Random Graphs

858L

Descriptive Studies of Biological Graphs

- What to networks that arise in nature "look like"? What are their characteristics?

- Related to random graphs because you need to know what features can arise just by chance.

- Need to describe their properties so that:
  - we can postulate how the graphs evolved.
  - understand trends in how they are put together, and perhaps how they perform their functions.
  - to create null models to distinguish between functional, conserved, important features, and those that are merely there by chance.

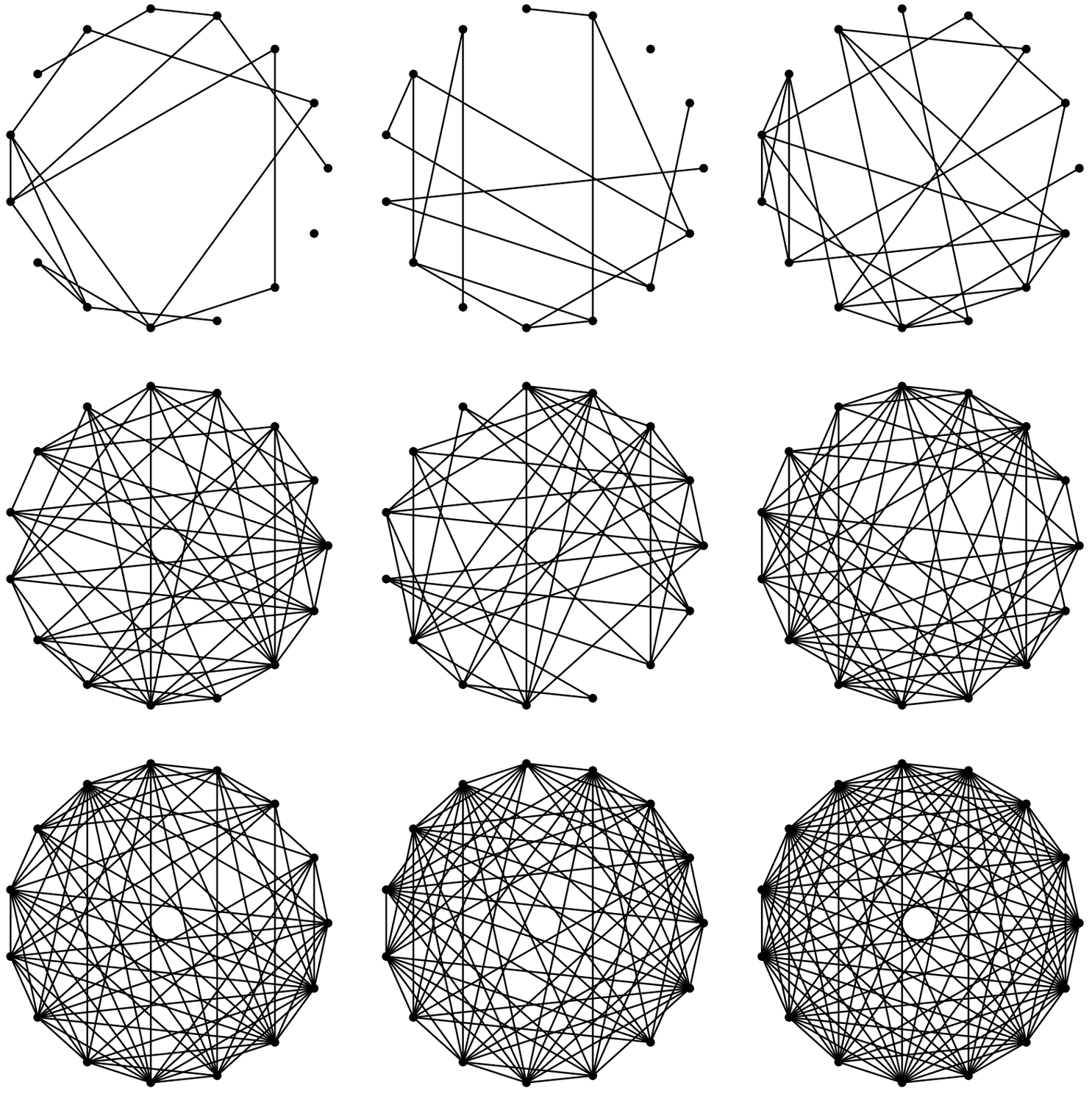
Random Models

Erdos-Renyi (1960):

Create n vertices

Between every pair of vertices {u,v}, add an edge with probability p.

Expected degree is $(n\text{-}1)p$



ShowGraphArray[Partition[Table[RandomGraph[15, p], {p, 0.1, 0.9, 0.1}], 3]]

Preferential Attachment

At time step t:

- Node u_t is added.
- Edge $\{v, u_t\}$ added independently with probability:

$$\frac{k_v}{\sum_w k_w}$$

Proposed in the context of the WWW: popular pages become more popular.

Expect a scale-free degree distribution.

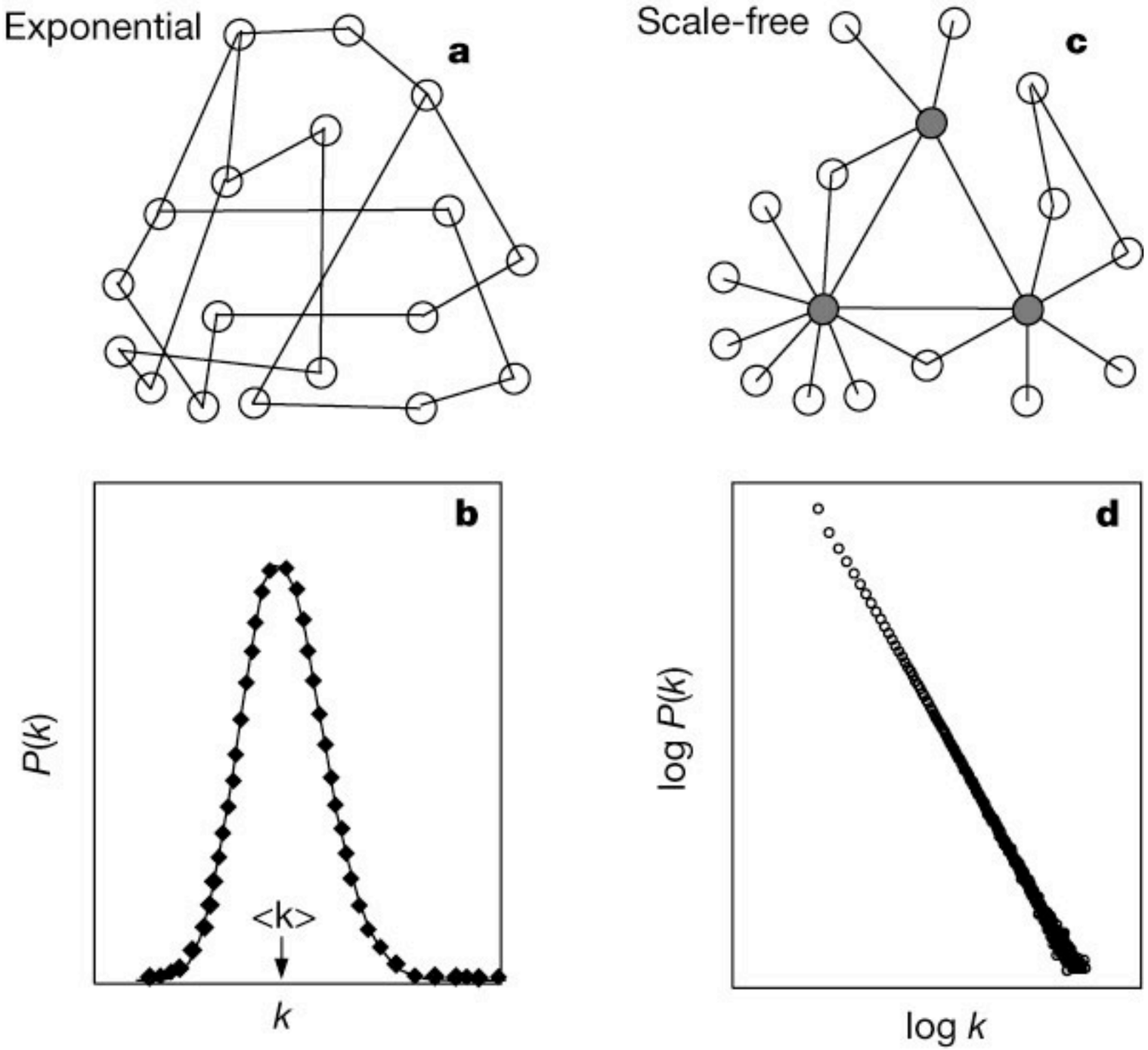
(Albert et al, Nature, 2000)

Degree Distribution

ER has an "average" node.

SF networks don't have a strong typical node

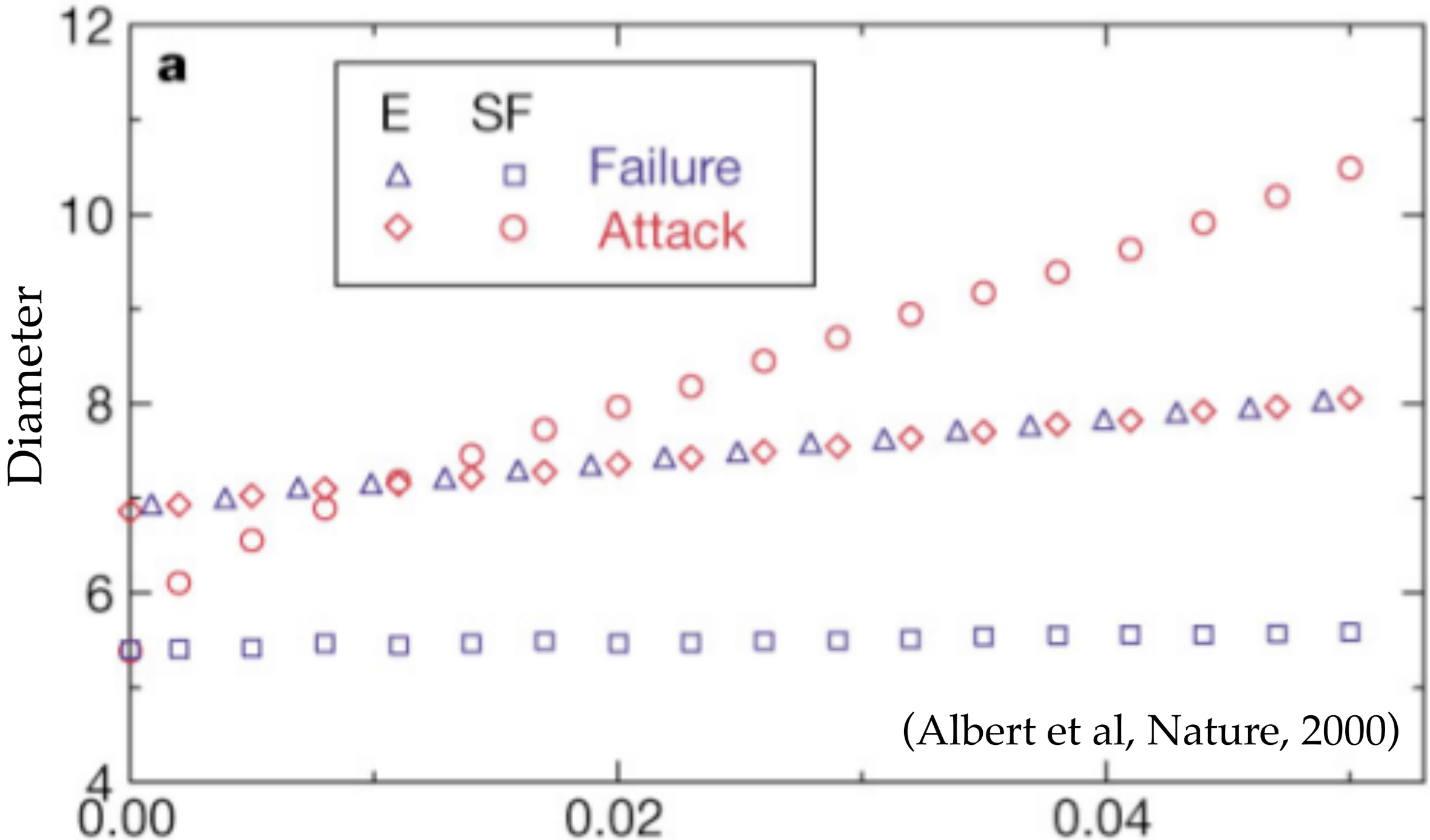
They have a longer tail: more nodes of higher degree.



(Jeong et al., *Nature*, 2000)

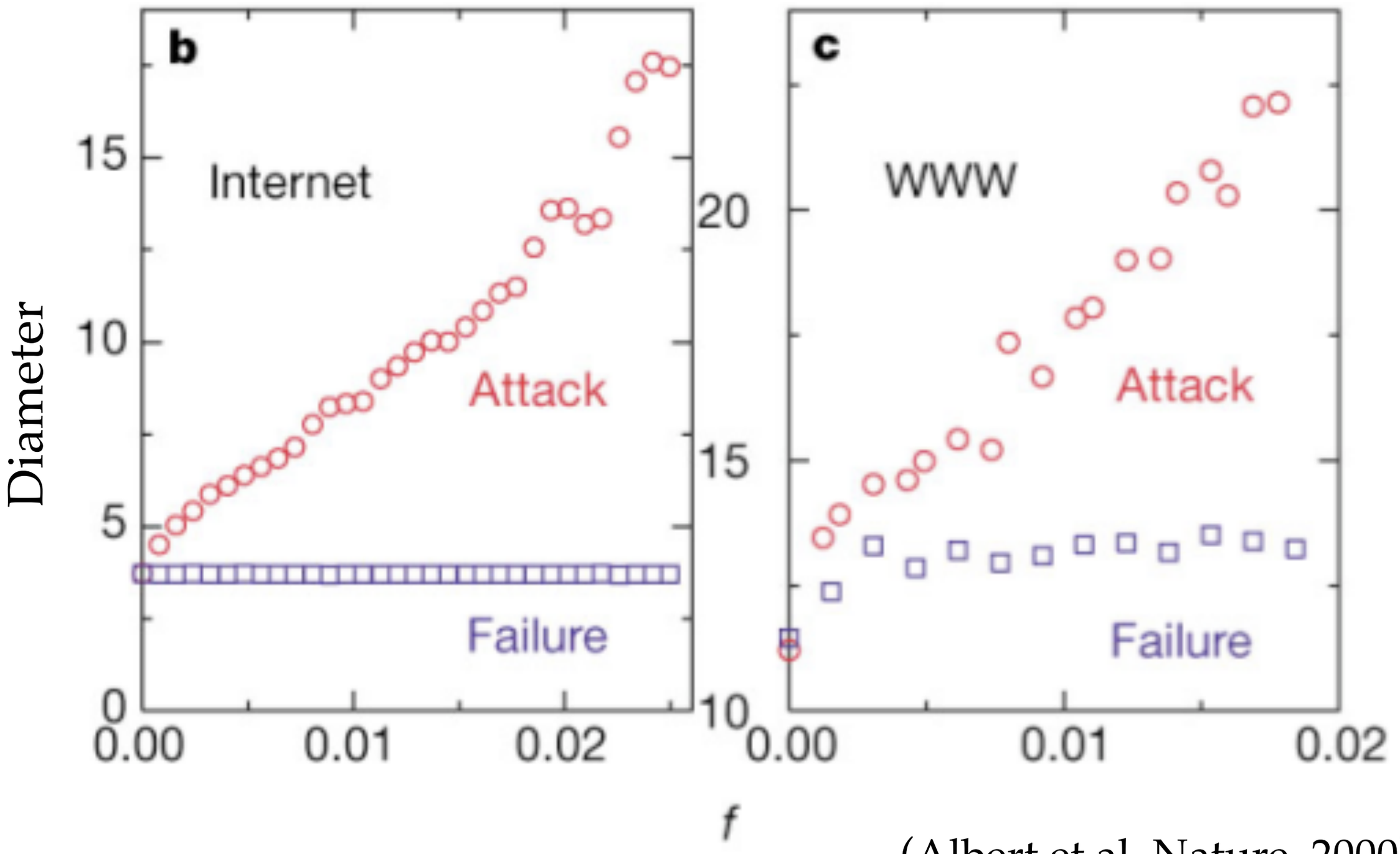
Attack vs. Failure: Random Graph

10,000 nodes; 20,000 edges; average degree 4



(Albert et al, Nature, 2000)

Attack vs. Failure on Real Graphs:

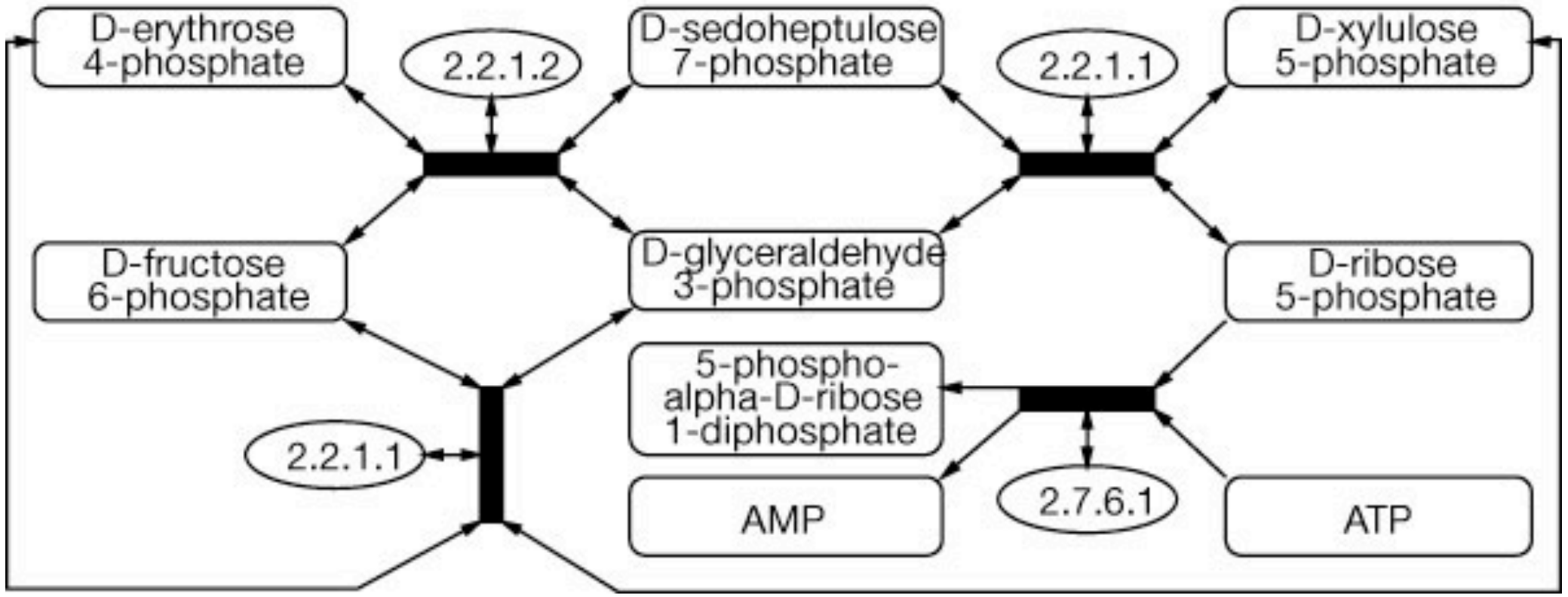


(Albert et al, Nature, 2000)

The large-scale organization of metabolic networks

Jeong et al., *Nature*, 2000.

Metabolic Network



• Nodes are *substrates* (small molecules, e.g.)

• Reactions are shown as black boxes. Reactions have input substrates and output substrates.

• Can think of reactions as "directed hyperedges".

• Reactions labeled with enzyme that catalyzes the reaction.

• Actually, the analysis won't really use the structure of the graph.

Metabolic Data Sets

43 organisms

Reactions from literature

Which organisms have particular reactions predicted from the genes they have (WIT database)

Expect that metabolic networks are not at all like really “random” networks.

Degree Distribution

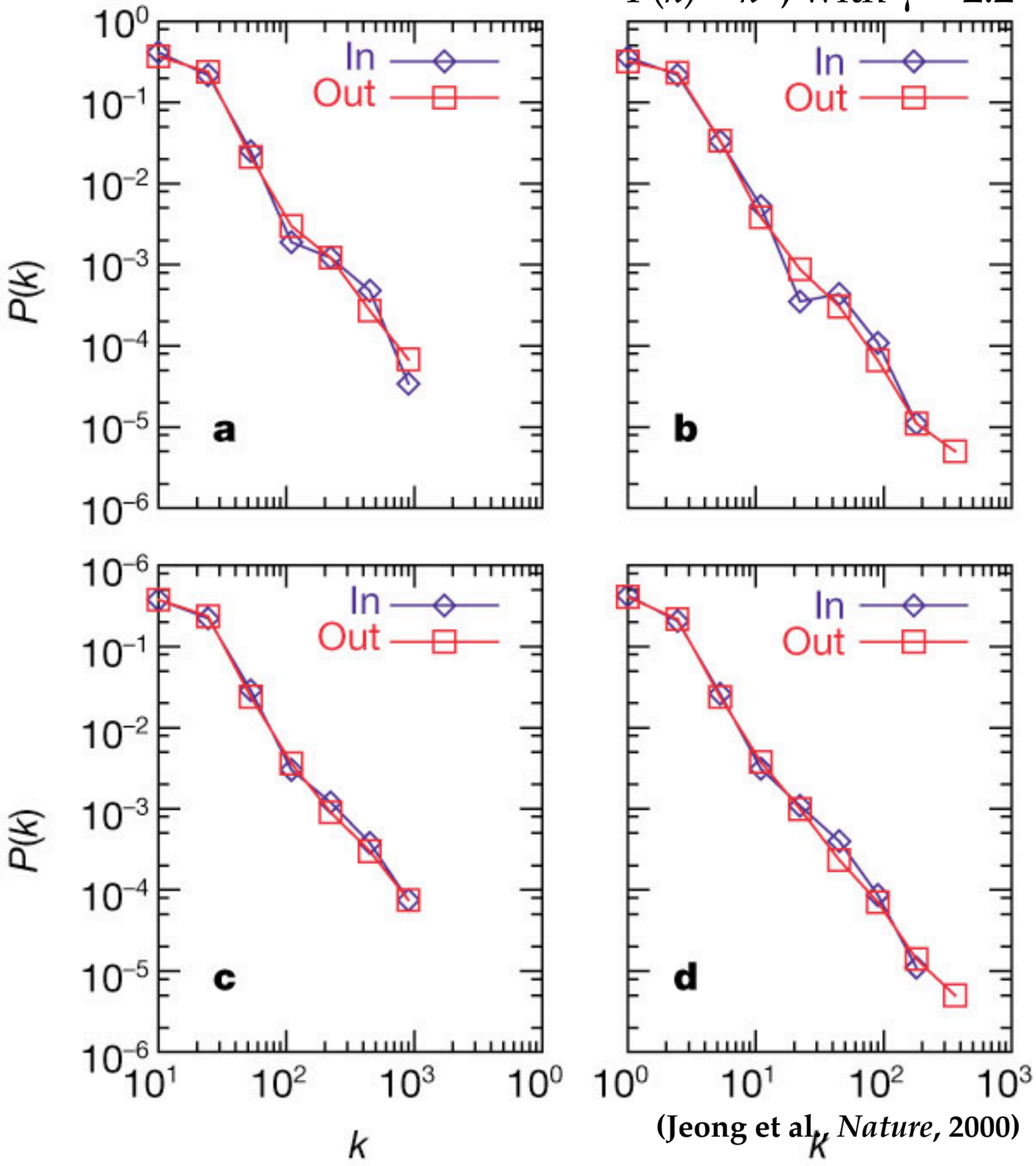
$P(k)$ is the fraction of nodes with in- or out-degree k.

(a) Archae
(b) E.coli
(c) Worm (c. elegans)
(d) Average

Clearly more similar to scale-free distribution & not like an ER random graph.

True for all 43 organisms

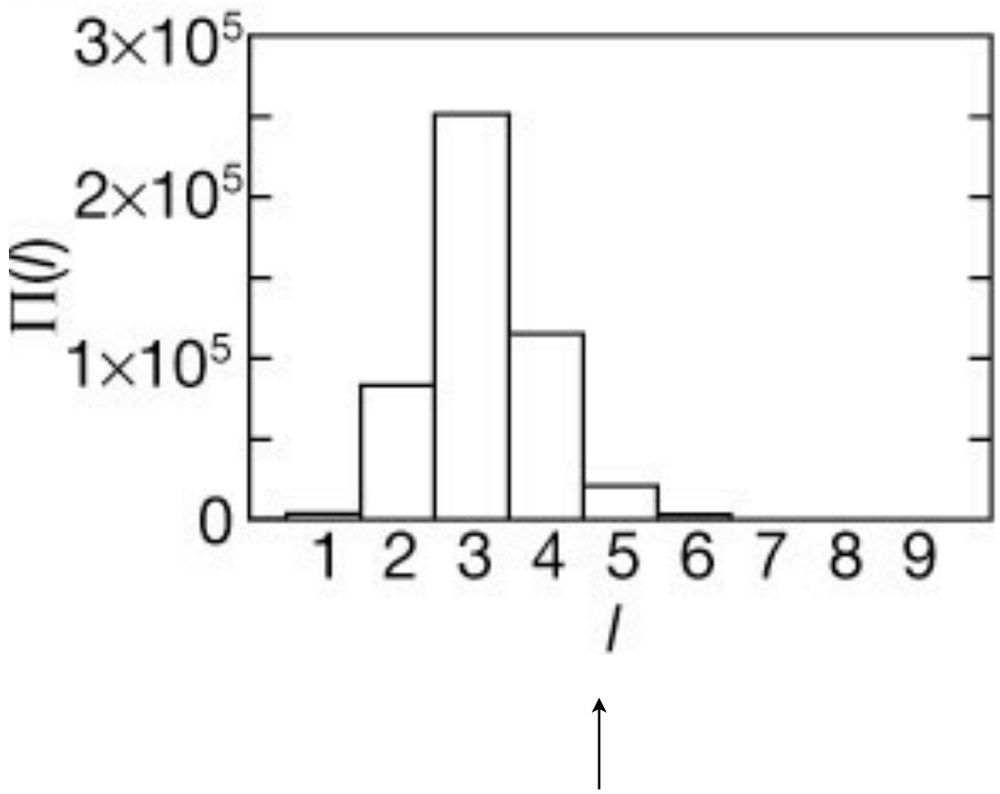
$P(k) \sim k^{-\gamma}$, with $\gamma = 2.2$



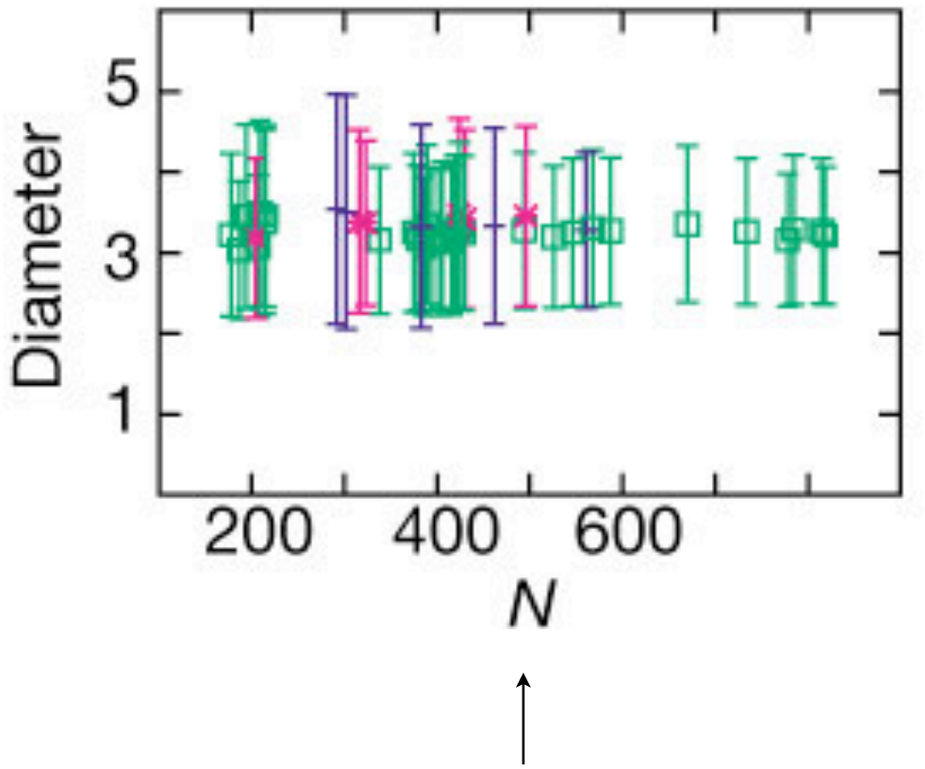
(Jeong et al., *Nature*, 2000)

Network Diameter

(Jeong et al., *Nature*, 2000)



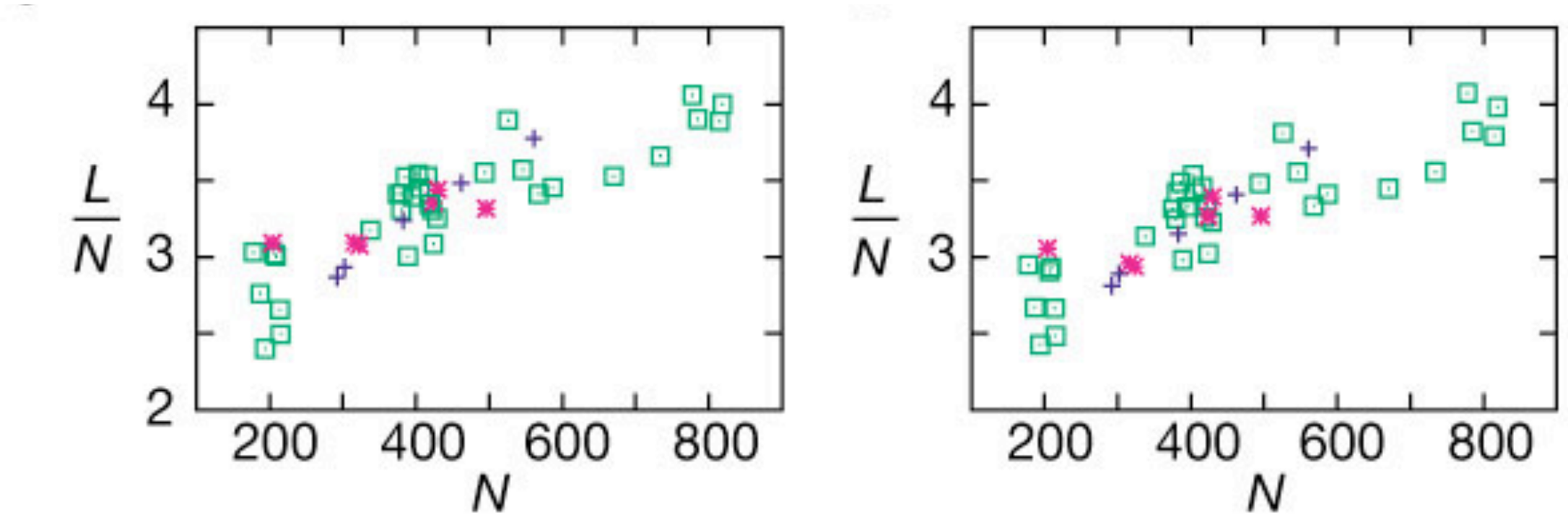
Histogram of path lengths between substrates for E. coli



Average path length (diameter) for each of the 43 organisms.

- Surprising that diameter is the same no matter how many substrates.
- More connections in complex organisms
- Perhaps advantageous to maintain a small diameter so that changes in one substrate can propagate quickly.

Indeed, higher average degrees for organisms with more substrates:



(Jeong et al., *Nature*, 2000)

Response to Mutations

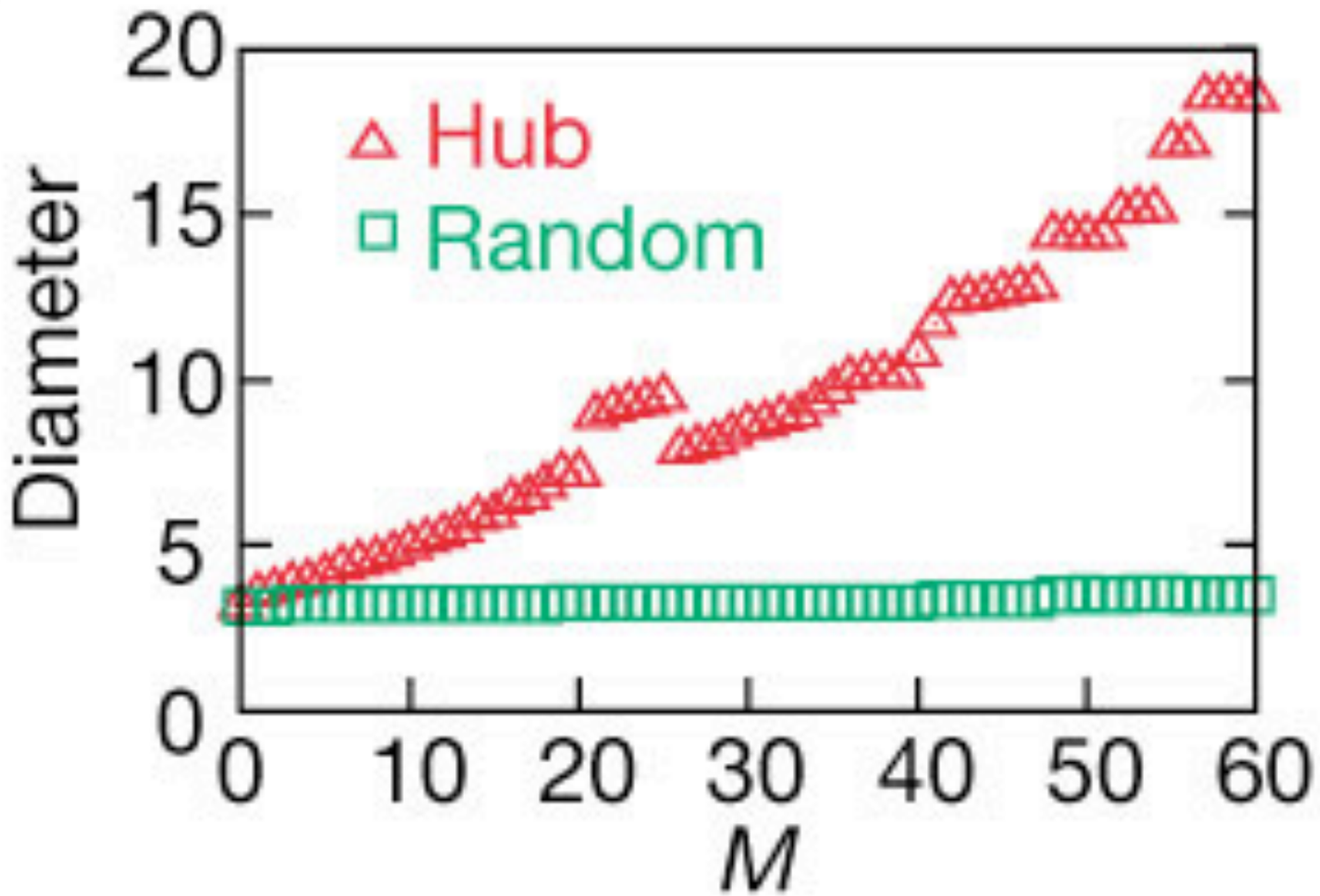
Scale-free networks are easy to destroy: knock out the hubs (high-degree nodes)

But they are very robust to random knock-outs

Random knockouts simulate random mutations.

Except the simulation knocks out substrates, and typically it's enzymes (edges) that are mutated.

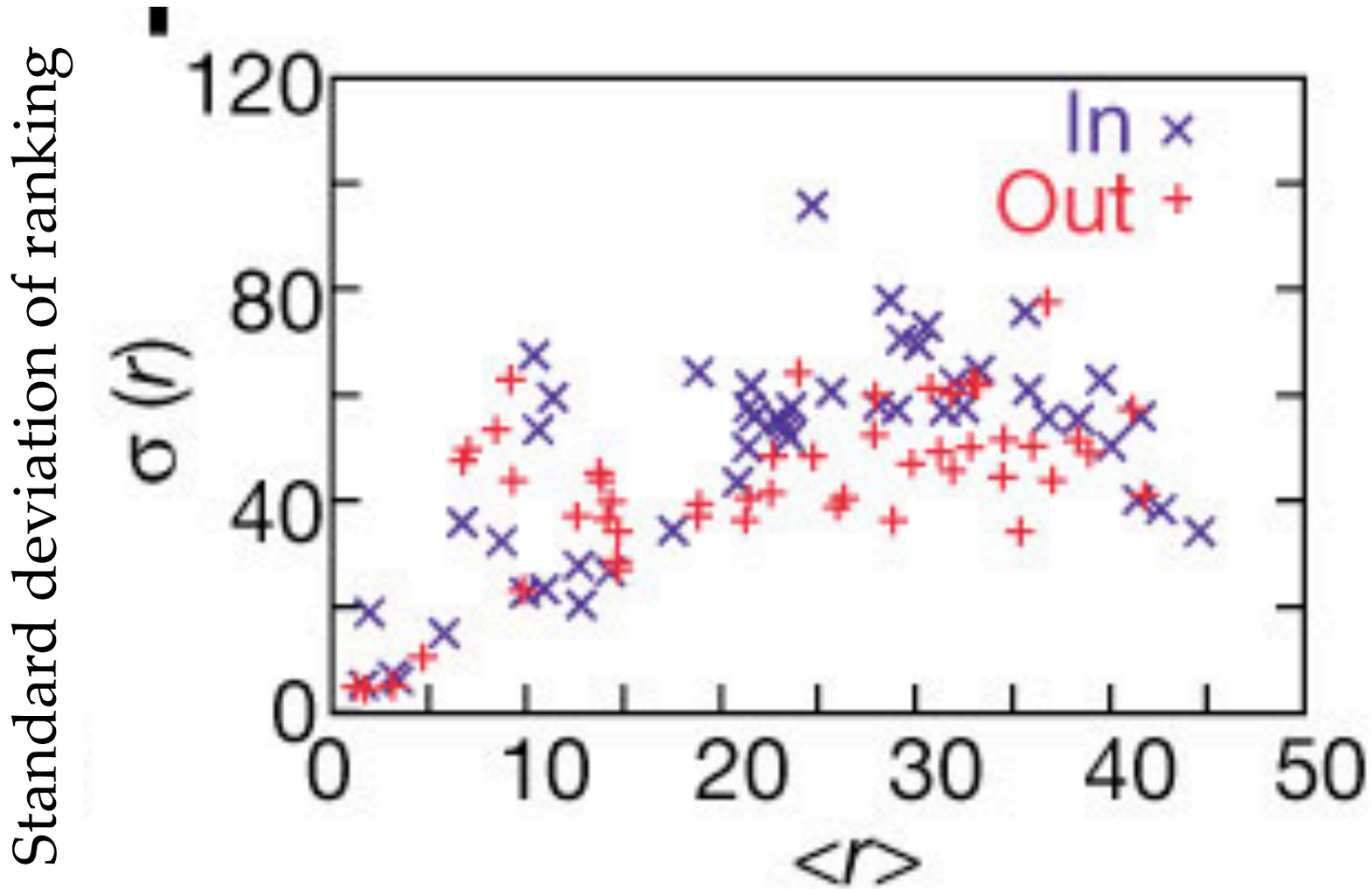
Simulation on E coli.



(Jeong et al., *Nature*, 2000)

Are The Same Substrates Always Hubs?

Variance increases as ranking increases: means that hubs are usually hubs, and non-hubs sometimes are more or less hubby.



Ties?

Average ranking (based on in- or out-degree)

(Jeong et al., *Nature*, 2000)