Seminar in Bioinformatics

Static properties and characteristics of networks and pathways
Introduction

• The aim: systematically catalogue all molecules and their interactions within a living cell.

• We would like to understand how these molecules and the interactions between them determine the function of this enormously complex machinery, both in isolation and when surrounded by other cells.
Why networks?

• A key challenge for biology in the twenty-first century is to understand the structure and the dynamics of the complex intercellular web of interactions that contribute to the structure and function of a living cell.

• Rapid advances in network biology indicate that cellular networks are governed by universal laws and offer a new conceptual framework that could potentially revolutionize our view of biology.

• You will see that the quantifiable tools of network theory offer surprising possibilities to understand the cell’s internal organization and evolution, fundamentally altering our view of cell biology.
Basic network nomenclature

• **Network / Graph** – we’ll look at components that interact with each other through pairwise interactions and we’ll reduce it to nodes and links (representing the interaction between to nodes).

• **Network types:**
  - protein–protein
  - protein–nucleic-acid
  - protein–metabolite
  - more complex functional interactions can also be considered within this representation
Architectural features of cellular networks
From random to scale-free networks

• First, let’s talk about **degree distribution**: The degree distribution - $P(k)$, gives the probability that a selected node has exactly $k$ links.  

(if there are $n$ nodes in total in a network and $n_k$ of them have degree $k$, we have $P(k) = n_k/n$)

• The degree distribution allows us to distinguish between different classes of networks.

• Now we can talk about **random networks** and **scale-free networks**.
Network models

• Random networks

• Scale-free networks

Poisson distribution

Power-law $P(k) \sim k^{-\gamma}$, where $\gamma$ is the degree exponent
Cellular networks are scale-free

- Most networks within the cell approximate a scale-free topology.
- The first evidence came from the analysis of metabolism, in which the nodes are metabolites and the links represent enzyme-catalysed biochemical reactions:

  a simple pathway, catalyzed by $Mg^{2+}$-dependant enzymes)
Cellular networks are scale-free

- Protein–protein interactions in diverse eukaryotic species also have the features of a scale-free network.

- Whereas most proteins participate in only a few interactions, a few participate in dozens — a typical feature of scale-free networks.
Cellular networks are scale-free

• Further examples of scale-free organization include genetic regulatory networks.
• However, not all networks within the cell are scale-free.
• Although the mathematical definition of a scale-free network requires us to establish that the degree distribution follows a power law, which is difficult in networks with too few nodes, the presence of hubs seems to be a general feature of all cellular networks.
• These hubs fundamentally determine the network’s behavior.
Small-world effect and assortativity

- **“Small world effect”** - any two nodes can be connected with a path of a few links only. This is a common feature of all complex networks.

- Although the “small-world effect” is a property of random networks, scale-free networks are **ultra small**— their path length is much shorter than predicted by the small-world effect 
  \( l \sim \log \log N \text{ vs. } \log N \)
Small-world effect and assortativity

• Within the cell, this ultra-small world effect was first documented for metabolism, where paths of only three to four reactions can link most pairs of metabolites.

• This short path length indicates that local perturbations in metabolite concentrations could reach the whole network very quickly.
Small-world effect and assortativity

- **Disassortativity** seems to be a property of all biological networks.

- Assortative nature of social networks vs. disassortative nature of all biological networks.

For example:
Small-world effect and assortativity

• Although the small- and ultra-small-world property of complex networks is mathematically well understood, the origin of disassortativity in cellular networks remains unexplained.
Evolutionary origin of scale-free networks

• It has emerged that two fundamental processes have a key role in the development of real networks:
  - most networks are the result of a growth process, during which new nodes join the system over an extended time period.
  - preferential attachment - nodes prefer to connect to nodes that already have many links.
Evolutionary origin of scale-free networks

- Indeed, 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.
Evolutionary origin of scale-free networks

• Growth and preferential attachment have a common origin in protein networks that is probably rooted in gene duplication.

• Proteins with a large number of interactions tend to gain links more often, as it is more likely that they interact with the protein that has been duplicated.
Evolutionary origin of scale-free networks

• Although the role of gene duplication has been shown only for protein interaction networks, it probably explains, with appropriate adjustments, the emergence of the scale-free features in the regulatory and metabolic networks as well.
Evolutionary origin of scale-free networks

• There is direct evidence that network growth is responsible for the observed topological features.

• The scale-free model predicts that the nodes that appeared early in the history of the network are the most connected ones, and indeed an inspection of the metabolic hubs confirms that.

• In the context of the protein interaction networks, cross-genome comparisons have found that, on average, the evolutionarily older proteins have more links to other proteins than their younger counterparts.
Motifs, modules and hierarchical networks
Motifs are elementary units of cellular networks

• Subgraphs capture specific patterns of interconnections that characterize a given network at the local level.

• Not all subgraphs occur with equal frequency.
Motifs are elementary units of cellular networks

- Some subgraphs, which are known as motifs, are overrepresented when compared to a randomized version of the same network.
- For example, triangle motifs, emerge in both transcription-regulatory and neural networks.
Modularity

• In general, modularity refers to a group of physically or functionally linked molecules (nodes) that work together to achieve a (relatively) distinct function.

• We would like not only to establish if a network is modular, but also to explicitly identify the modules and their relationships in a given network.
Modules

• Modules are groups of highly interconnected nodes.

• We can try to identify them directly from the graph’s topology and correlate these topological entities with their potential functional role.
Hierarchy organization of topological modules

• Module identification is complicated by the fact that at face value the scale-free property and modularity seem to be **contradictory**.

• Isolation vs. hubs in scale-free network

• Clustering and hubs naturally coexist, however, which indicates that topological modules are not independent, but combine to form a **hierarchical network**.
Network robustness
Robustness

• A key feature of many complex systems is their robustness, which refers to the system’s ability to respond to changes in the external conditions or internal organization while maintaining relatively normal behavior.
Topological robustness

• Intuition tells us that disabling a substantial number of nodes will result in an inevitable functional disintegration of a network.

• Random networks vs. complex systems
Topological robustness

• Scale-free networks do not have a critical threshold for disintegration — they 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.

• This is because random failure affects mainly the numerous small degree nodes, so we can see that there is a reliance on hubs.
Topological robustness

- This reliance on hubs, on the other hand, induces a so-called **attack vulnerability** — the removal of a few key hubs splinters the system into small isolated node clusters.

- This double-edged feature of scale-free networks indicates that there is a strong relationship between the **hub status** of a molecule (for example, its number of links) and its **role in maintaining the viability** and/or growth of a cell.

  
  **Essential proteins:**
  ~10% - less than 5 links
  60% - more than 15 interactions
Functional and dynamical robustness

• In a cellular network, each node has a slightly different biological function and therefore the effect of a perturbation cannot depend on the node’s degree only.

• The functional and dynamical robustness of cellular networks is supported by recent results that indicate that several modules are robust to many varied perturbations.
Beyond topology: characterizing the links
Characterizing the links

• Despite their successes, purely topology-based approaches have important intrinsic limitations.

• Therefore, an ultimate description of cellular networks requires that both the strength and the temporal aspects of the interactions are considered.
Characterizing the links

• In metabolic networks, the flux of a given metabolic reaction, which represents the amount of substrate that is being converted to a product within a unit of time, offers the best measure of interaction strength.

• A striking feature of the flux distribution of E. coli is its overall heterogeneity.

• This is captured by the flux distribution for E. coli (power law) - this indicates that most reactions have quite small fluxes, coexisting with a few reactions with extremely high flux values.
Characterizing the links

• The biochemical activity in both the metabolic and genetic networks is dominated by several ‘hot links’ that represent high activity interactions that are embedded into a web of less active interactions.

• The origin of this seemingly universal property of the links is probably rooted again in the network topology.
Conclusions
Conclusions

• The cell can be approached from the bottom up, moving from molecules to motifs and modules, or from the top to the bottom, starting from the network’s scale-free and hierarchical nature and moving to the organism-specific modules and molecules.

• It is impossible to ignore the apparent universality we have witnessed by delving into the totality of pairwise interactions among the various molecules of a cell.