Network Preliminaries

Héctor Corrada Bravo

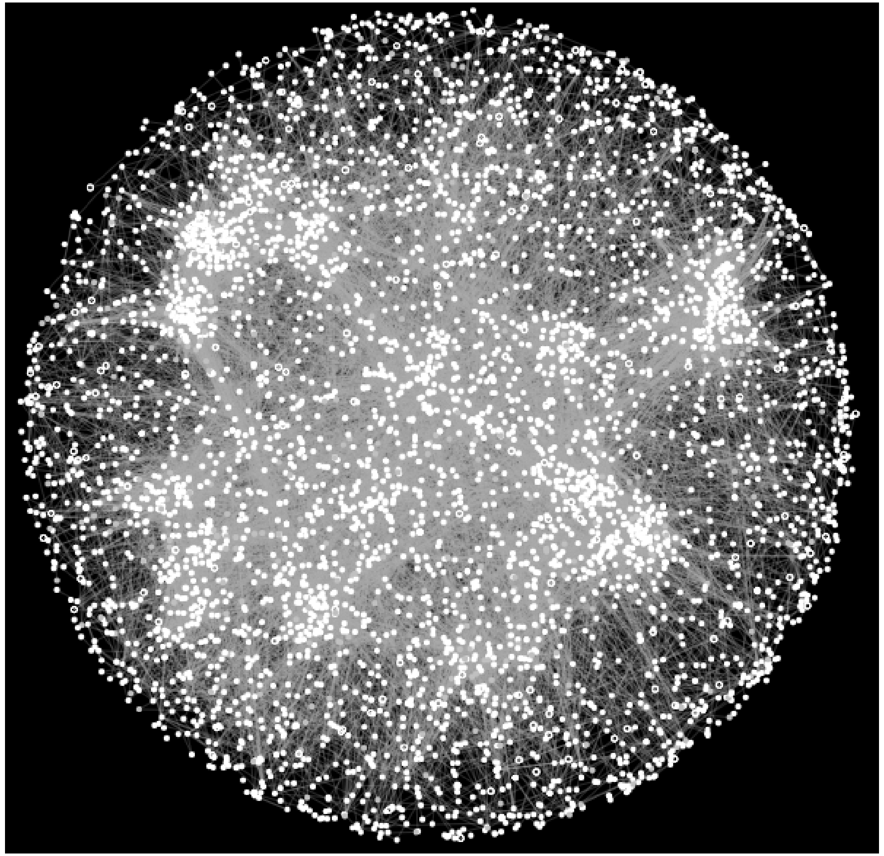
University of Maryland, College Park, USA

CMSC828O 2018-09-01

Genetic Interaction Network

- Yeast high-throuput double-knockdown assay
- ~5000 genes
- ~800k interactions

http://www.geneticinteractions.org/



Costanzo et al. (2016) Science. DOI: 10.1126/science.aaf1420

Genetic Interaction Network

- Yeast high-throuput double-knockdown assay
- ~5000 genes
- ~800k interactions

http://www.geneticinteractions.org/

Costanzo et al. (2016) Science. DOI: 10.1126/science.aaf1420

Genetic Interaction Network

- Number of vertices: 2803
- Number of edges: 67,268

Preliminaries

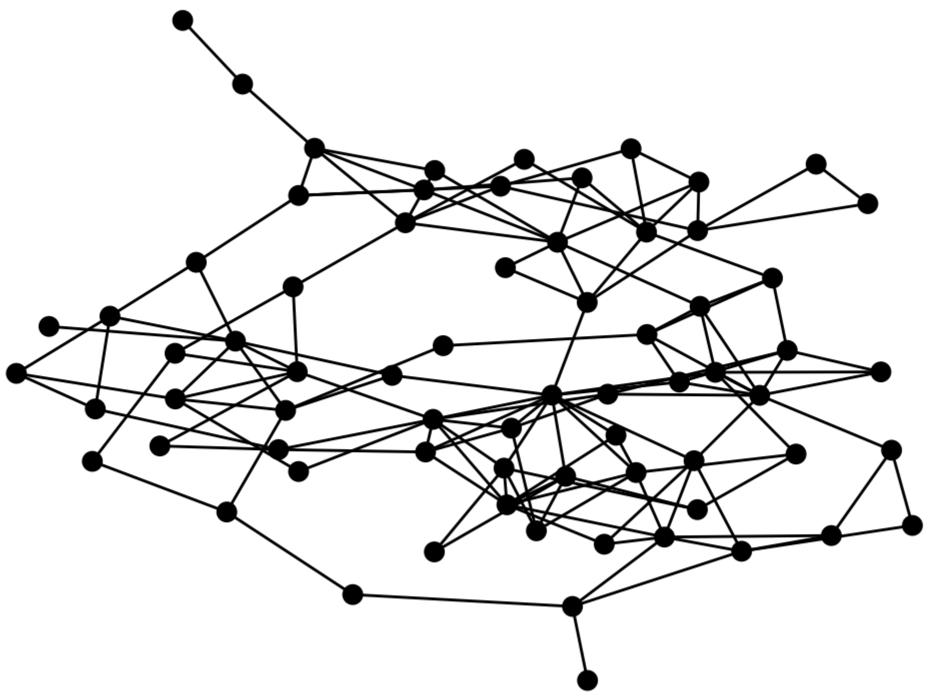
Network: abstraction of *entities* and their interactions

Graph: mathematical representation

vertices: nodes

edges: links

Unirected graph

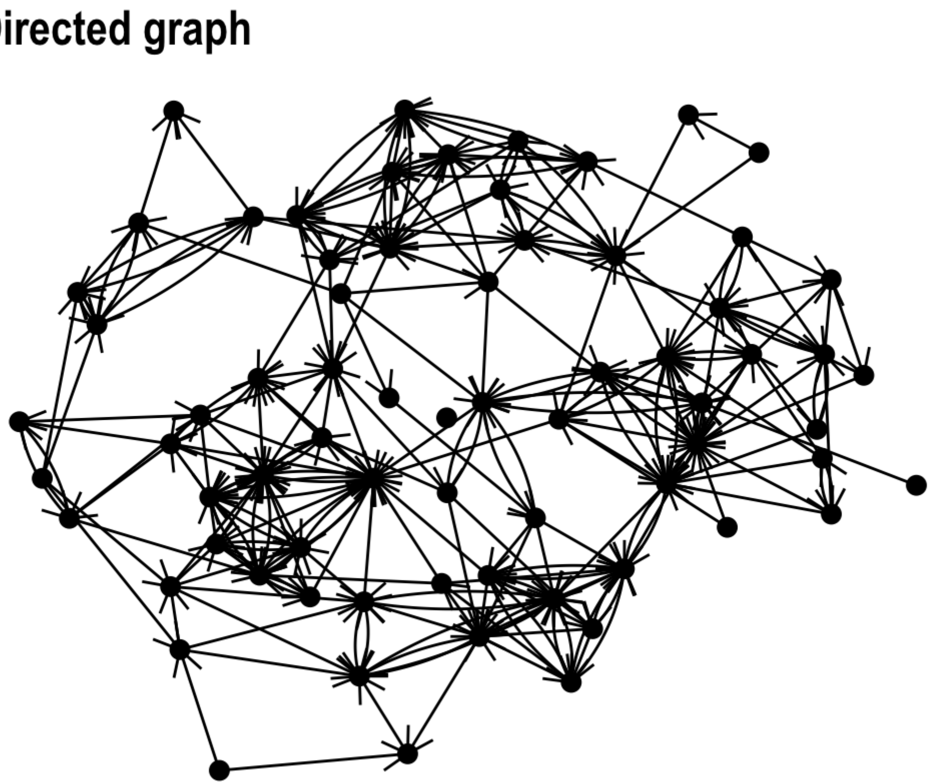


Preliminaries

Network: abstraction of *entities* and their interactions

Graph: mathematical representation

vertices: nodes
edges: links



Network statistics: notation

Number of vertices: n

In our example: *number of genes*

Network statistics: notation

Number of vertices: n

In our example: *number of genes*

Number of edges: m

In our example: *number of genetic interactions*

Network statistics: notation

Number of vertices: n

In our example: *number of genes*

Number of edges: m

In our example: *number of genetic interactions*

Degree of vertex i: k_i

Number of genetic interactions for gene i

Network statistics: notation

On the board:

- Calculate number of edges m using degrees k_i (for both directed and undirected networks)
- Calculate *average degree* c
- Calculate *density* ρ

Network statistics: notation

On the board:

- Calculate number of edges m using degrees k_i (for both directed and undirected networks)
- Calculate *average degree* c
- Calculate *density* ρ

In our example:

Average degree: 47.9971459

Density: 0.0171296

(On the board)

Number of edges using degrees (undirected)

$$m = \frac{1}{2} \sum_{i=1}^{n} k_i$$

Number of edges using degrees (directed)

$$m = \sum_{i=1}^{n} k_i^{\text{in}} = \sum_{i=1}^{n} k_i^{\text{out}}$$

(On the board)

Average degree

$$c = \frac{1}{n} \sum_{i=1}^{n} k_i$$

Density

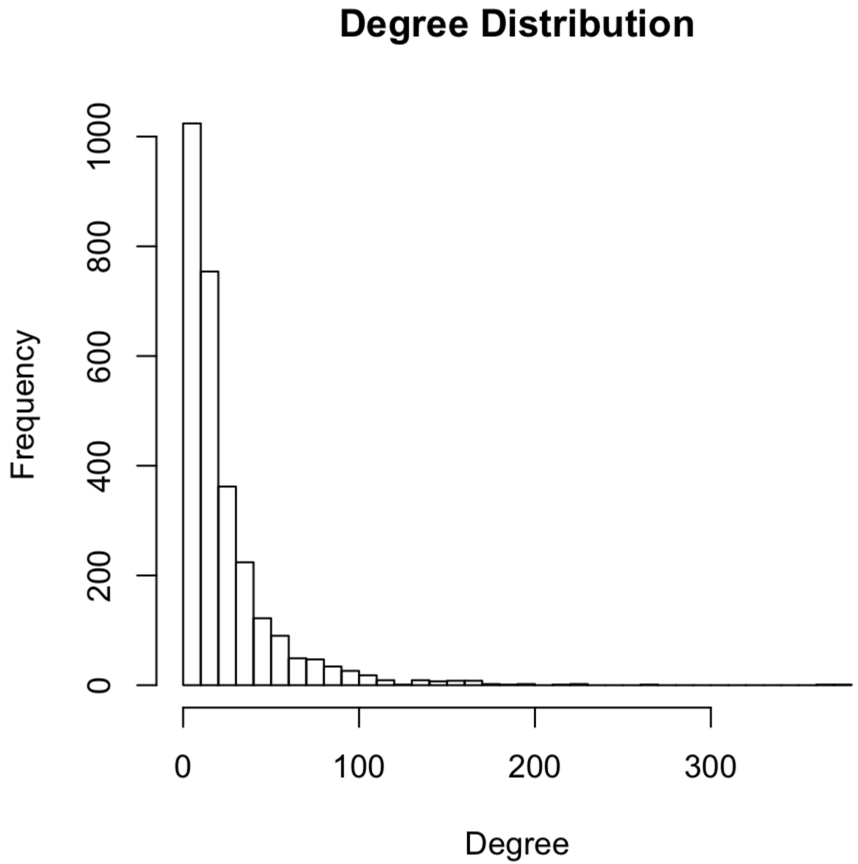
$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1} \approx \frac{c}{n}$$

Degree distribution

Fundamental analytical tool to characterize networks

p_k: probability randomly chosen vertex has degree k

On the board: how to calculate p_k and how to calculate average degree c using degree distribution.



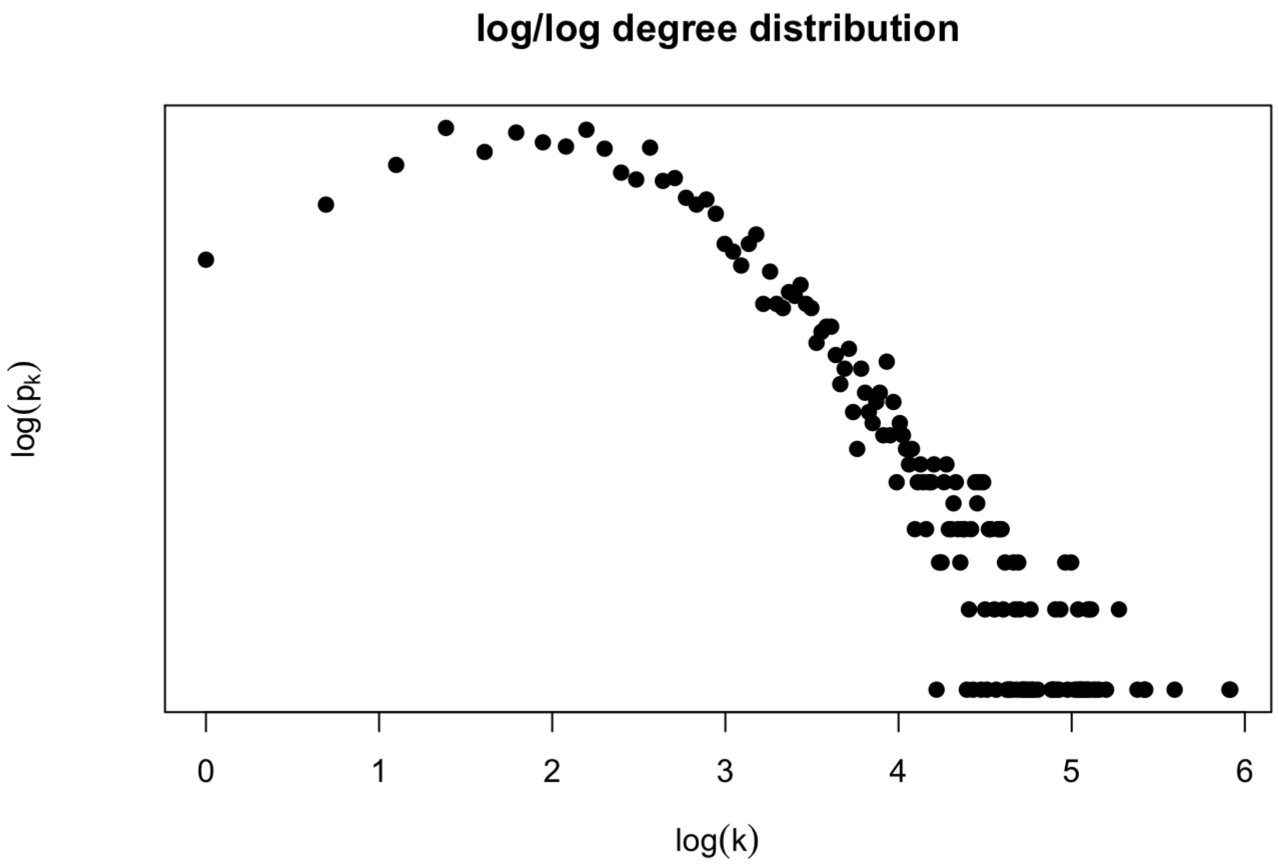
(On the board)

Degree distribution

$$p_k = \frac{n_k}{n}$$

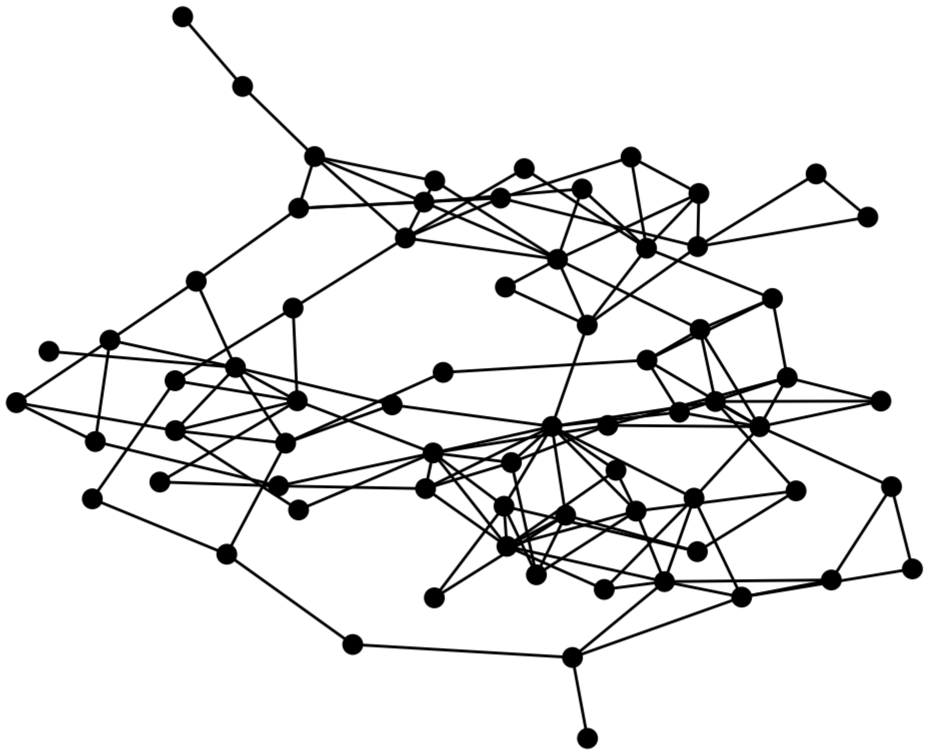
n_k: number of nodes in graph with degree k

Degree Distribution



Paths and Distances

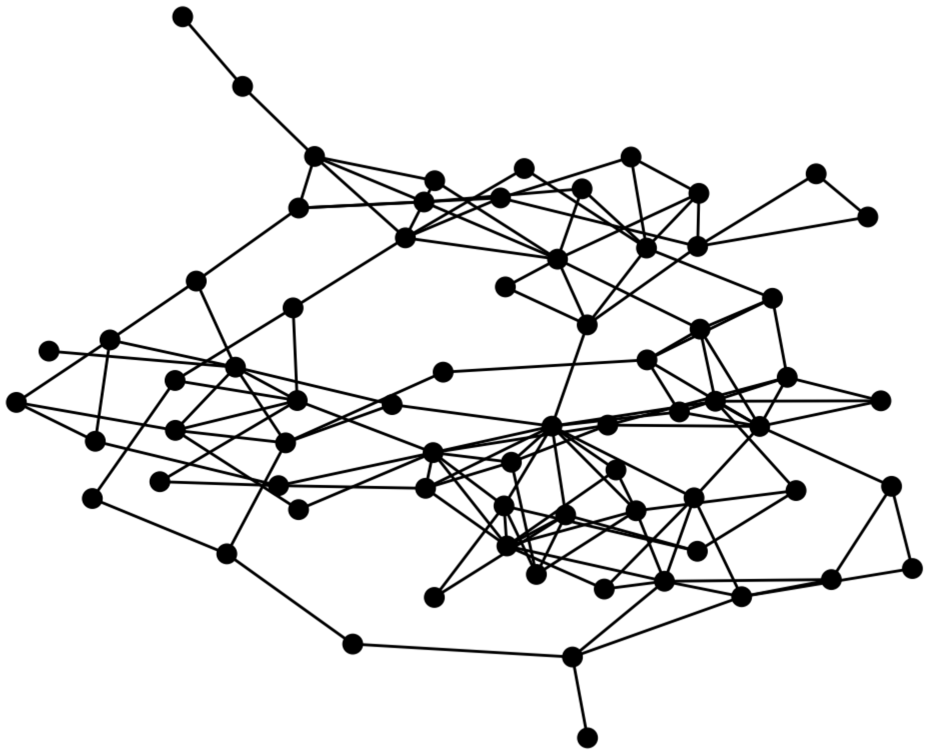
Distance d_{ij}: length of **shortest** path betwen vertices i and j.



Paths and Distances

Distance d_{ij}: length of **shortest** path betwen vertices i and j.

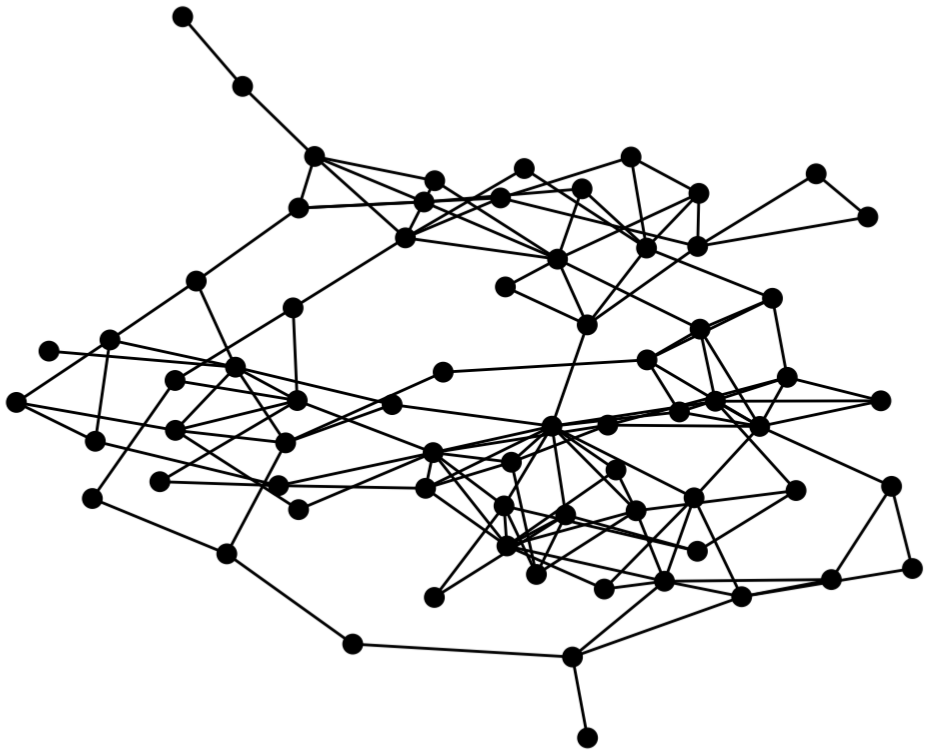
Diameter: longest shortest path $\max_{ij} d_{ij}$



Paths and Distances

Distance d_{ij}: length of **shortest** path betwen vertices i and j.

On the board: average path length

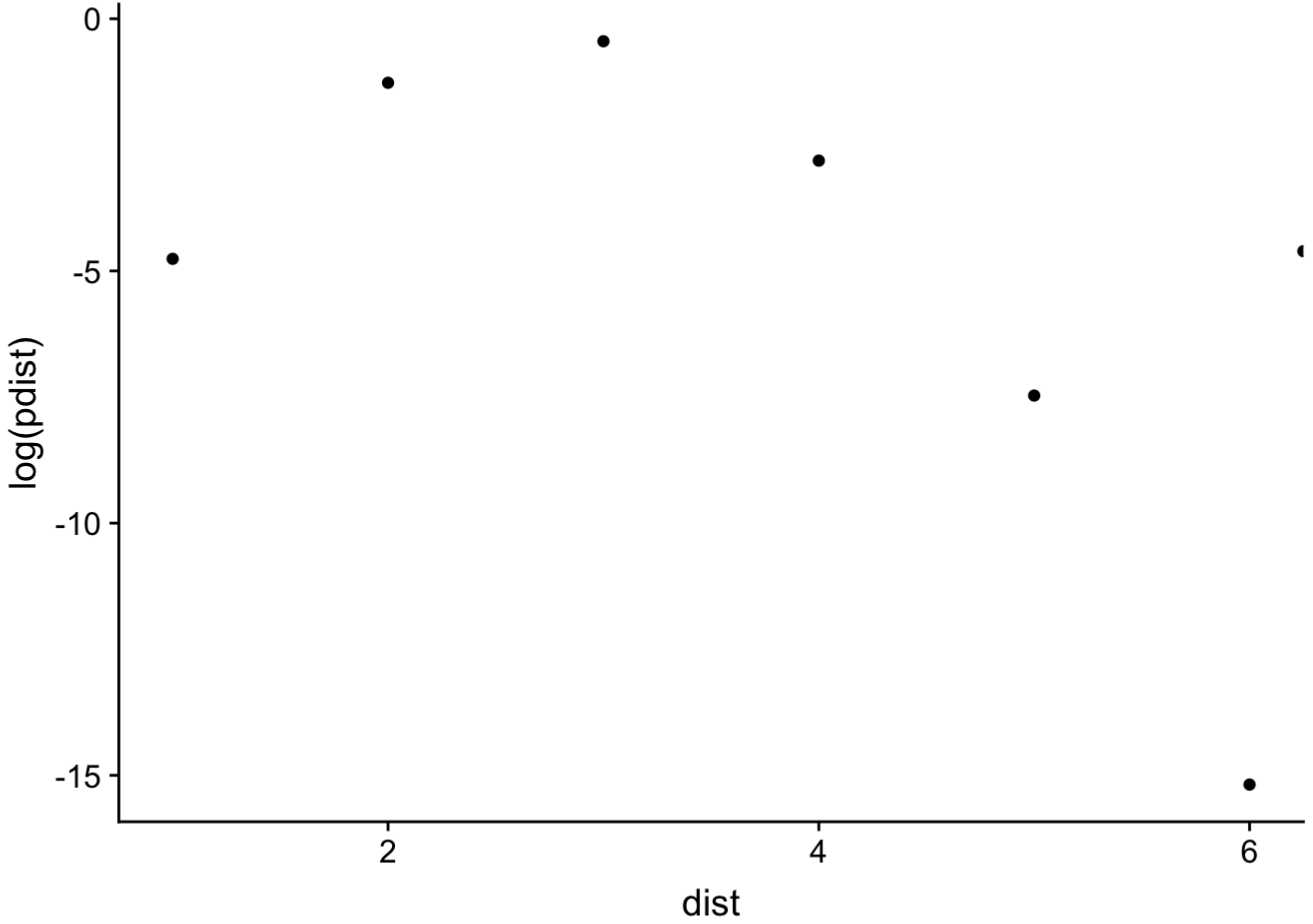


(On the board)

Average path length

$$\bar{d} = \frac{1}{n(n-1)} \sum_{i,j;i \neq j} d_{ij}$$

Distance Distribution



Distances and paths

By convention: if there is no path between vertices i and j then $d_{ij} = \infty$

Distances and paths

By convention: if there is no path between vertices i and j then $d_{ij} = \infty$

Vertices i and j are *connected* if $d_{ij} < \infty$

Distances and paths

By convention: if there is no path between vertices i and j then $d_{ij} = \infty$

Vertices i and j are *connected* if $d_{ij} < \infty$

Graph is connected if $d_{ij} < \infty$ for all i, j

Distances and paths

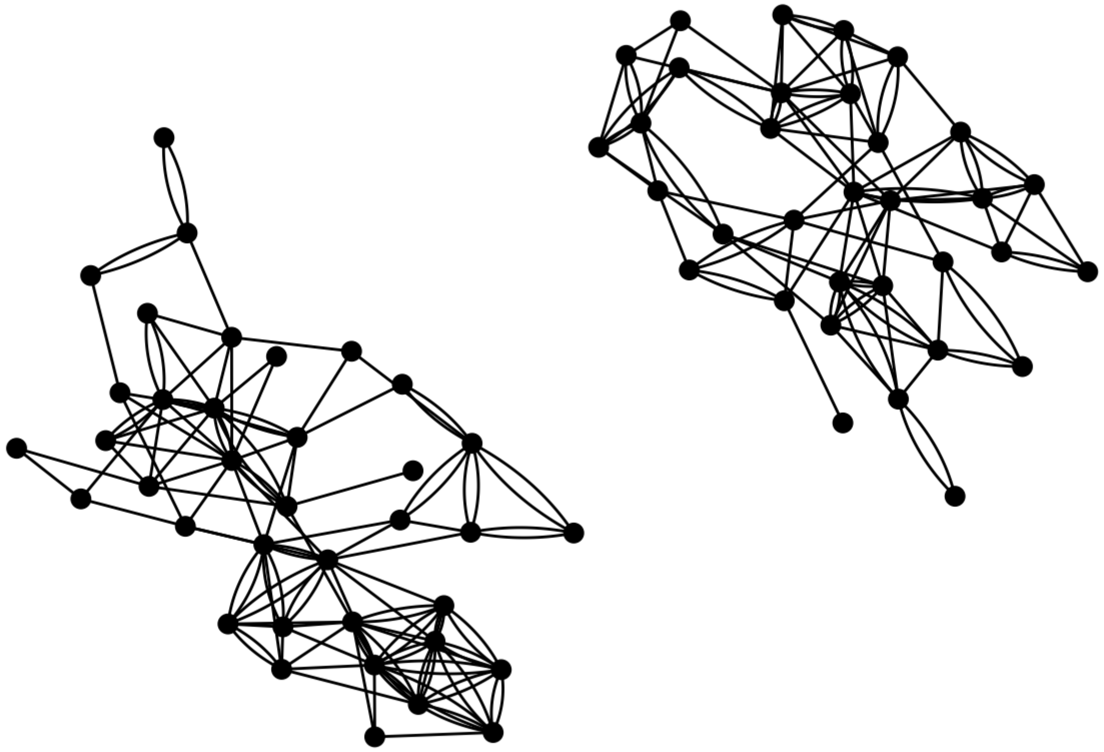
By convention: if there is no path between vertices i and j then $d_{ij} = \infty$

Vertices i and j are *connected* if $d_{ij} < \infty$

Graph is connected if $d_{ij} < \infty$ for all i, j

Components maximal subset of connected components

Components



Clustering Coefficient

One last quantity of interest: how dense is the neighborhood around vertex i?

Do the genes that interact with me also interact with each other?

Definition on the board

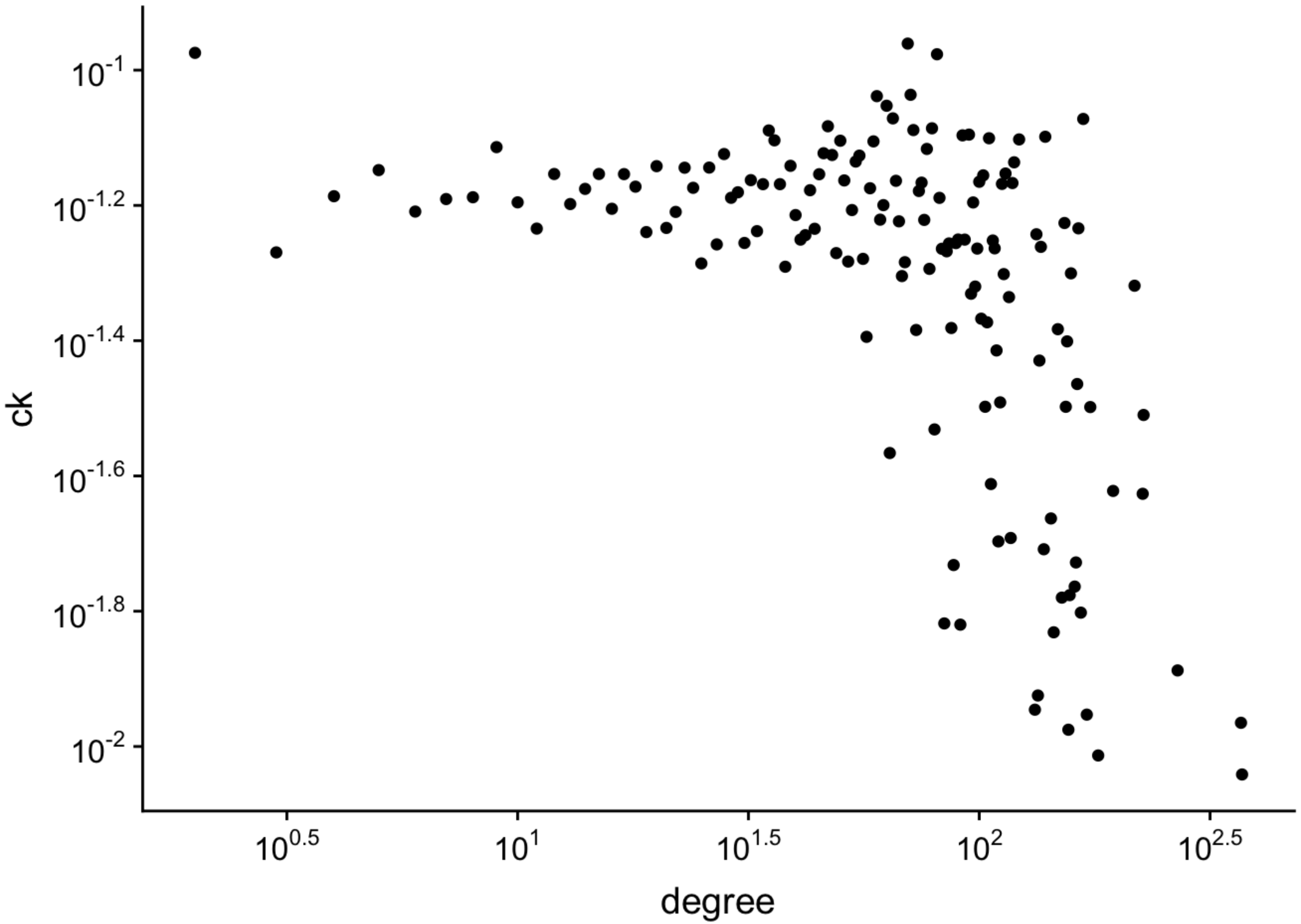
(On the board)

Clustering coefficient

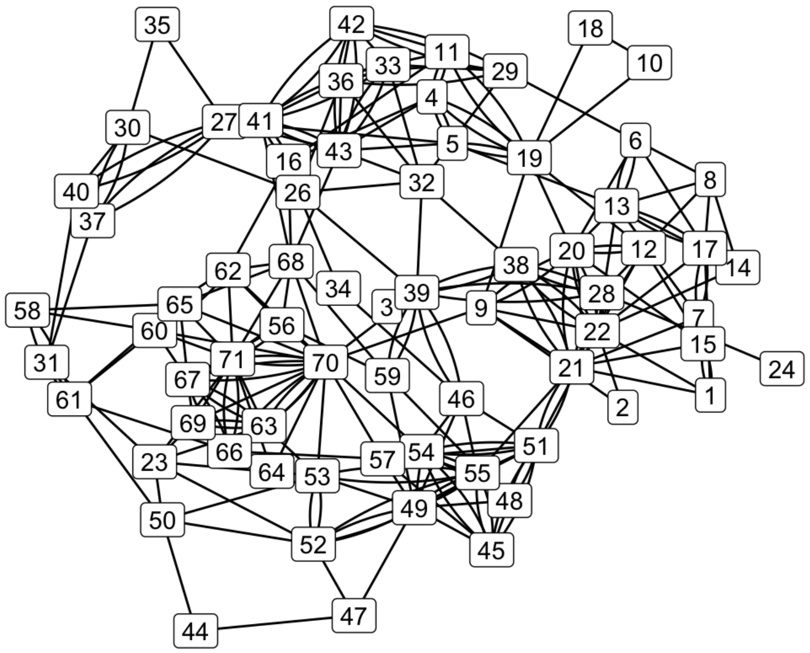
$$c_i = \frac{2m_i}{k_i(k_i - 1)}$$

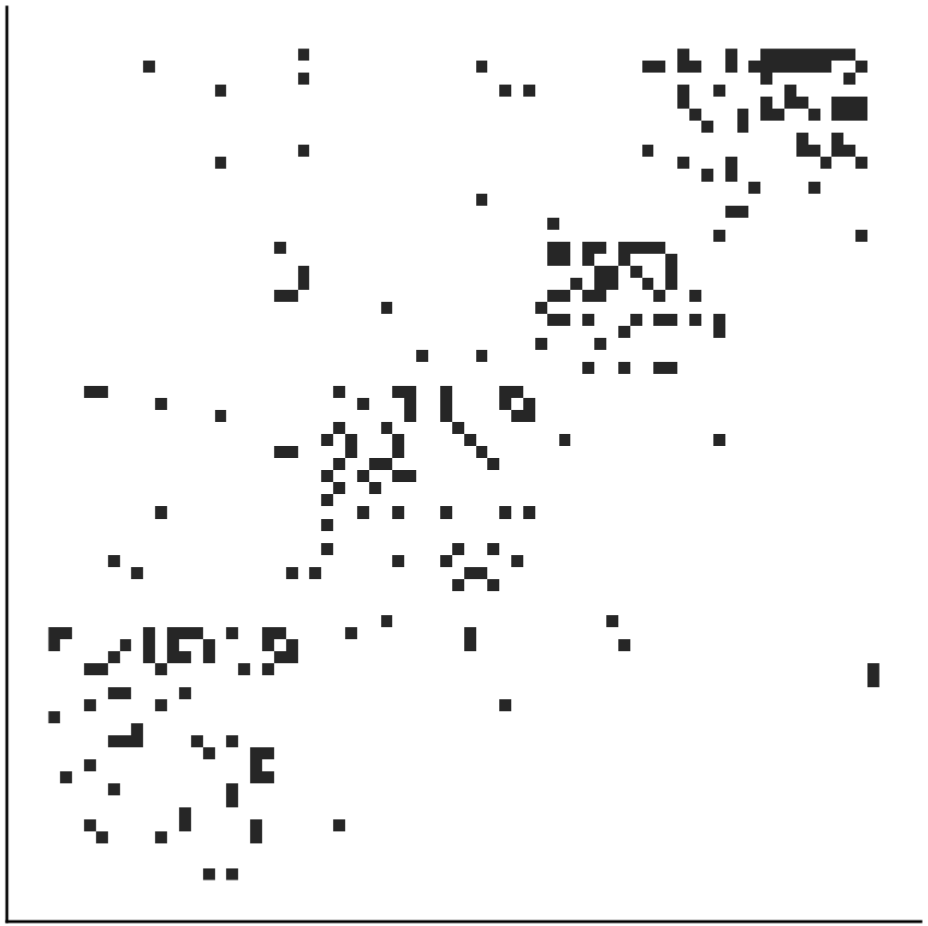
m_i: number of edges between neighbors of vertex i

Clustering coefficient



Adjacency Matrix





Adjacency Matrix

On the board:

- Definition
- Computing degree with adj. matrix
- Computing num. edges m with adj. matrix
- Computing paths with adj. matrix