Lecture Title and Date

Network Topology - Network Quantities

3/3/2025

Objectives of the Lecture

- Understand the key network quantities and how they are calculated
- Analyze network metrics and their implications for biological and computational network
- Explore the network motifs and cliques to understand the network structures.

Key Concepts and Definitions

- **Degree of a node**: The number of edges incident on the node
- **Indegree**: number of things going into something
- **Outdegree**: number of things going out of something
- **Clustering coefficient**: Ratio of existing links to maximum number of links for neighboring nodes
- **Average coefficient**: average of clustering coefficient of all nodes n
- **Path length**: Number of edges along a path
- **Shortest path length**: The minimum number of edges required to travel from one node to another.
- **Network motifs**: Regulatory modules within the network
- **Cliques**: Fully connected sub-components
- **K-core:** A subgraph in which each node has at least k connections
- **Defective clique**: A method to predict missing interaction in a network by analyzing partial cliques.

Main Content/Topics

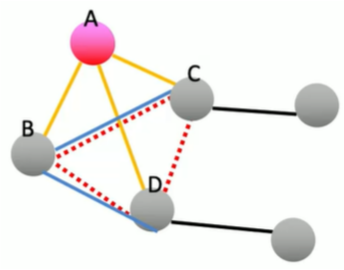
Degree and connectivity:

- Example: A node with 3 incoming edges and 5 outgoing edges has Indegree = 3, Outdegree = 5.
- In biological networks, a transcription factor (TF) may regulate several genes (outdegree), while a gene may be regulated by multiple TFs (indegree).

Clustering Coefficient:

- Definition: Measures the probability that a node's neighbors are also connected.

- Example Calculation:
    - Consider a node A with 3 neighbors.
    - 2 existing links (blue) out of 3 possible links (red).
    - Clustering coefficient for A, $C_A = 2/3$.

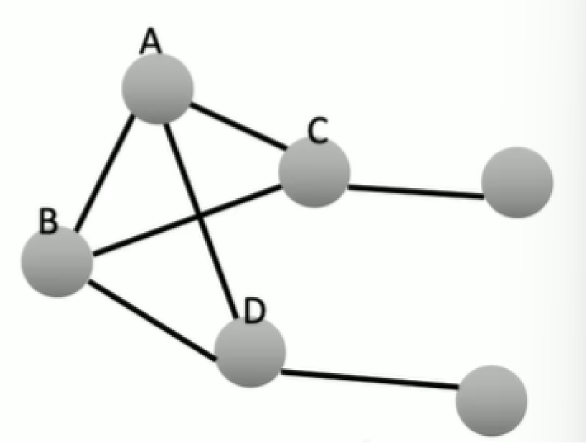


Average Clustering Coefficient

- Definition: The average clustering coefficient across all nodes in a network.

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

- Example Calculation:
    - Nodes A, B, C, D each have a maximum of 3 possible links.
    - A and C have 2 existing links, B and D have 1 existing link.
    - $C_A = 2/3$, $C_B = 2/3$, $C_C = 1/3$, $C_D = 1/3$.
    - Average Clustering Coefficient = $(C_A + C_B + C_C + C_D) / 4 = 0.5$.
- Importance: Measures network interconnectedness and global structure



Path Length and Small-World Networks

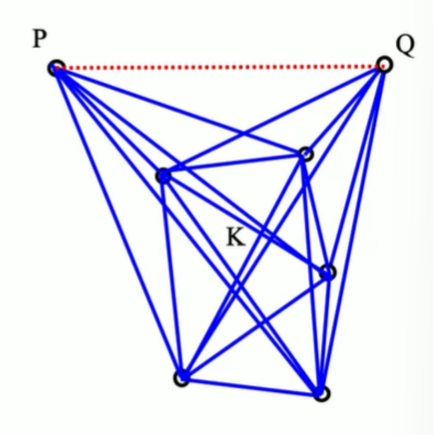
- Path length: The number of edges along a path between two nodes. Can be consider as number of intermediate TFs to reach final target
- Shortest path length (L(i,j)): The minimum number of edges needed to travel from node i to node j.
- Average Path Length:
  - The mean of shortest path lengths across all node pairs.
  - Networks with small L values exhibit small-world properties, meaning most nodes can be reached efficiently.

Network Motifs

- Definition: Small, recurring patterns in networks that carry out specific functions.
- Feed-forward loop (FFL): A common motif in regulatory networks, where a transcription factor regulates a second one, and both regulate a target gene.
  - Enhances network robustness and response control.

Cliques and K-Cores

- Definition: A clique is a fully connected subgraph.
- K-core: A subgraph where all nodes have at least k edges.
- Issues in Network Analysis:
  - High-throughput experiments can miss interactions, leading to incomplete networks.
  - Solution: Defective Cliques
    - If proteins P and Q interact with all members of clique K, they are likely to interact with each other.
    - Helps in predicting missing protein interactions.
    - Example: A partially observed clique with missing edges. If proteins **P** and **Q** interact with all other members of a clique, we infer they likely interact.



Discussion/Comments

- Network topology provides a fundamental framework for analyzing the structure and function of complex systems, from biological networks to computational models. Using degree distributions, clustering coefficients, and path lengths allows us to quantify the interconnectedness and efficiency of a network.
- The concept of network motifs offers a way to recognize recurring functional units within larger structures, helping us identify key regulatory elements in biological networks
- Understanding these network properties not only helps in modeling and predicting system behavior but also has applications in fields such as disease modeling, social network analysis, and technological systems design.

Suggest Readings and Other Helpful References of Key Concepts

There is only one suggested reading for this lecture:

1. McGillivray, P., Clarke, D., Meyerson, W., Zhang, J., Lee, D., Gu, M., Kumar, S., Zhou, H., & Gerstein, M. (2018). *Annual Review of Biomedical Data Science, 1(1)*, 153–180. Network analysis as a grand unifier in biomedical data science. https://doi.org/10.1146/annurev-biodatasci-080917-013444 *(Section 2.)*

Are the readings for the class useful? If so, are the specific subsections useful or would change.

If not, are there other references you could suggest? Please suggest one.

Yes, this reading is useful, especially for the part: 2.4: Network motifs, network logic and network stability as it talks about the evolutionary significance of network motifs like negative autoregulation and feedforward loops. It helps to expand the knowledge and enhance what we learn from class.

References ISL/ESL (if any)

- No relevant references
- First part of network is discussed in the lecture summary 25m10a