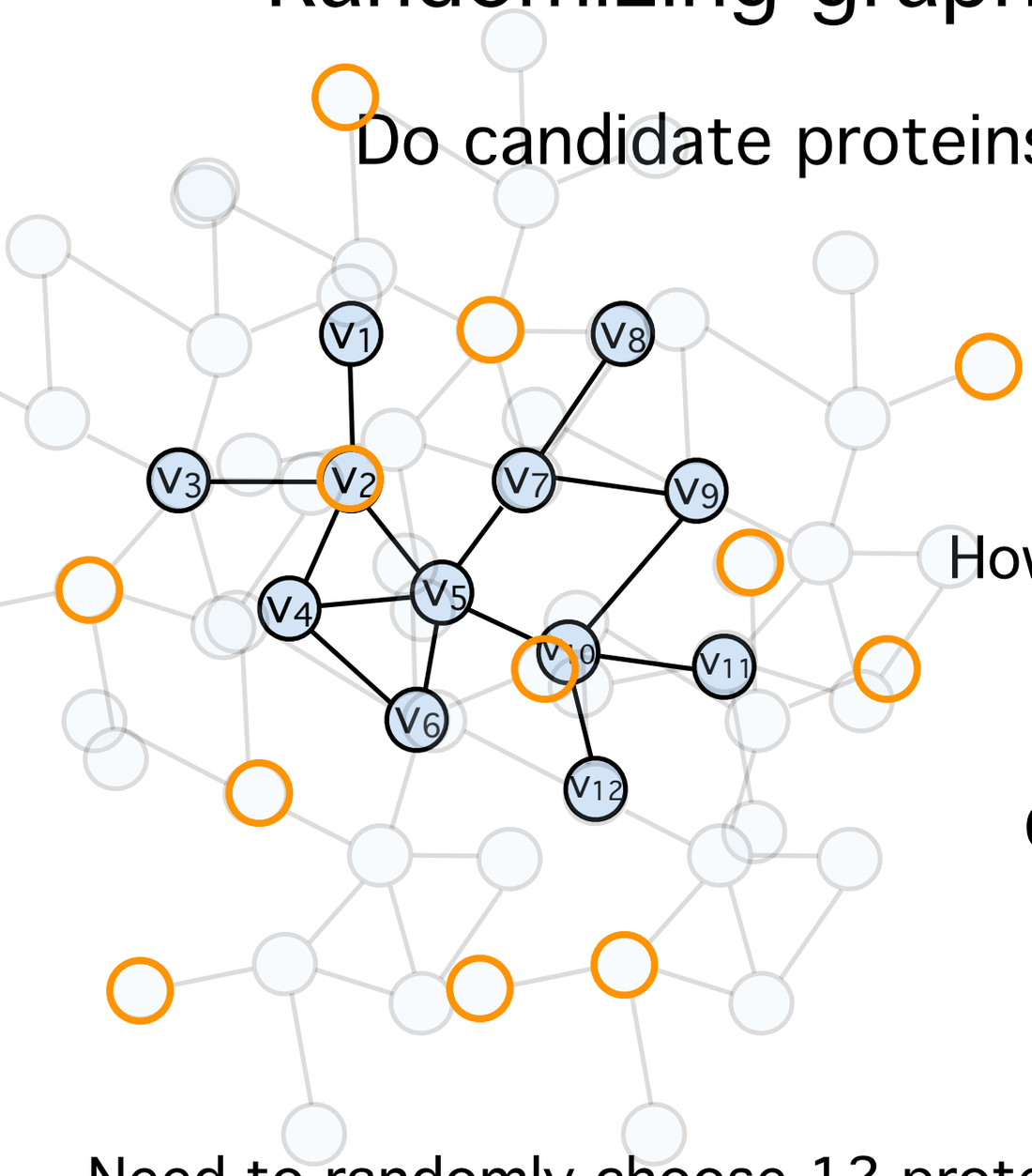
Can I randomly choose any 12 proteins in the network?

Possible scenario



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?

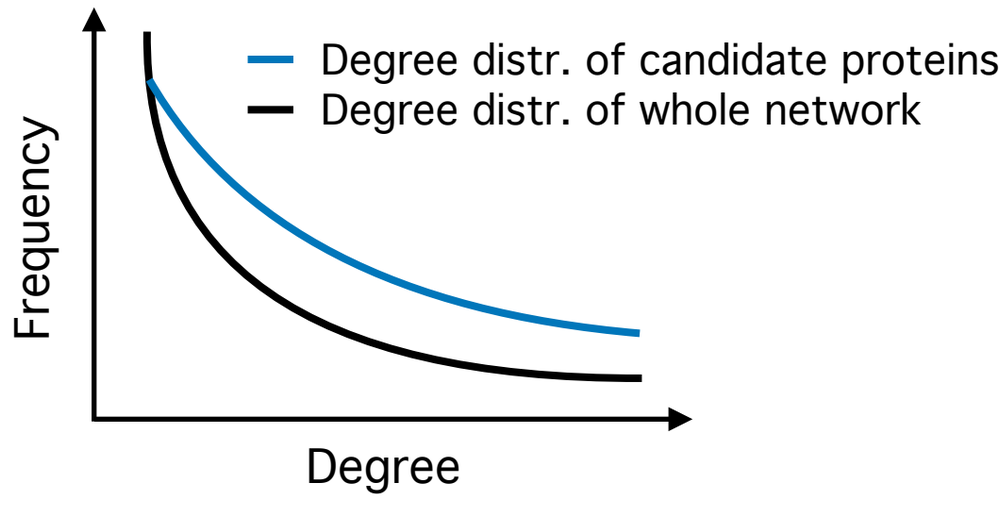
↓

Can I randomly choose any 12 proteins in the network?

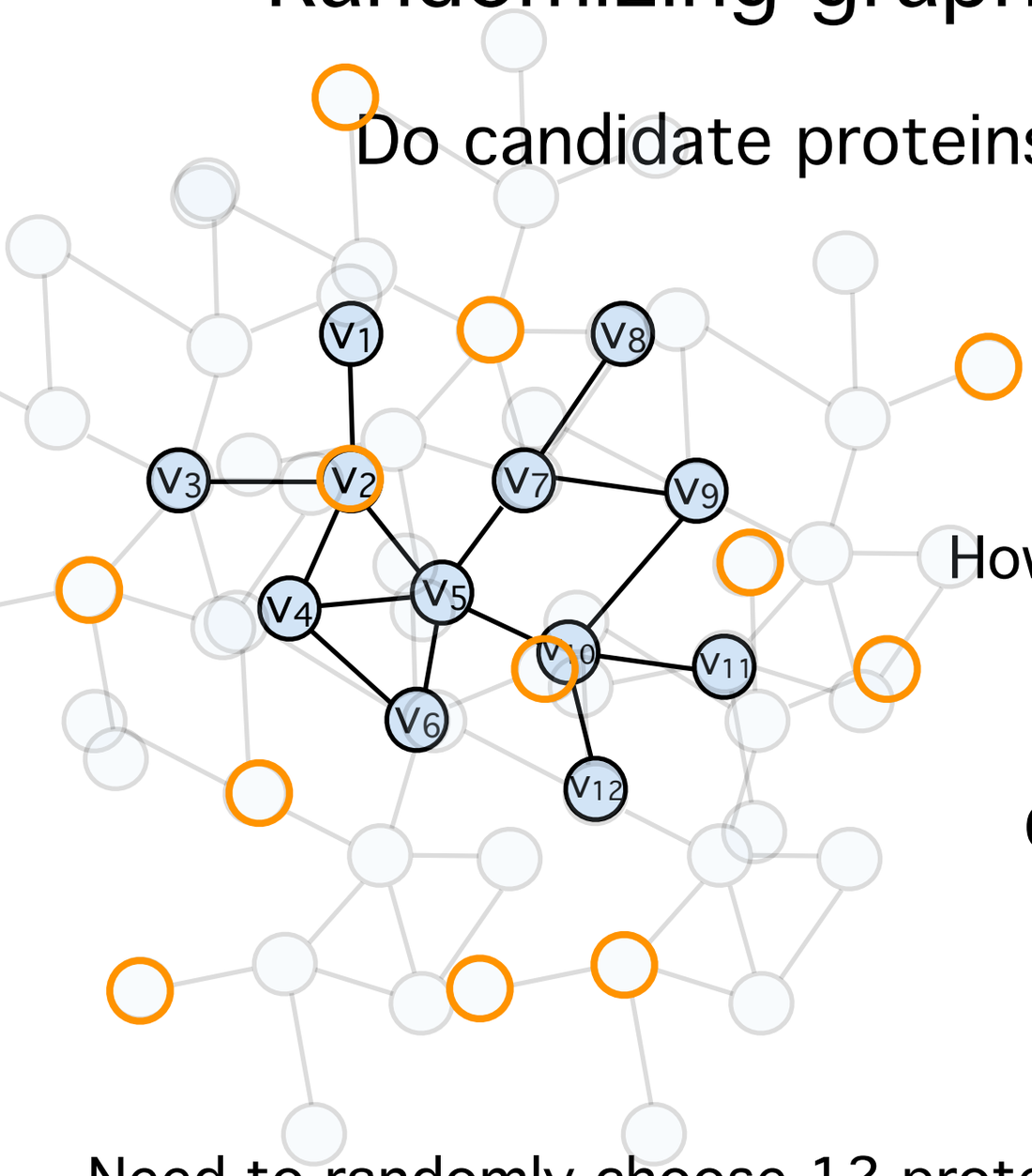
Need to randomly choose 12 proteins with the same degree distribution like candidate proteins

←

Possible scenario



Randomizing graphs to compute significances



Do candidate proteins tend to interact with each other?

Number of edges: 14
Average shortest path: 2.17

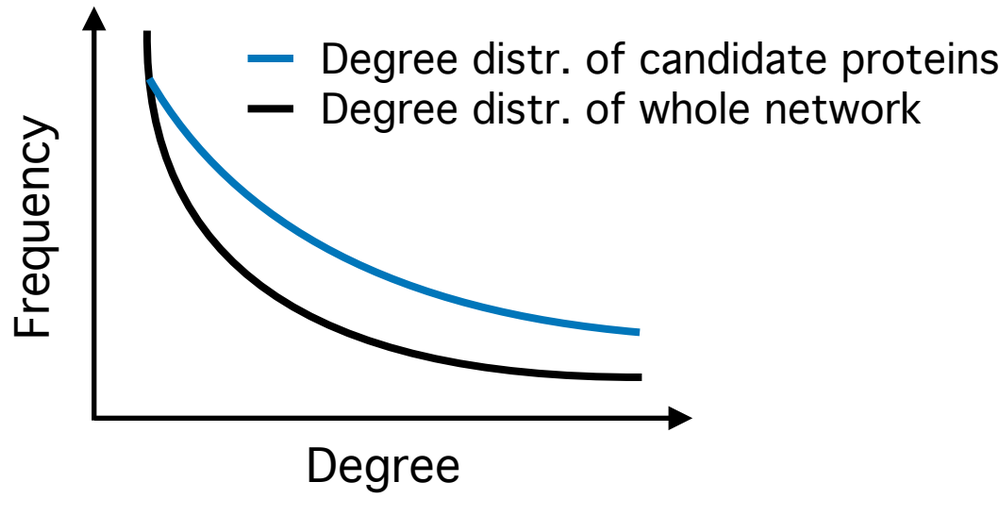
↓

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

↓

Can I randomly choose any 12 proteins in the network?

Possible scenario



←

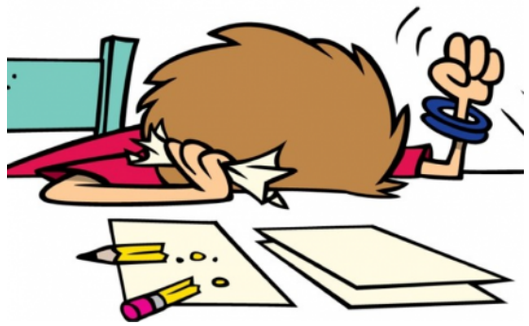
Need to randomly choose 12 proteins with the same degree distribution like candidate proteins

Hard

Randomizing graphs to compute significances

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

Hard



Randomizing graphs to compute significances

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

Hard

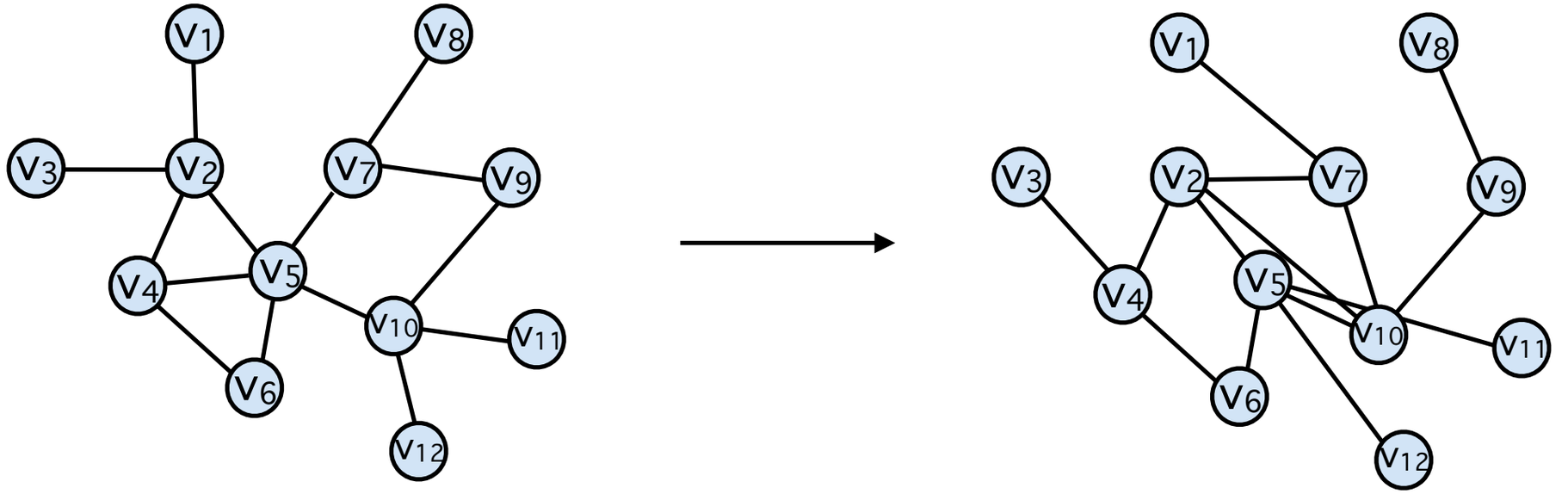
↓

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

Randomizing graphs to compute significances

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

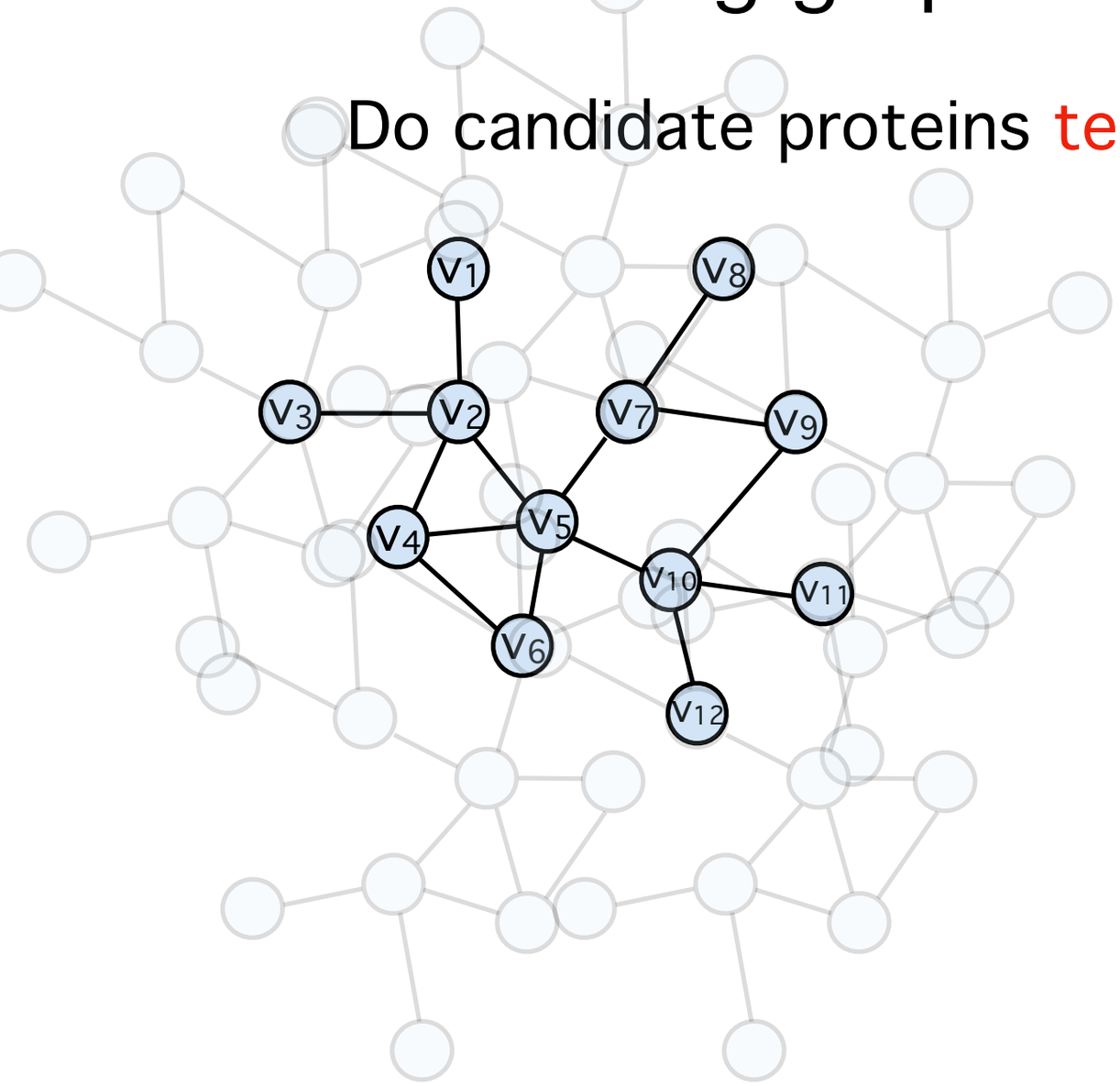
Hard



Edges are shuffled such that every vertex maintains its degree

Randomizing graphs to compute significances

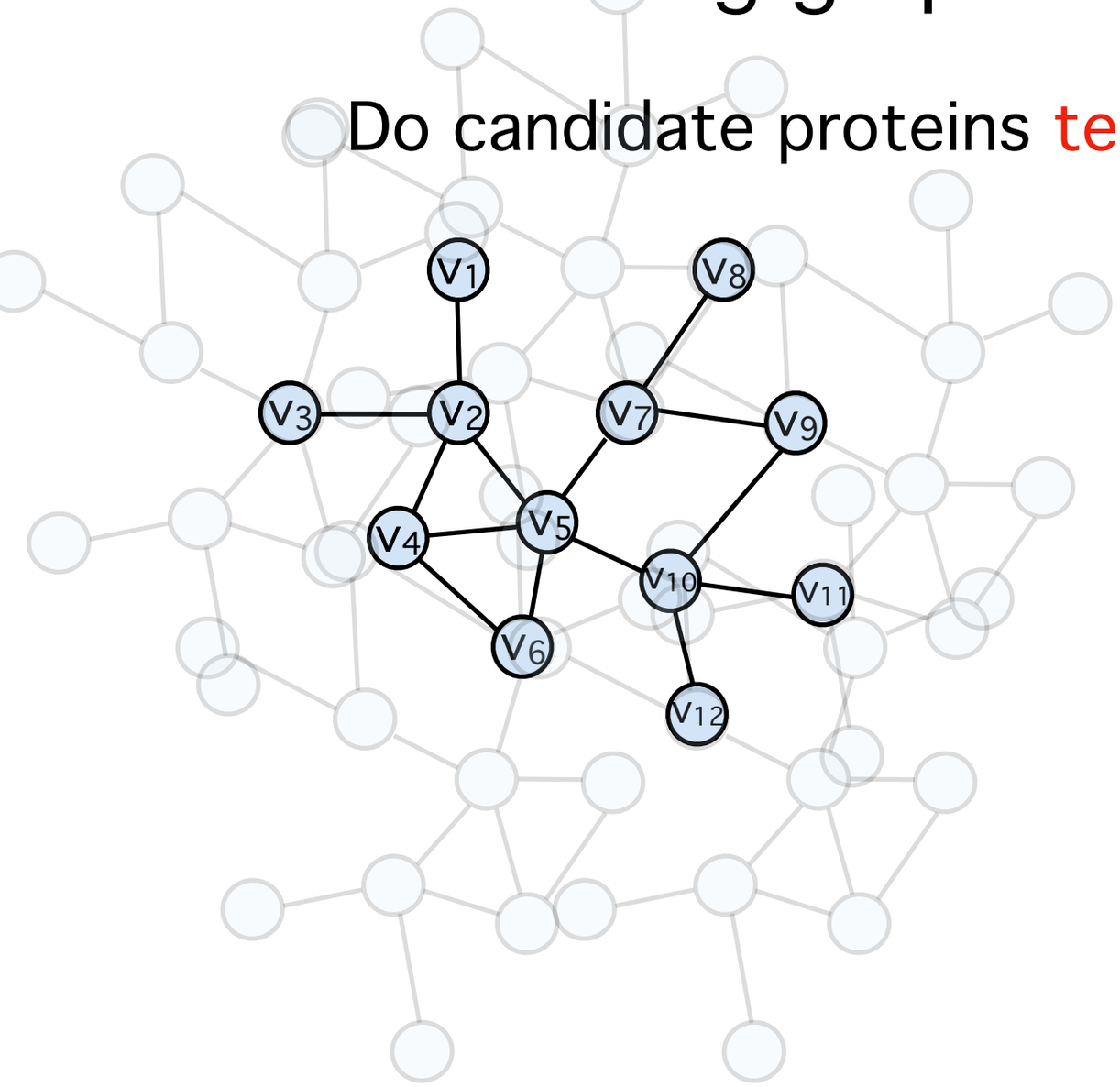
Do candidate proteins tend to interact with each other?



Number of edges: 14
Average shortest path: 2.17

Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



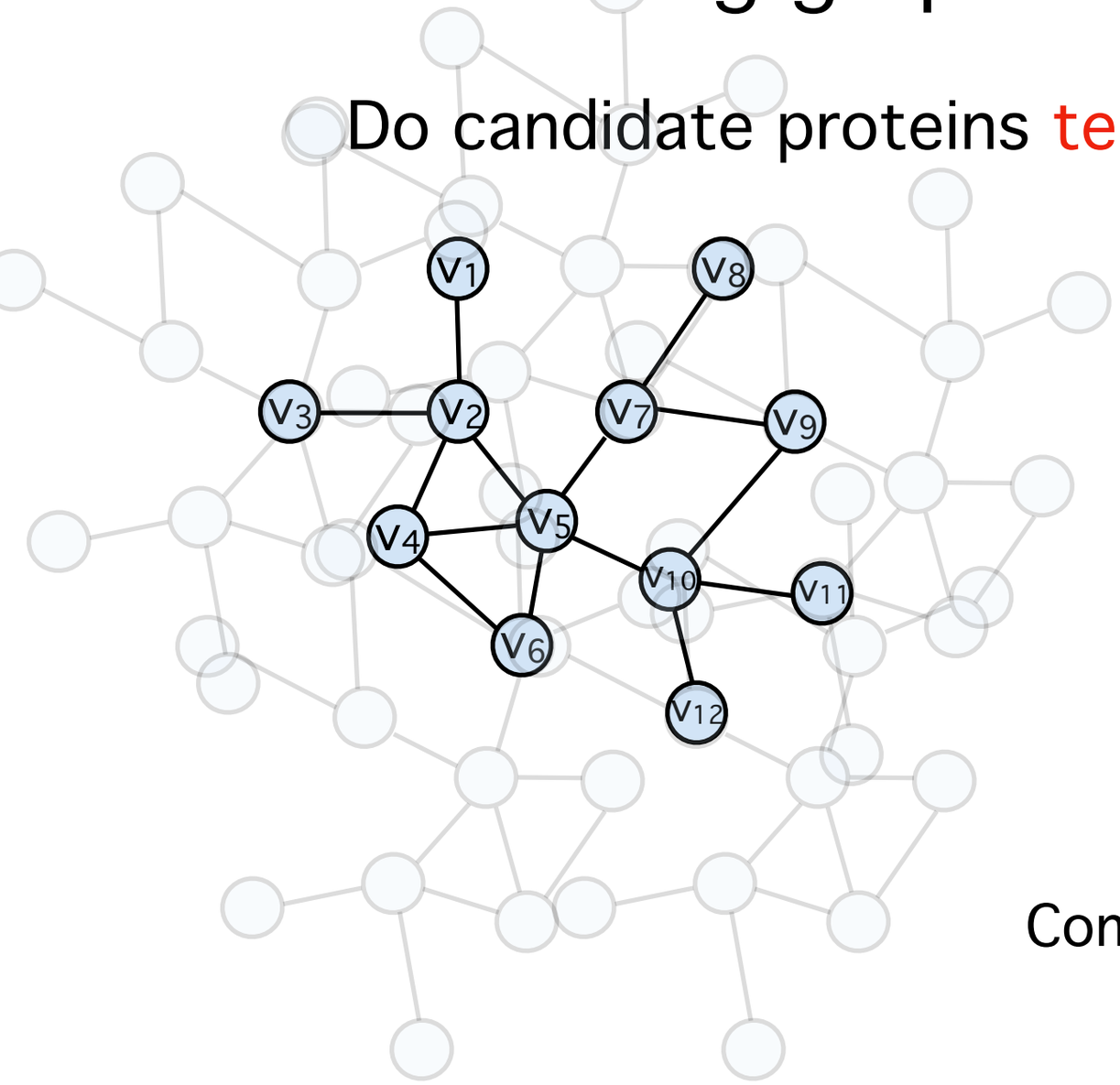
Number of edges: 14
Average shortest path: 2.17

↓

Generate a high number of degree-controlled randomized networks

Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



Number of edges: 14
Average shortest path: 2.17

↓

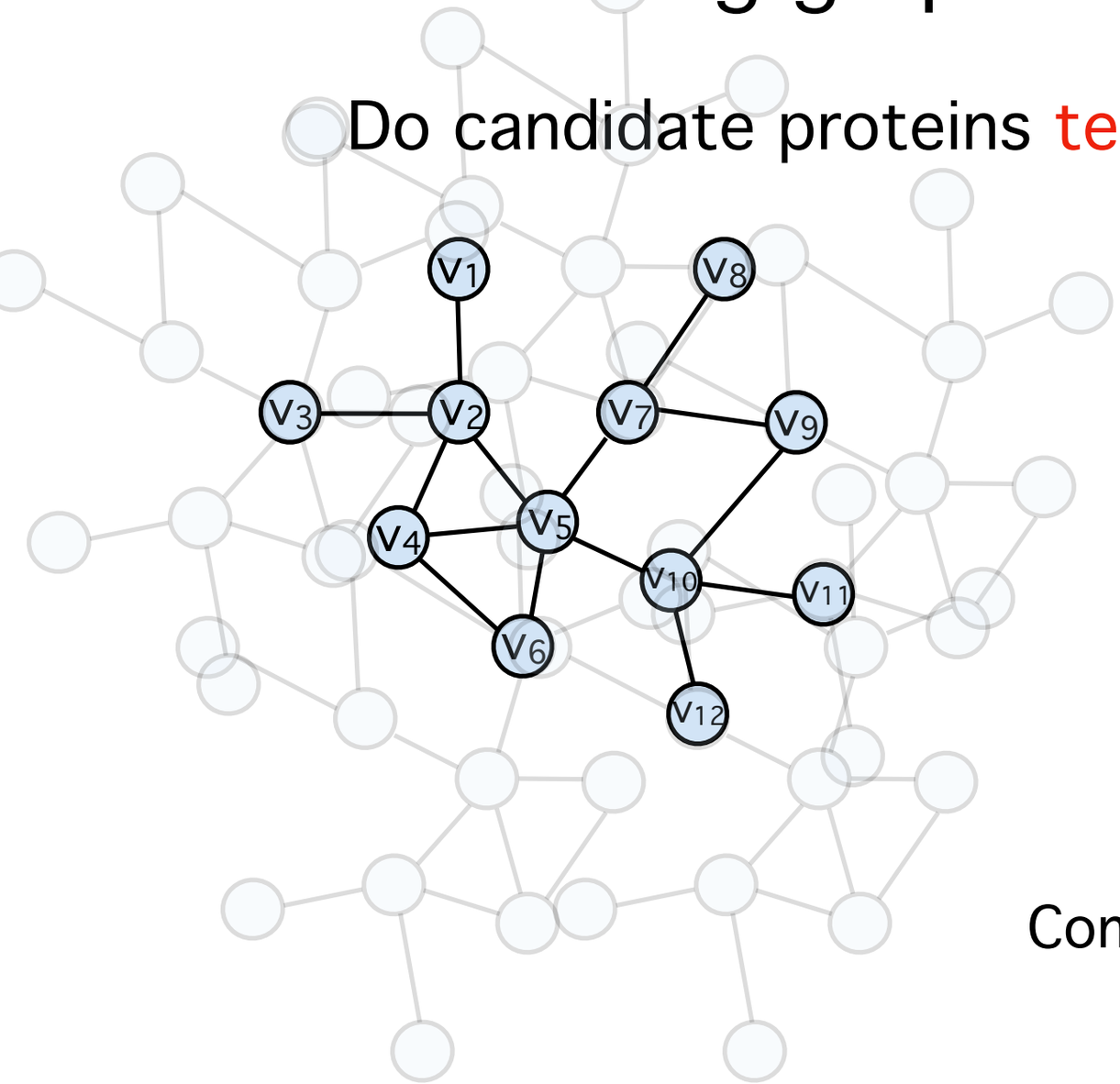
Generate a high number of degree-controlled randomized networks

↓

Compute closeness of candidate proteins in each of them

Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



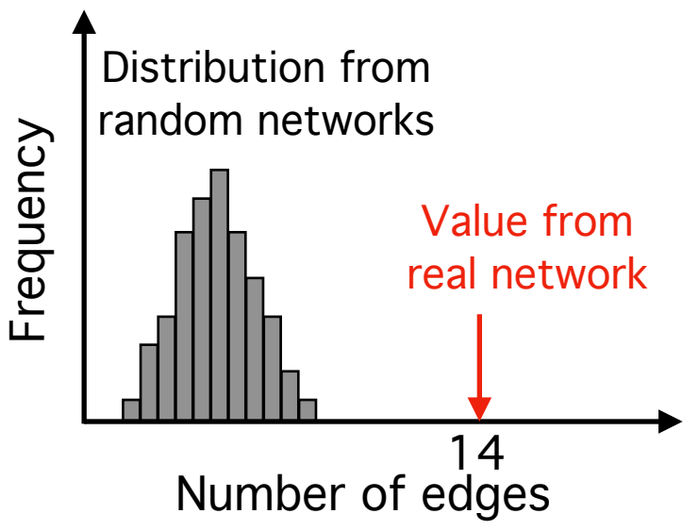
Number of edges: 14
Average shortest path: 2.17

↓

Generate a high number of degree-controlled randomized networks

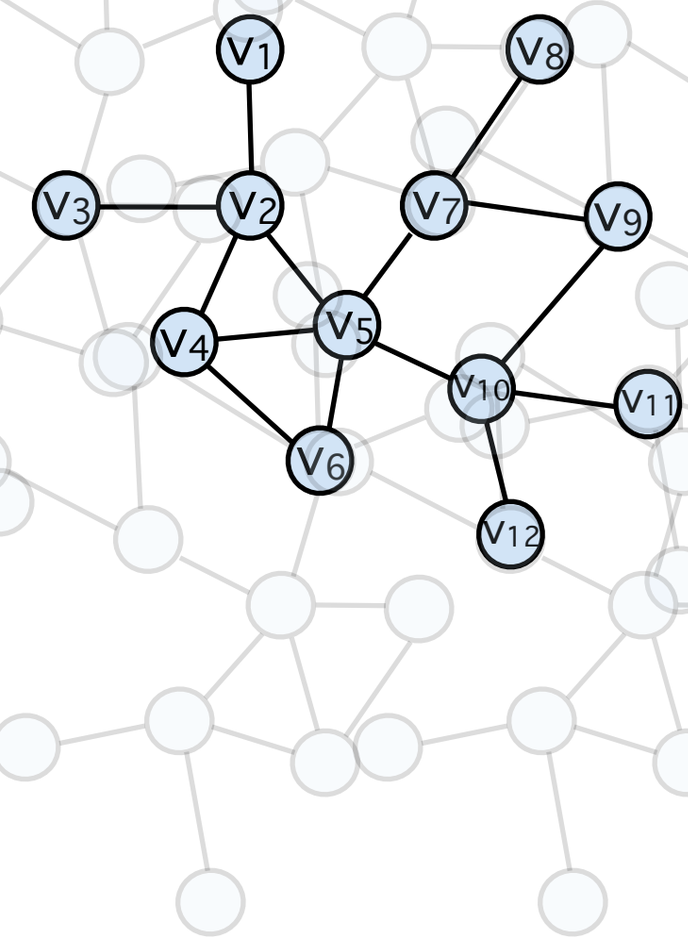
↓

Compute closeness of candidate proteins in each of them



Randomizing graphs to compute significances

Do candidate proteins tend to interact with each other?



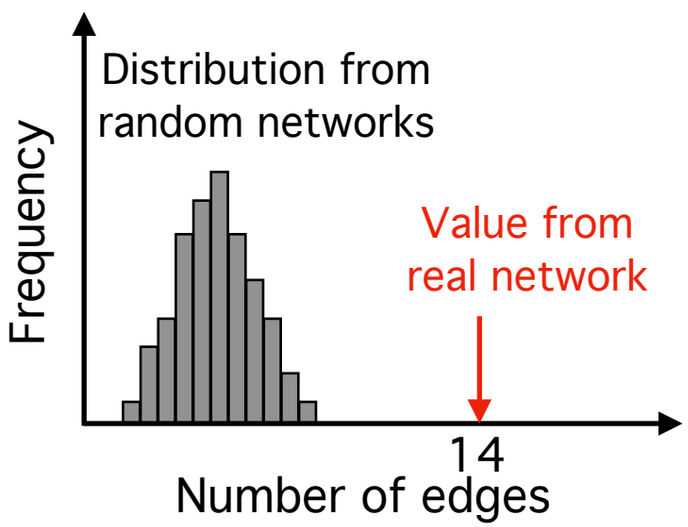
Number of edges: 14
Average shortest path: 2.17

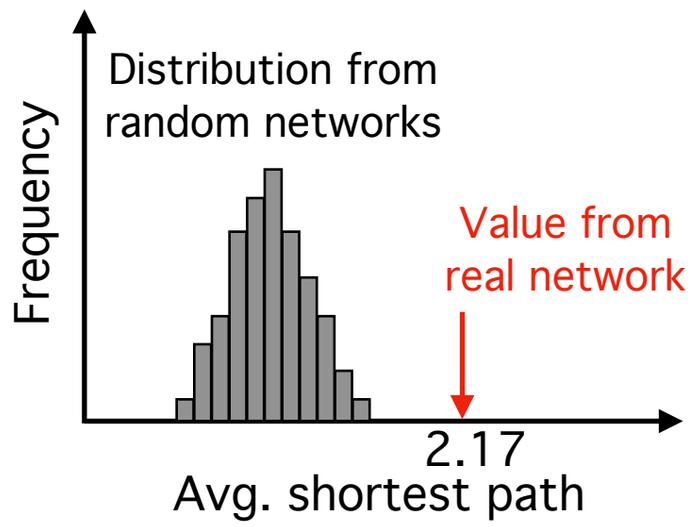
↓

Generate a high number of degree-controlled randomized networks

↓

Compute closeness of candidate proteins in each of them



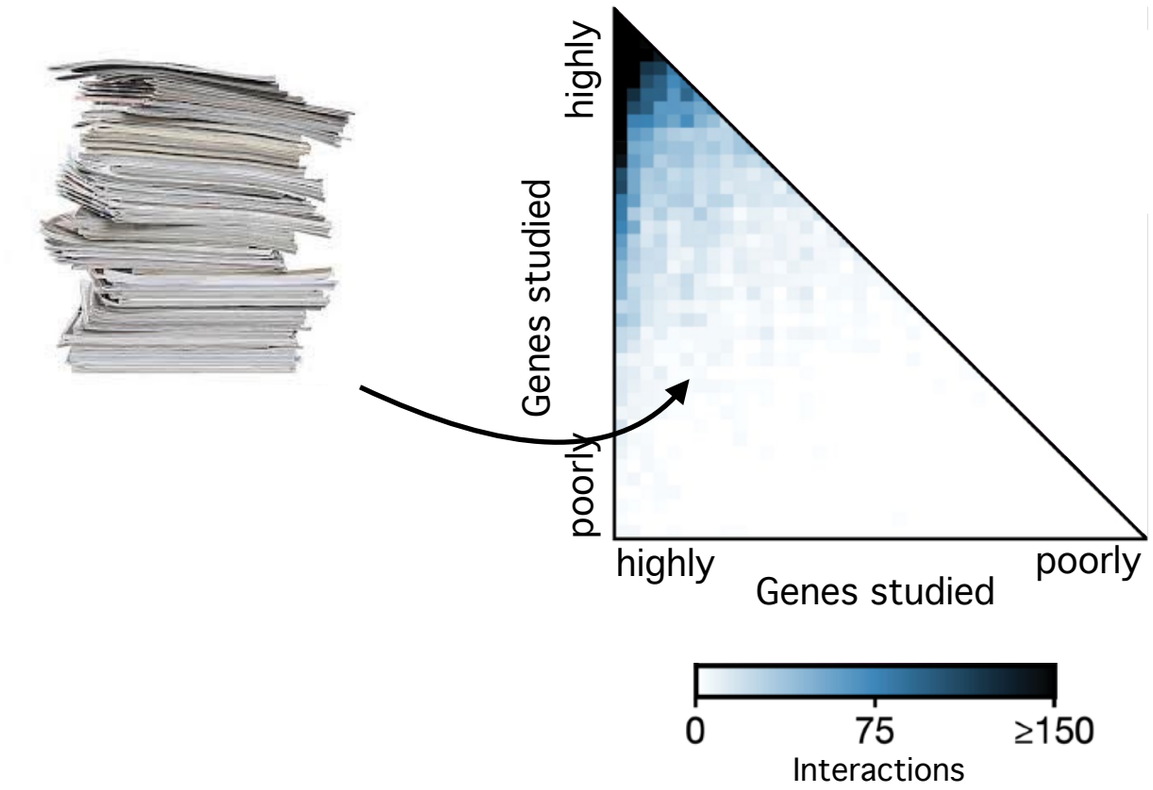


Study bias in curated protein interaction data can falsify network analyses

Luck et al *Nature* 2020

Study bias in curated protein interaction data can falsify network analyses

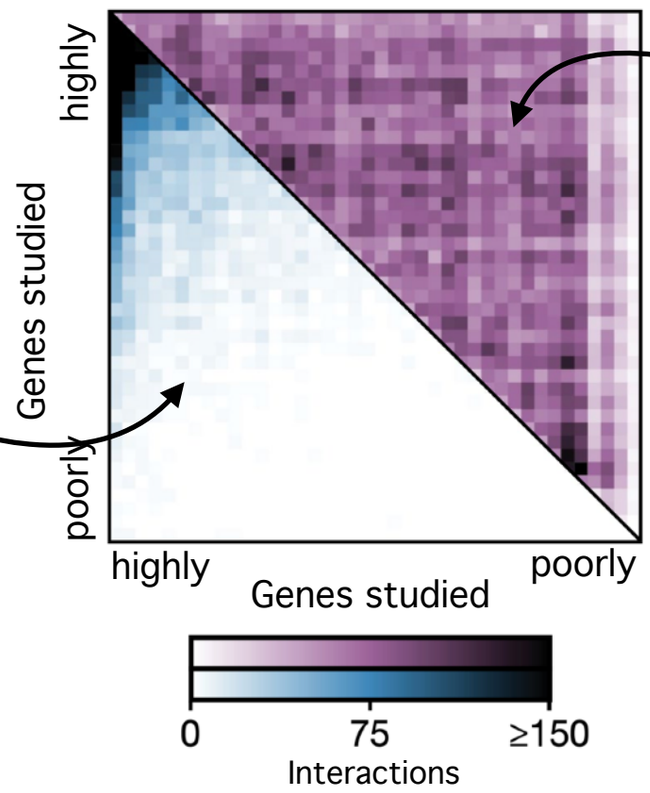
Literature curation



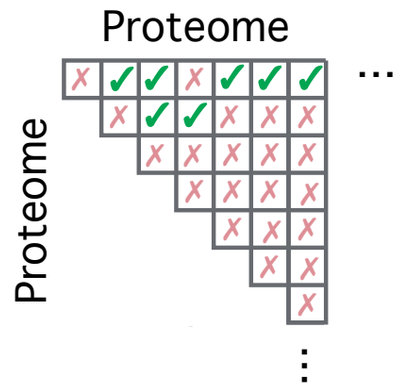
Luck et al *Nature* 2020

Study bias in curated protein interaction data can falsify network analyses

Literature curation



Systematic mapping

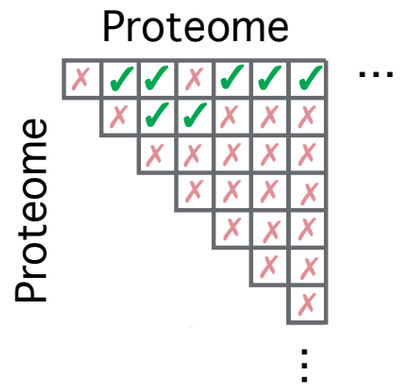


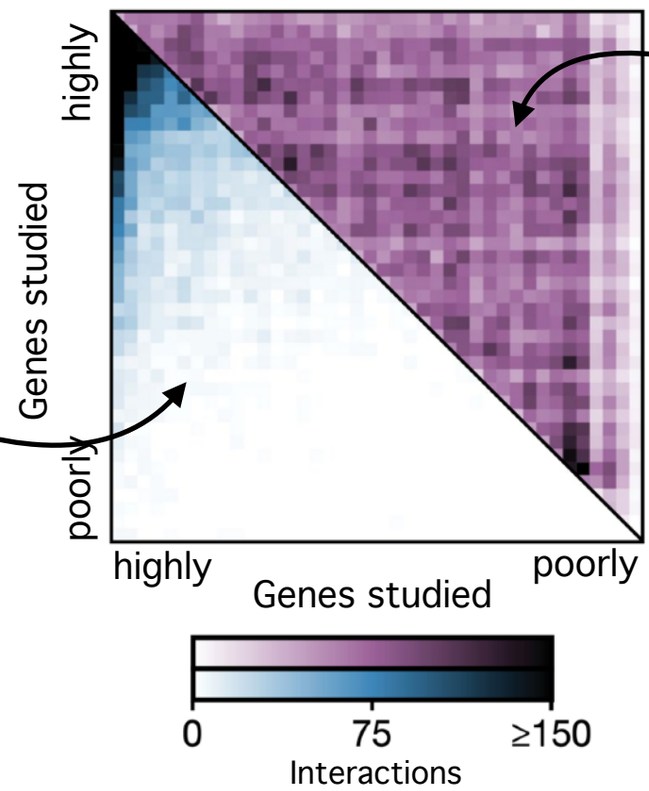
Luck et al *Nature* 2020

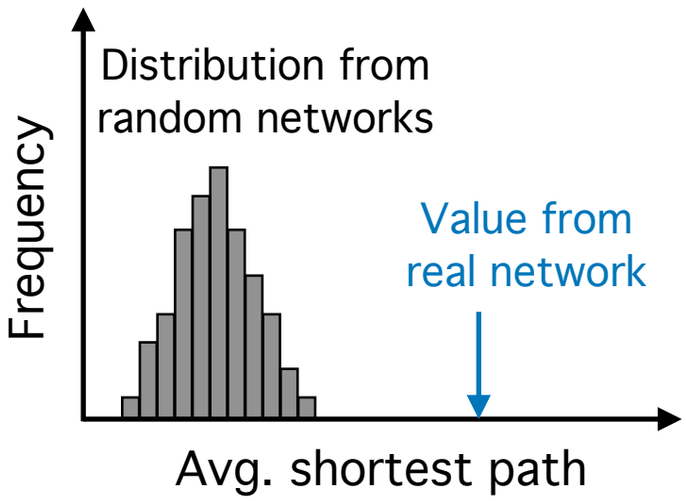
Study bias in curated protein interaction data can falsify network analyses

Literature curation

Systematic mapping





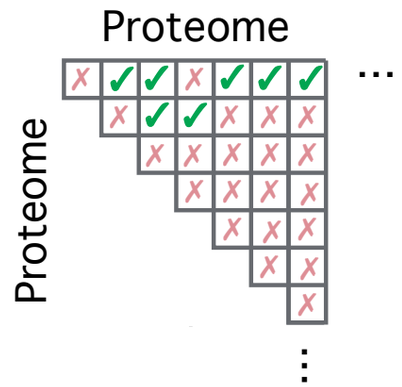


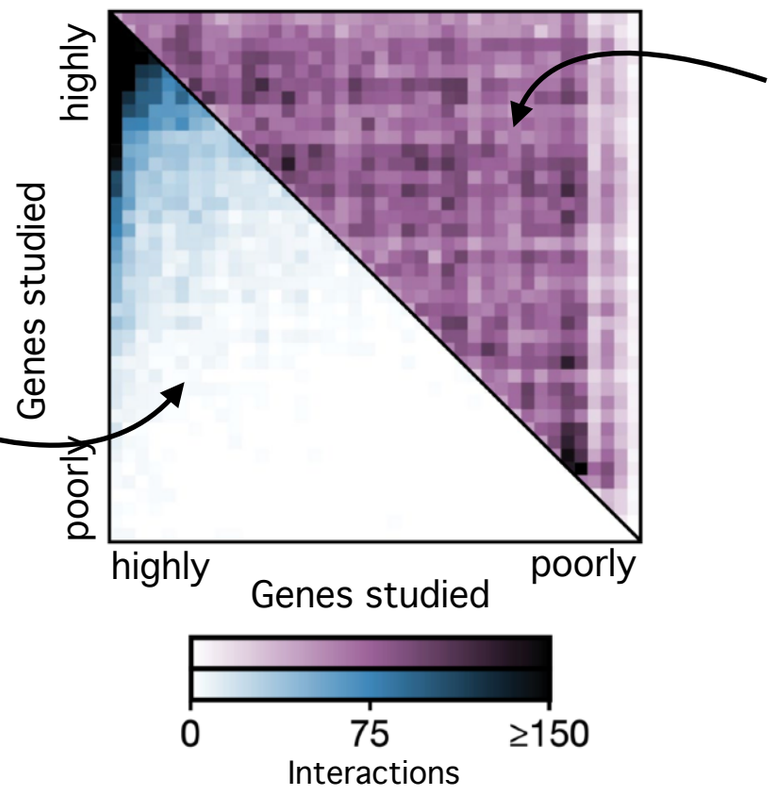
Luck et al *Nature* 2020

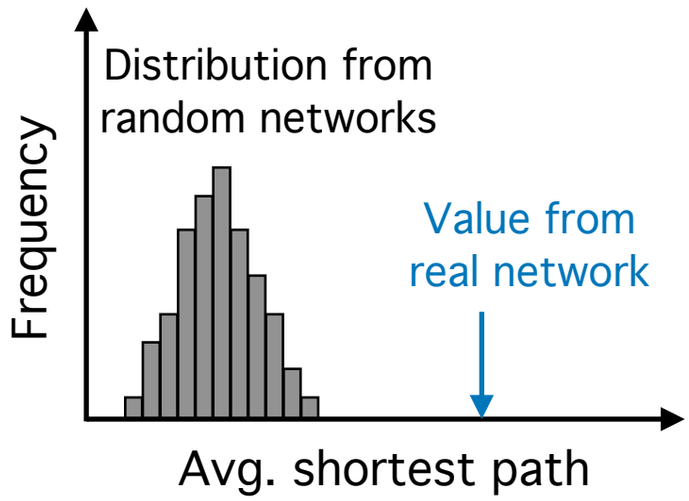
Study bias in curated protein interaction data can falsify network analyses

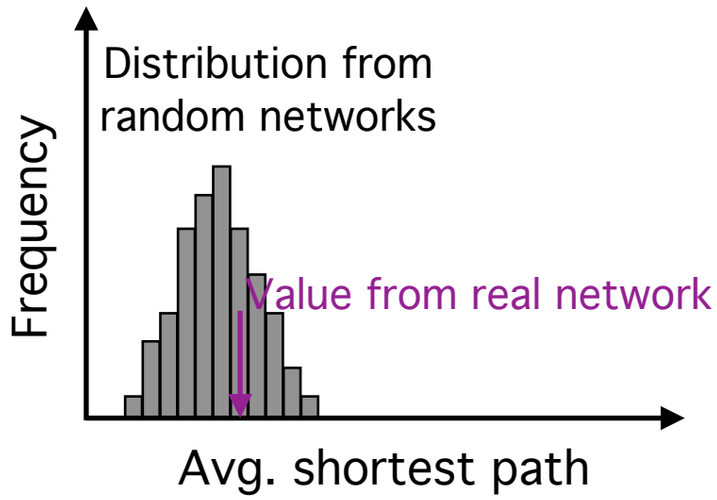
Literature curation

Systematic mapping







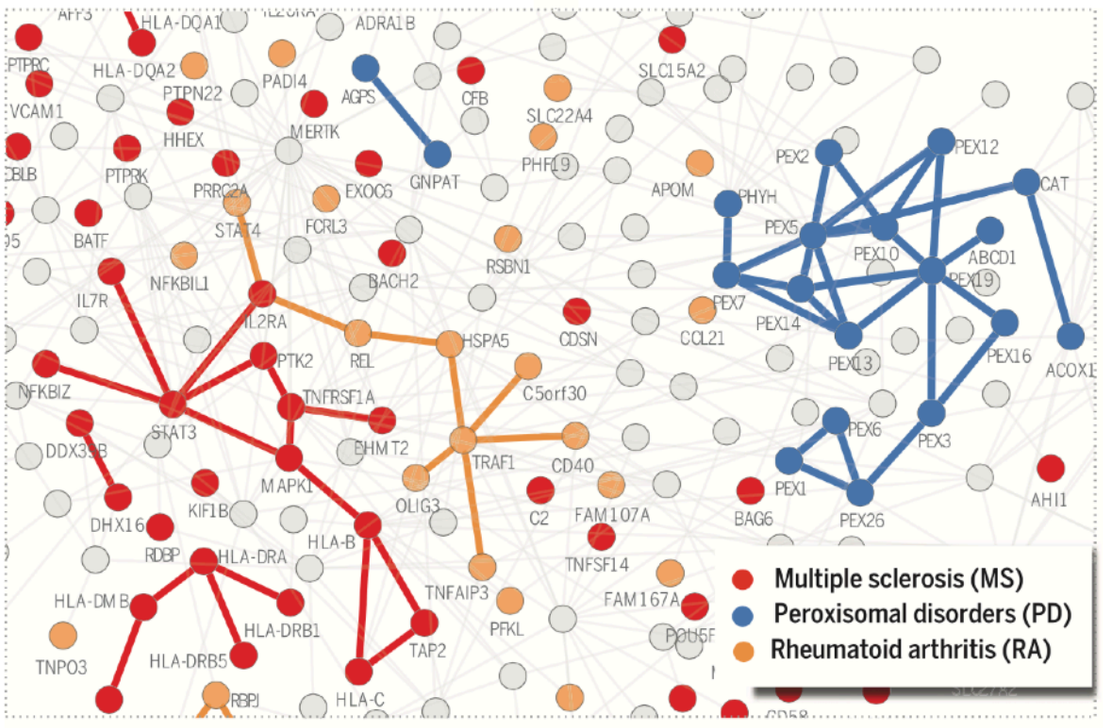


Luck et al *Nature* 2020

Network closeness of disease genes and tissue-specific proteins

Uncovering disease-disease relationships through the incomplete interactome

Jörg Menche, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, Albert-László Barabási*



Science 2015

A reference map of the human binary protein interactome

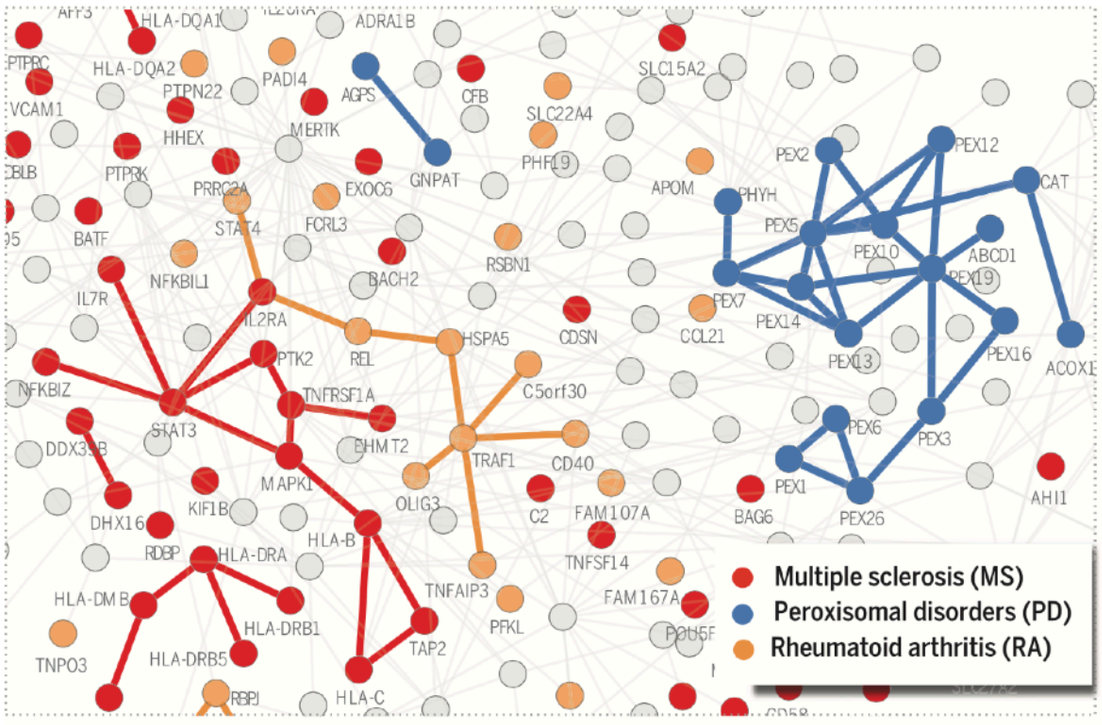
Katja Luck[1,2,3,33], Dae-Kyum Kim[1,4,5,6,33], Luke Lambourne[1,2,3,33], Kerstin Spirohn[1,2,3,33], David E. Hill[1,2,3]✉, Marc Vidal[1,2]✉, Frederick P. Roth[1,4,5,6,16,32]✉ & Michael A. Calderwood[1,2,3]✉

Nature 2020

Network closeness of disease genes and tissue-specific proteins

Uncovering disease-disease relationships through the incomplete interactome

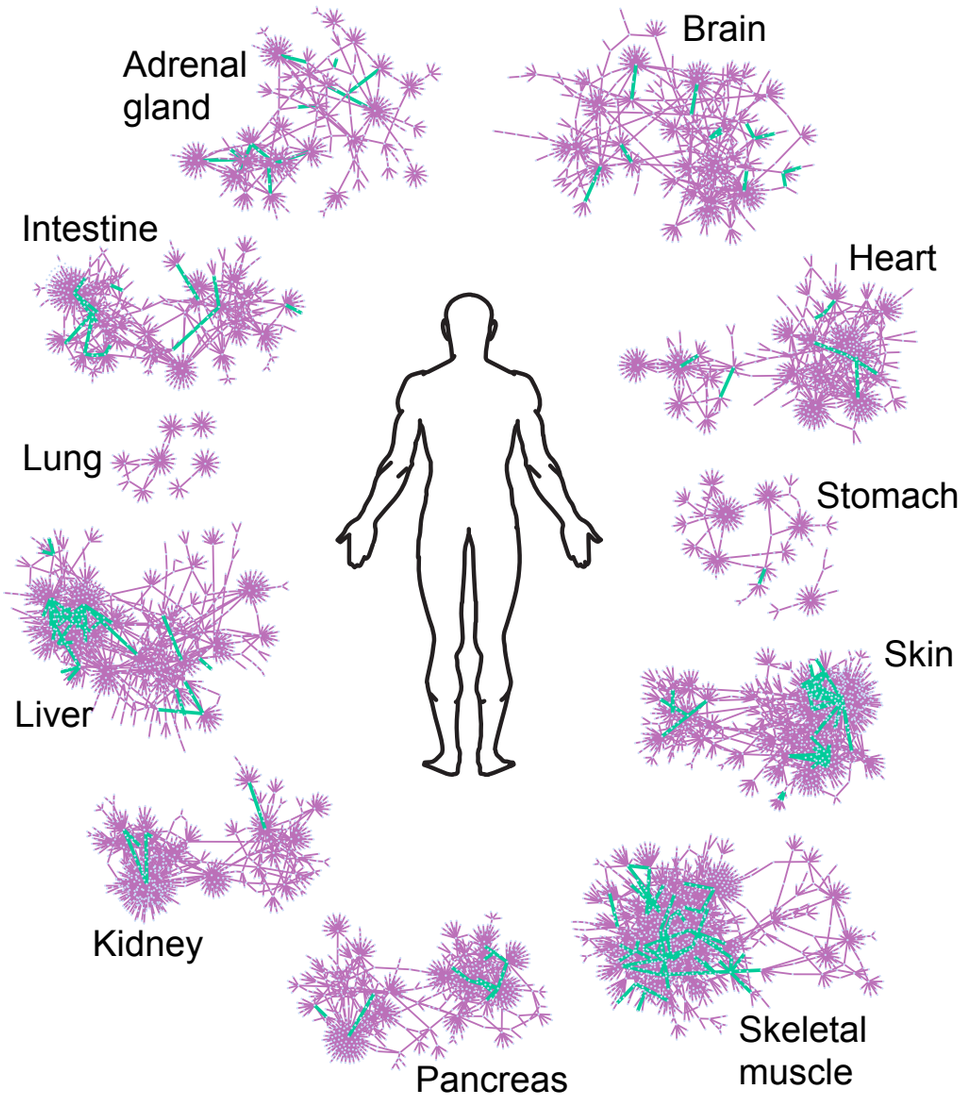
Jörg Menche, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, Albert-László Barabási*



Science 2015

A reference map of the human binary protein interactome

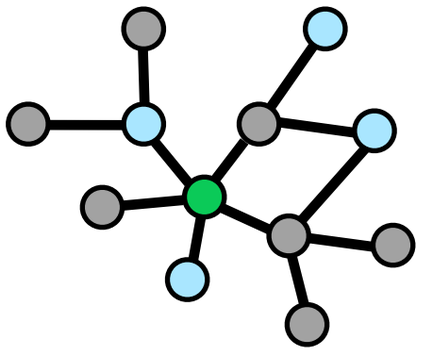
Katja Luck[1,2,3,33], Dae-Kyum Kim[1,4,5,6,33], Luke Lambourne[1,2,3,33], Kerstin Spirohn[1,2,3,33], David E. Hill[1,2,3]✉, Marc Vidal[1,2]✉, Frederick P. Roth[1,4,5,6,16,32]✉ & Michael A. Calderwood[1,2,3]✉



Nature 2020

What is the function of my gene of interest?

Guilt-by-association

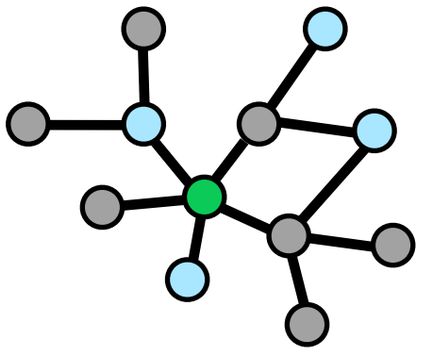


- Candidate protein
- Known apoptosis function
- other

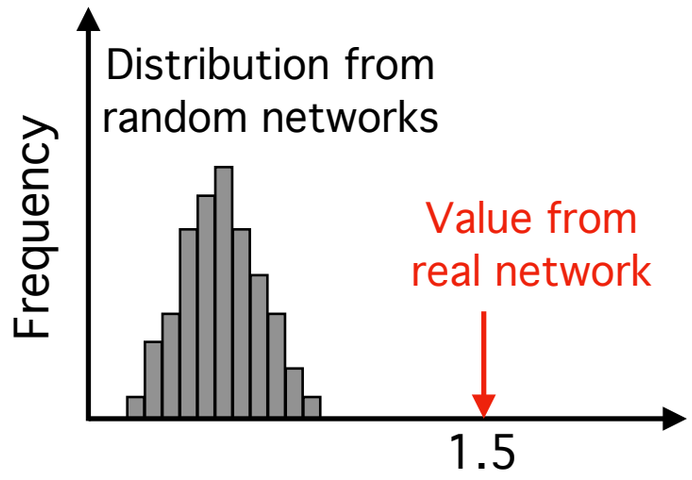
Fuxman Bass et al *Nature Methods* 2013, Luck et al *Nature* 2020

What is the function of my gene of interest?

Guilt-by-association



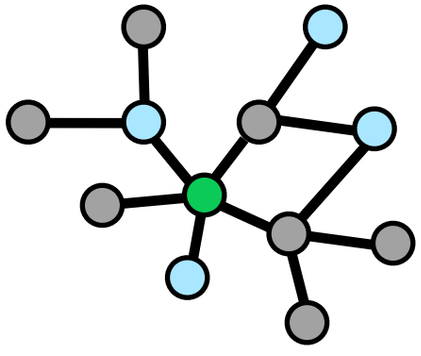
Candidate protein
Known apoptosis function
other



Fuxman Bass et al *Nature Methods* 2013, Luck et al *Nature* 2020

What is the function of my gene of interest?

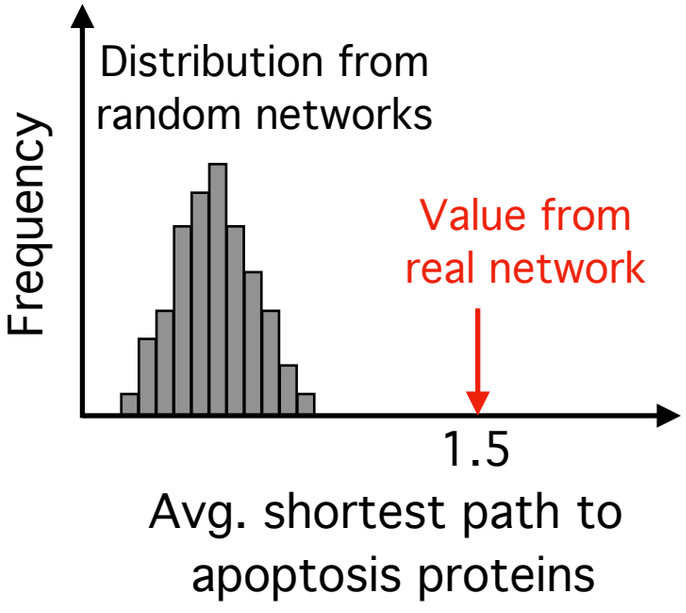
Guilt-by-association



Candidate protein

Known apoptosis function

other

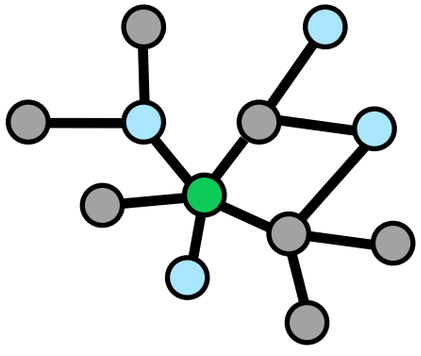


OTU deubiquitinase 6A

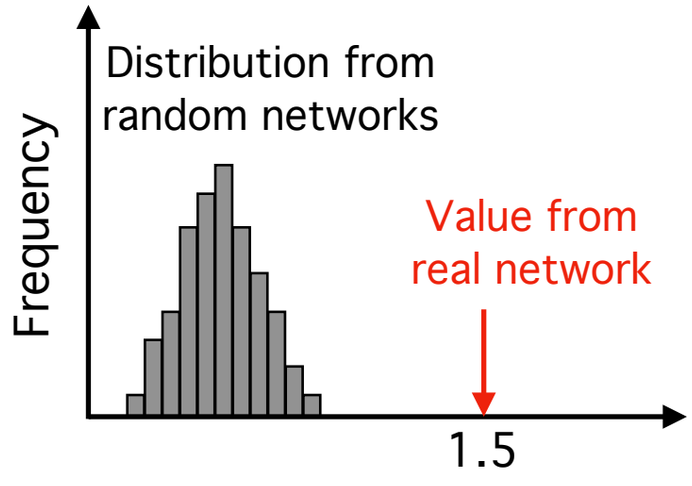
Fuxman Bass et al *Nature Methods* 2013, Luck et al *Nature* 2020

What is the function of my gene of interest?

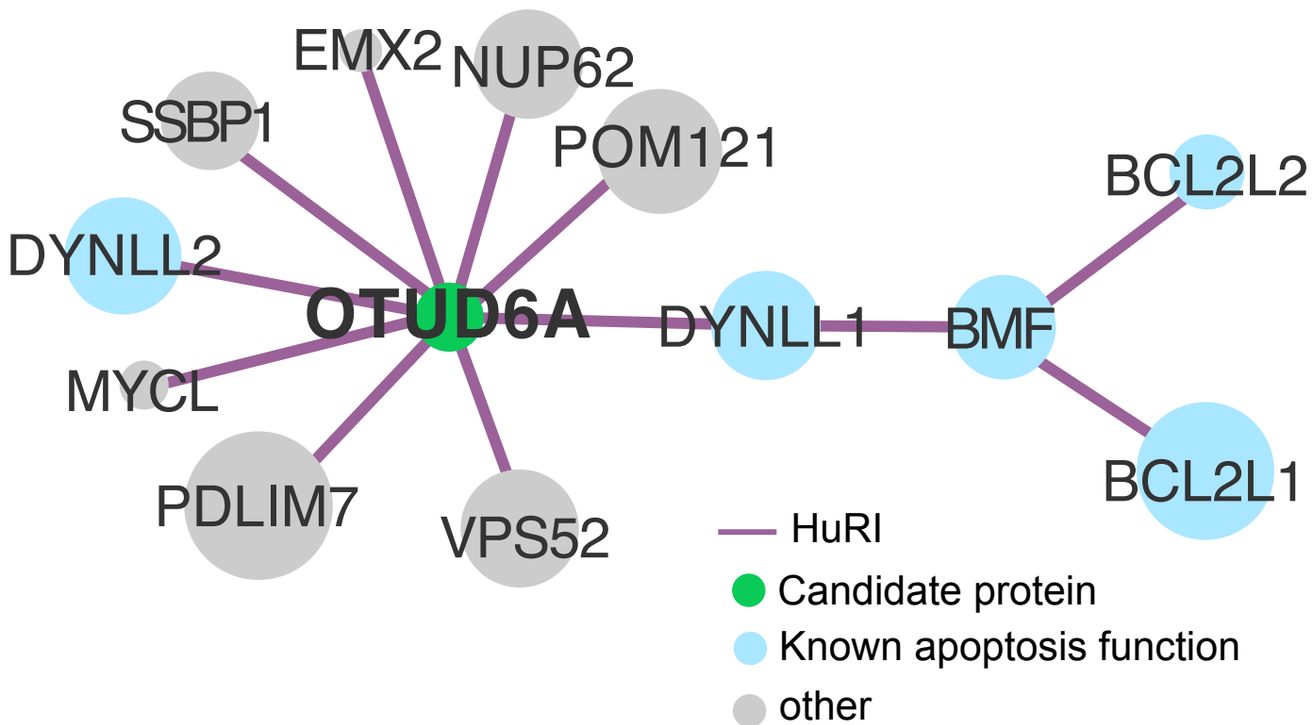
Guilt-by-association



Candidate protein
Known apoptosis function
other



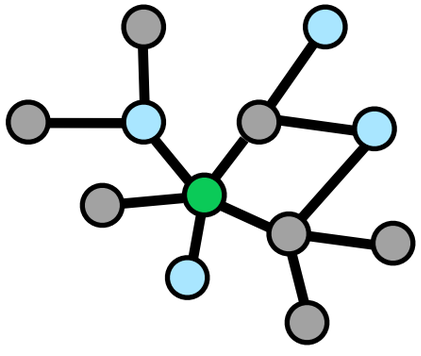
OTU deubiquitinase 6A



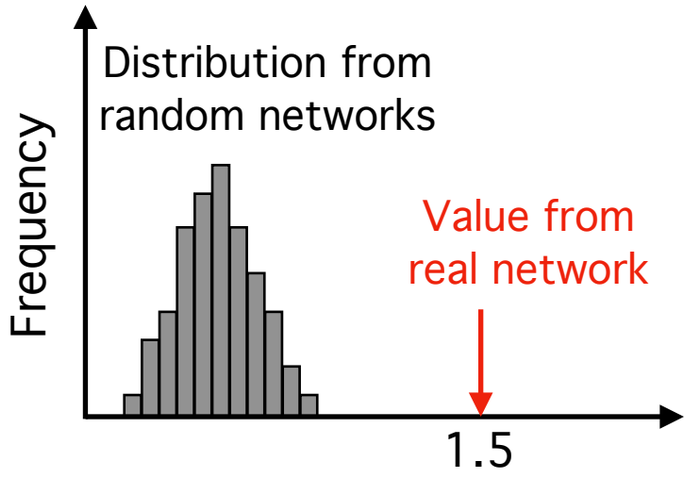
Fuxman Bass et al *Nature Methods* 2013, Luck et al *Nature* 2020

What is the function of my gene of interest?

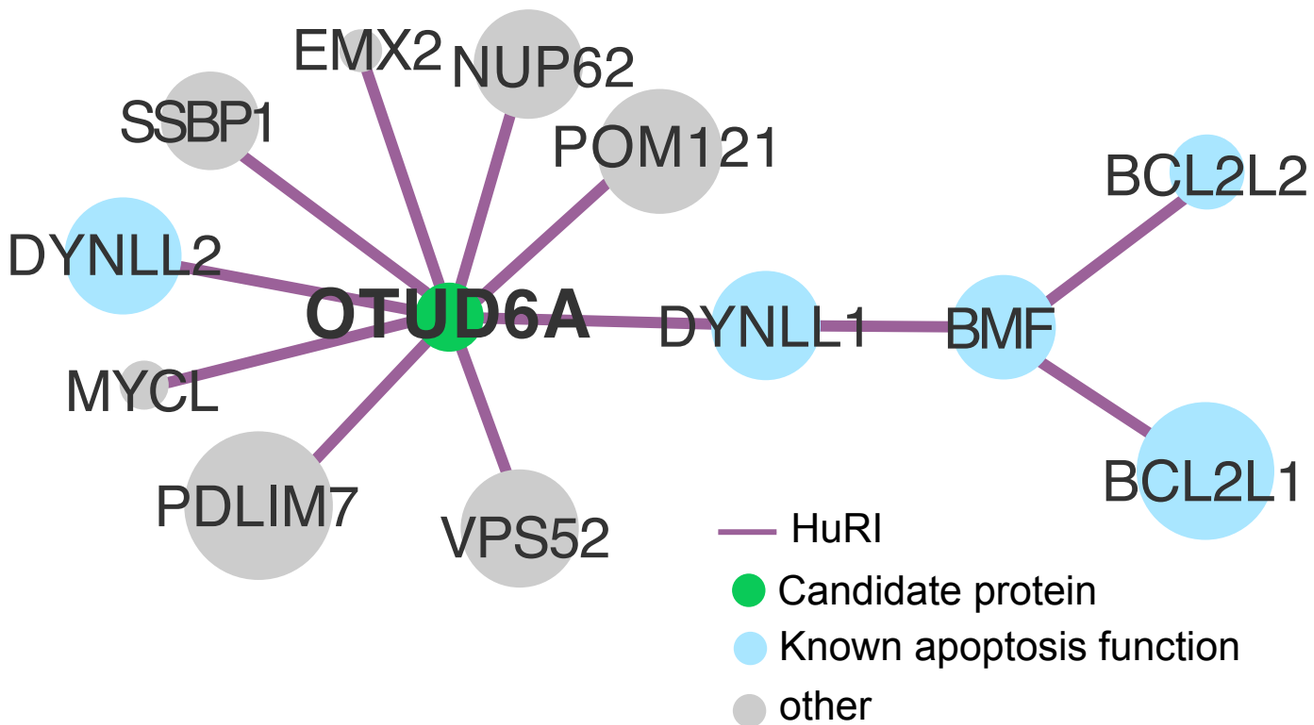
Guilt-by-association



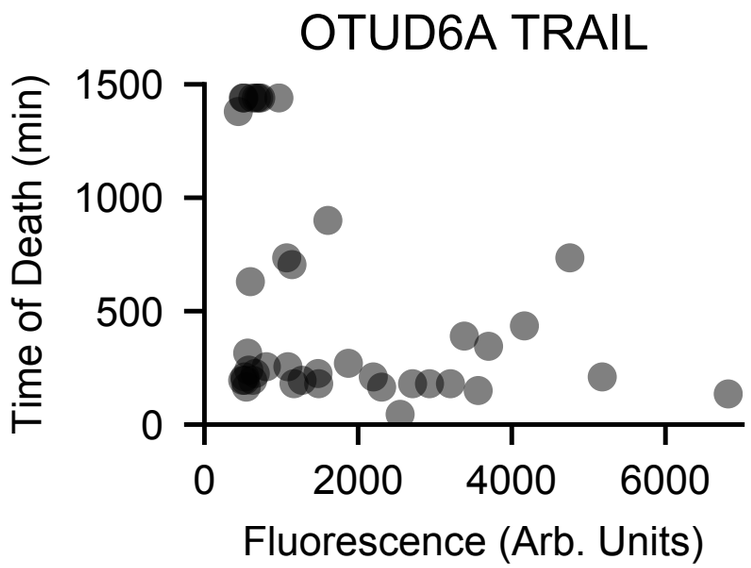
Candidate protein
Known apoptosis function
other



OTU deubiquitinase 6A



OTUD6A expression results in earlier cell death



Fuxman Bass et al *Nature Methods* 2013, Luck et al *Nature* 2020

Summary

- Molecular interaction data can be represented as graphs
- Biological networks are scale-free
- Use degree-controlled randomized networks to look for trends
- Trends in literature-curated networks can be falisified
- Guilt-by-association is a method to predict functions of proteins using interaction data