Network Biology:
Understanding the cell's functional organization

Albert-László Barabási

Zoltán N. Oltvai

Outline:

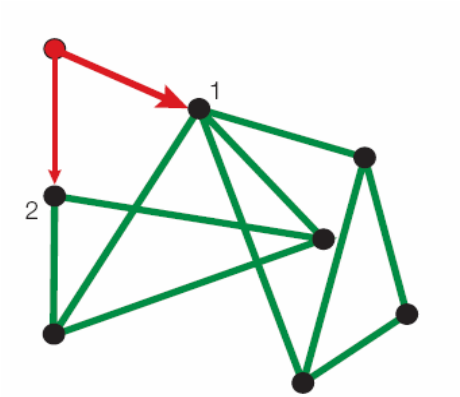
- Evolutionary origin of scale-free networks
- Motifs, modules and hierarchical networks
- Network robustness
- Beyond topology: characterizing the links
- Future directions in network biology

Evolutionary origin of scale-free networks

- Two fundamental processes have a key role in the development of real networks
  - Growth process
  - Preferential attachment
- Both are jointly responsible for the emergence of the scale-free property in complex networks

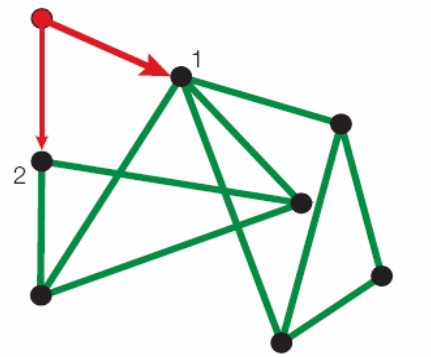
Growth Process

- The network emerges through the subsequent addition of new nodes
  - World Wide Web



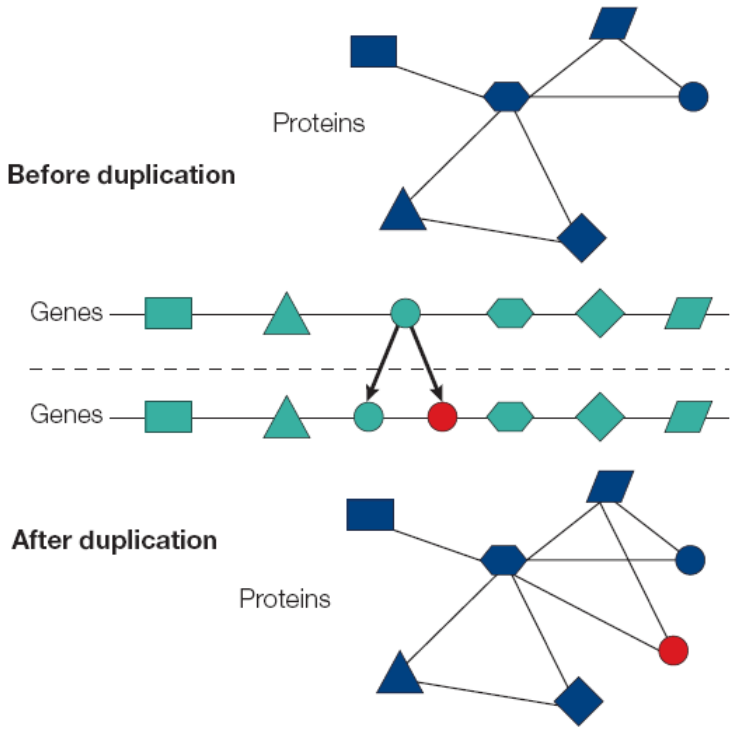
Preferential attachment (rich gets richer)

- Nodes prefer to connect to more connected nodes.
- If a node has many links, new nodes will tend to connect to it with a higher probability.
- This node will therefore gain new links at a higher rate than its less connected peers and will turn into a **hub.**



Gene duplication

- Common origin of growth and preferential attachment in protein networks.



Evidence of preferential attachment

- Highly connected proteins have a natural advantage:
  - More likely to have a link to a duplicated protein.
  - More likely to gain new links if a randomly selected protein is duplicated.
- Origin of the scale free topology traces back to gene duplication.

Evidence of network growth

- The nodes that appeared early in the history of the network are the most connected ones.
  - Among the most connected substrates of the metabolic networks:
    - Coenzyme A, NAD, GTP
- Cross-genome comparisons:
  - the evolutionarily older proteins have more links to other proteins than their younger counterparts.

Modules

- Modularity: group of physically or functionally linked molecules (nodes) that work together to achieve a distinct function:
  - protein-protein & protein-RNA complexes.
  - Temporally coregulated groups of molecules:
    - governing cell cycle.
    - Signal amplification in a signaling pathway.

Identifying topological & functional modules

- Hierarchical modularity:
  - Metabolic
  - protein-protein interaction
  - Regulatory networks
- Can the modules that are present in a cellular network be determined in an automated and objective fashion?
  - Clustering methods

Module Identification Methods:

- Using the network's topological description.
  - Identification of functional modules from the genomic association.
- Combining the topology with integrated functional genomic data

Hierarchical modularity

- Modules do not have a characteristic size
- The network is as likely to be partitioned into a set of clusters of 10-20 components as into fewer (larger) modules.
- Key issue in network biology:
  - Identification of groups of modules of various sizes that together carry out a specific cellular function.

Network robustness

- The system's ability to respond to changes in the external conditions or internal organization while maintaining relatively normal behavior.
  - ***Topological robustness.***
  - ***Functional & dynamical robustness.***

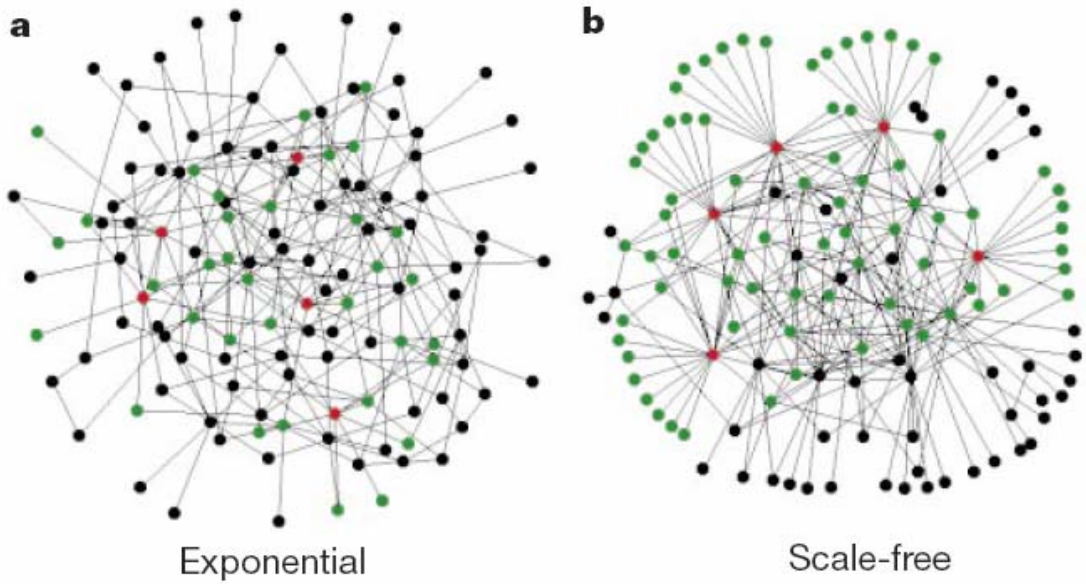
Topological robustness

- disabling a substantial number of nodes will result in an inevitable functional disintegration of a network.

- This is certainly true for a random network:
  - if a critical fraction of nodes is removed, the network breaks down into tiny, non-communicating islands of nodes.

Topological robustness

- Topology has an important role in generating topological robustness.



Albert et al. Error and attack tolerance of complex networks. Nature **406,** 378-382(2000).

Topological robustness

- Scale-free networks are amazingly robust against accidental failures:
  - even if 80% of randomly selected nodes fail, the remaining 20% still form a compact cluster with a path connecting any two nodes.
  - Random failure affects mainly the numerous small degree nodes.
- **Attack vulnerability:** reliance on hubs
  - the removal of a few key hubs splits the system into small isolated node clusters.

Topological robustness

- There is a strong relationship between the hub status of a molecule and its role in maintaining the viability and/or growth of a cell.
  - *S. cerevisiae* only ~10% of the proteins with less than 5 links are essential, but this fraction increases to over 60% for proteins with more than 15 interactions,
- The protein's degree of connectedness has an important role in determining its deletion phenotype.
  - Only ~ 18.7% of *S. cerevisiae* genes (~14.4% in *E. coli*) are lethal when deleted individually.
  - Simultaneous deletion of many *E. coli* genes is without substantial phenotypic effect.

Functional & dynamical robustness

- In a cellular network, each node has a slightly different biological function.
- The effect of a perturbation cannot depend on the node's degree only.
- Functional role of the whole complex determines the deletion phenotype of the individual proteins.

Functional & dynamical robustness

- Similar to topological robustness, dynamical and functional robustness are also selective:
  - some important parameters remain unchanged under perturbations, but others may vary widely.
  - For example, the adaptation time or steady-state behavior in chemotaxis show strong variations in response to changes in protein concentrations.

Functional & dynamical robustness

- First, adaptation and robustness are inherent network properties, and not a result of the fine-tuning of a component's characteristics.

- Second, robustness is inevitably accompanied by vulnerabilities:
  - Many cellular networks are well adapted to compensate for the most common perturbations, but they collapse when well-selected components are disrupted.

Functional & dynamical robustness

- Third, the ability of a module to evolve also has a key role in developing or limiting robustness.

- Forth, modularity and robustness are considerably quite intertwined,
  - with the weak communication between modules probably limiting the effects of local perturbations in cellular networks.

Beyond topology:

- Purely topological-based approaches have important limitations.
  - Ex: the activity of various metabolic reactions differ widely.

- Description of cellular network requires that both the **intensity** and the **temporal aspects** of interaction are considered.

Metabolic networks

- Flux: the amount of substrate that is being converted to a product within a unit of time.

- Flux distribution of e. coli is heterogeneous:
  - Reactions with flux that span several orders of magnitude coexist under same conditions.
  - Most reactions have small fluxes, coexisting with a few reactions with extremely high flux values.

Genetic regulatory interactions

- In perturbed transcriptome of *S. cerevisiae:* the degree to which each pair of genes is coexpressed or local similarities indicates that the functional organization of genetic regulatory networks might also be highly uneven.

- Most pairs have weak correlation coefficient, few pairs show quite a significant correlation coefficient.

Hot links

- Biochemical activity in both the metabolic and genetic networks is dominated by **hot links**.
- High activity interactions that are embedded into a web of less active interactions.
- Origin of this property of links is rooted in network topology.

Further directions

- Development of new theoretical methods
- Characterize the network topology
- Insights into the dynamics of motif clusters and biological function.
- Enhance data collection abilities.

Determinants of actual interactions:

- Cell's internal state
- position in the cell cycle
- Intracellular environment
- 3D shape
- Anatomical architecture
- Compartmentalization
- State of cells cytoskeleton

- The cell can be approached from
  - Bottom-up
    - Moving from molecules to motifs and modules
  - Top to the bottom
    - Starting from the network's scale-free and hierarchical nature and move to the organism-specific modules and molecules.
- Structure, topology, network usage, robustness and function are deeply interlinked.

Ultimate Aim of Network Biology:

- Most studies have focused on different subsets of the complex cellular networks.

- Integrated studies of all interactions which will offer further insights into how the **network of networks** contributes to the **cell's behavior**.

Effect of sampling on topology predictions of protein-protein interaction networks

Jing-Dong J Han
Denis Dupuy
Nicolas Bertin
Michael E Cusick
Marc Vidal

Overview:

- Available protein-protein interaction (PPI) networks obtained from:
  - Yeast two-hybrid
  - Co-AP/MS
- These partial networks show scale-free topologies

- Is the complete interactome scale-free?

Network Topologies:

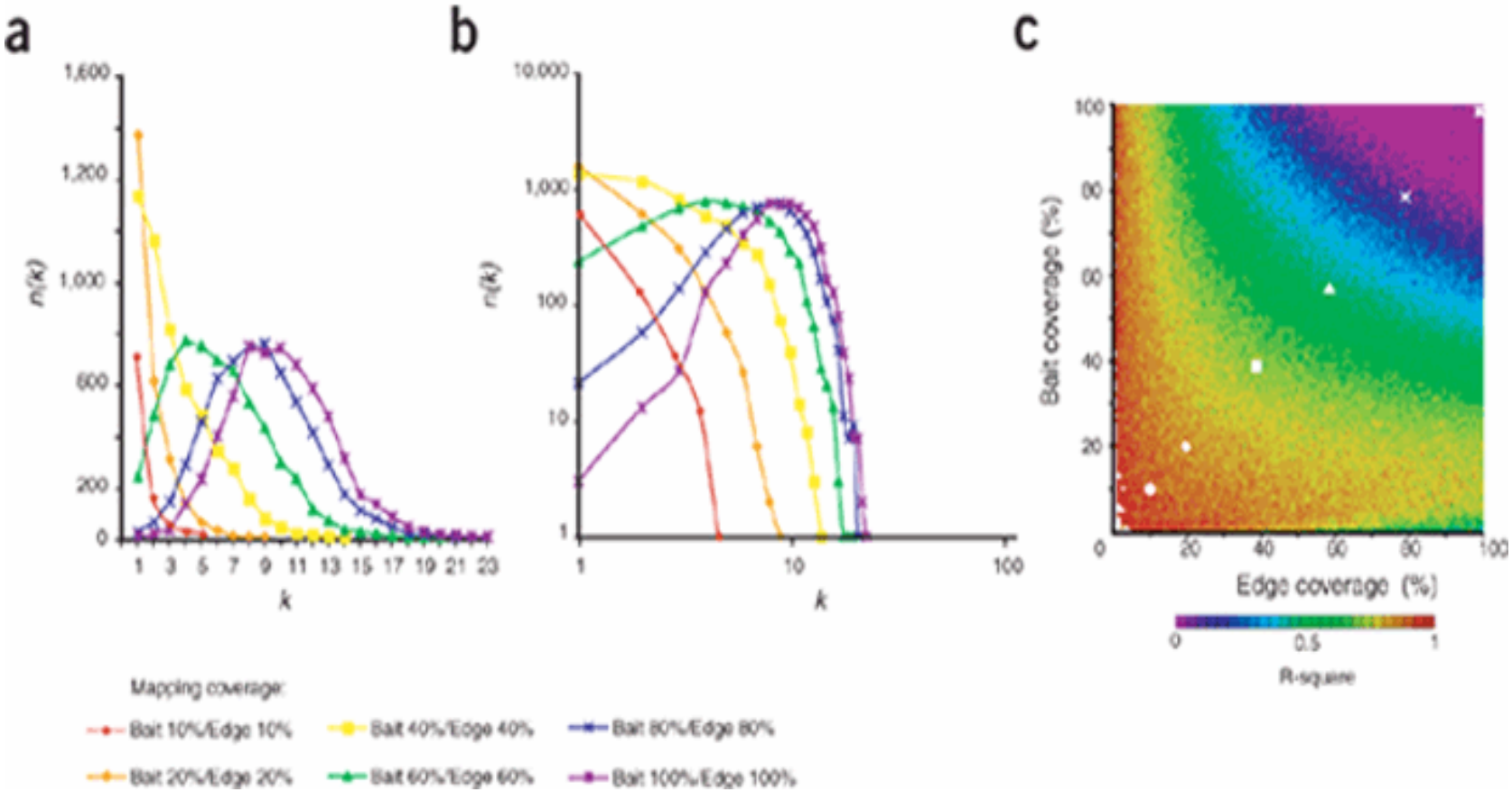
- 4 theoretical interaction networks are generated:
  - Random
  - Exponential
  - Power law
  - Truncated normal
- Sample from these networks.

Network parameters:

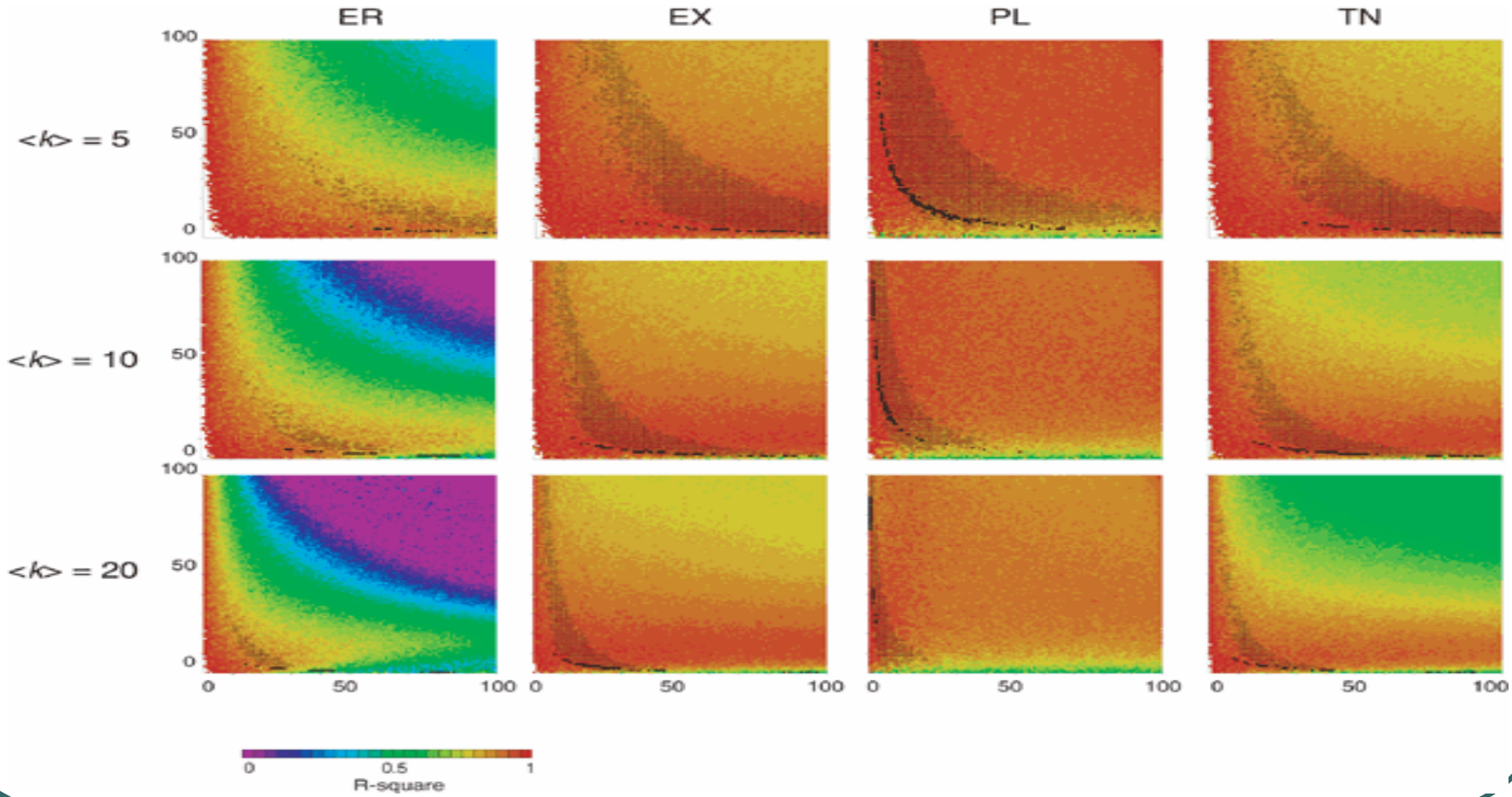
- Degree distribution

- log(n(k)) & log(k) linear regression R-square
- γ
- Fraction of nodes in the main component
- <k>

Sampling of an ER random network

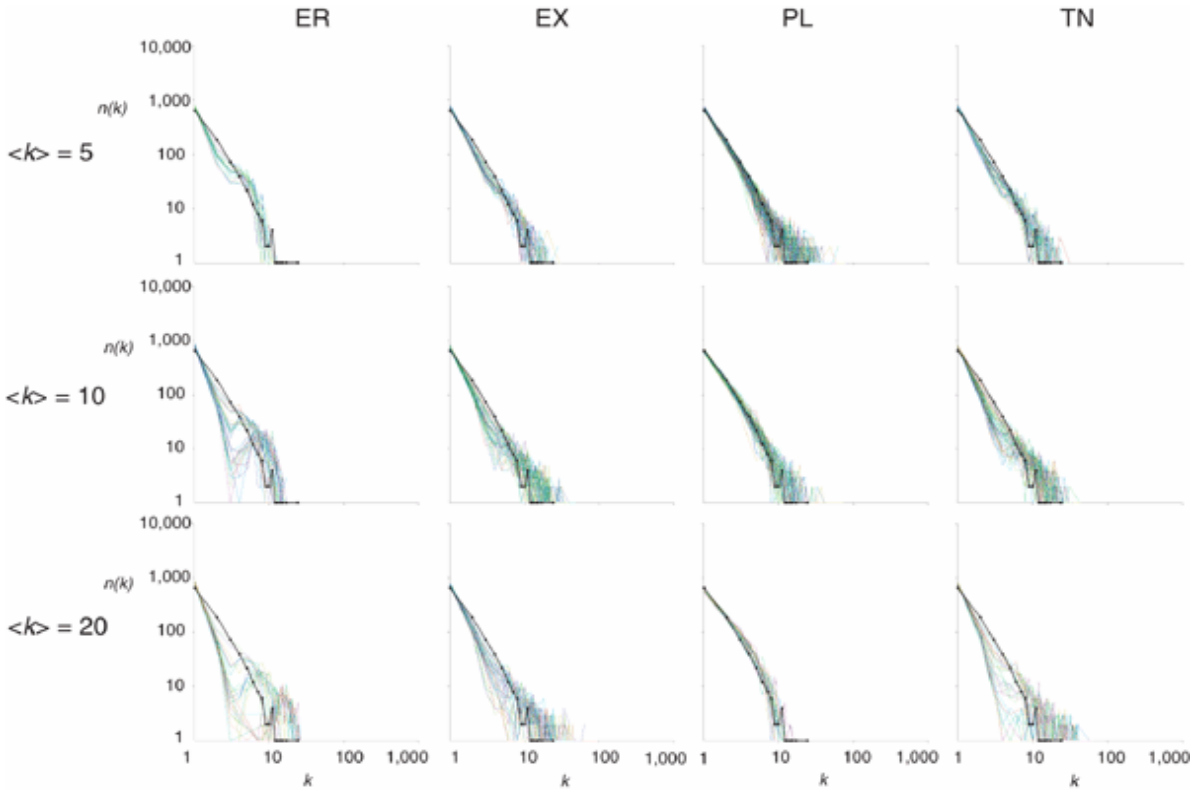


Sampled networks:



Degree distribution of sampled networks:

- # of sampled networks of comparable size to experimental Y2H maps: likelihood of a particular topology



Results:

- Partial sampling of these networks resulted in sub-networks with topological characteristics that were virtually indistinguishable from those of currently available Y2H-derived partial interactome maps.

- The observed scale-free topology of existing interactome maps can not be extrapolated to complete interactome