Lecture Title and Date

Network Topology - Intro
March 3, 2025

Objectives of the Lecture

- Understand the primary motivation for using network topology in computational biology and how it serves as a midpoint
- Explain how networks can be used in biological and non-biological disciplines
- Understand how networks can go beyond simple visual representations and allow us to make predictions

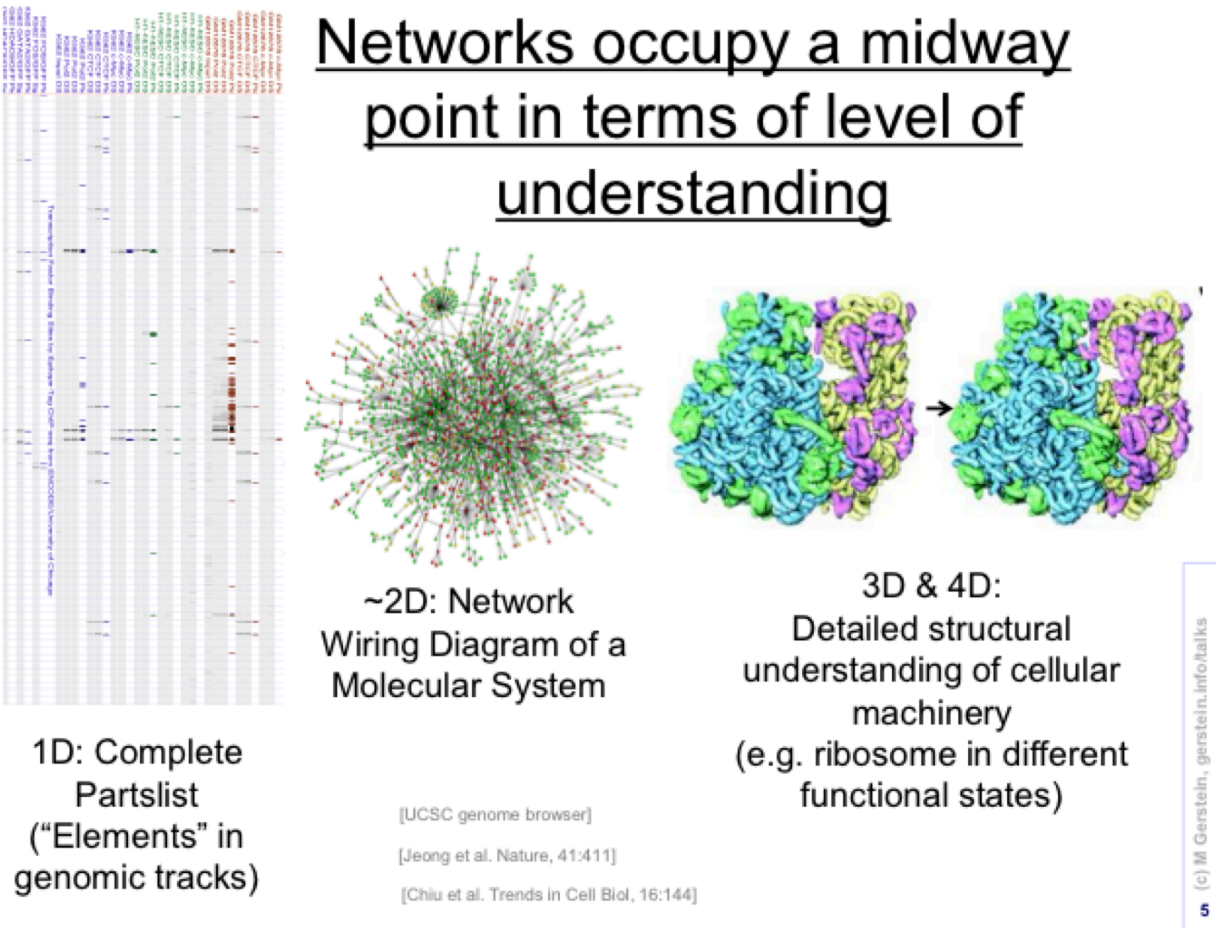
Key Concepts and Definitions

- **Network/Network topology**: two-dimensional systems-level representation of biological or non-biological data as nodes and relationships as edges
- **Linear track:** One-dimensional representation of genomic data where features (genes, assays) are displayed as separate tracks without explicit relationships
- **Mechanistic model:** High dimensional models of genomic data
- **Guilt by association**: principle underlying network science, using the relative position of an entity in a network and its connectivity to infer its function.
- **Interactome:** Application of networks in bioinformatics, explicitly represents molecular interactions within a biological system

Main Content/Topics

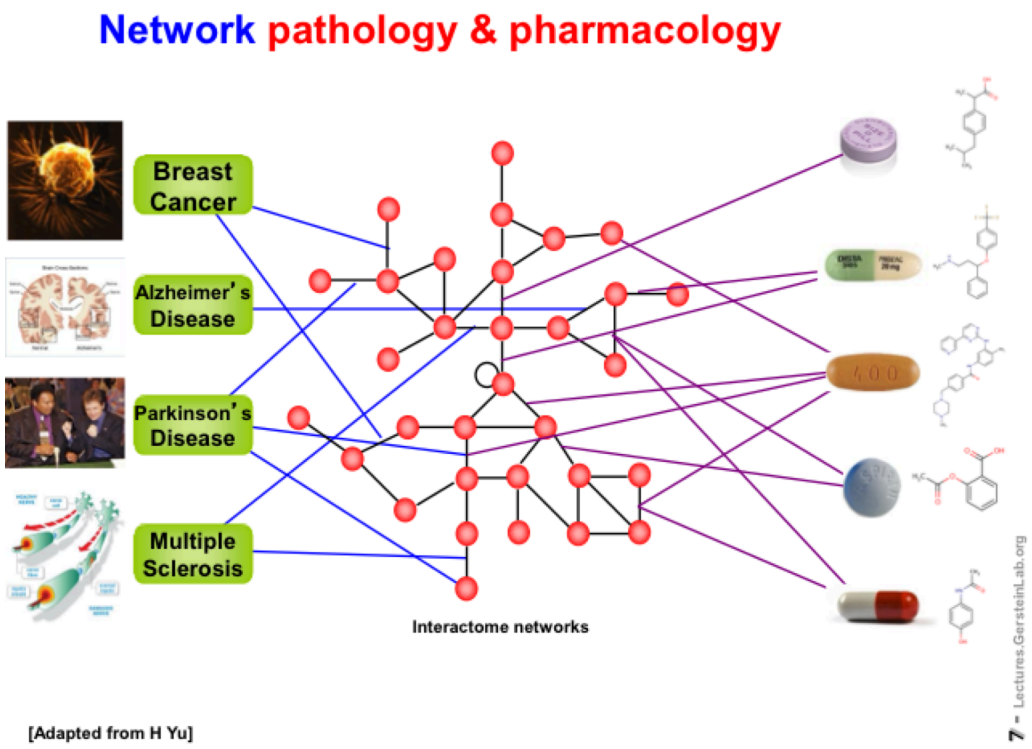
- Genomic data is often represented as thousands of one-dimensional **linear tracks**, each track representing a different type of data, including gene expression, protein binding, and regulatory elements. This is a very “parts” view. We list all the different genes, but we don’t know how they work together.
    - For example, in the slide on the left, each element represents a CHIP-Seq assay of transcription factor binding from the UCSC genome browser.
    - Although this approach may be useful in communicating results for a small-scale experiment, as the volume of data increases this representation will not scale effectively and might be too simplistic for our needs.
- By contrast, we can also represent biological data as detailed **mechanistic models** in higher dimensions.
    - For example, in the slide on the right side of the slide, we can represent ribosomes in the genome as three-dimensional and four-dimensional mechanistic structures.[1]
    - However, this approach is too complex and might be too ambitious for our needs

- Rather than considering transcription factors as individual parts or tracks in isolation, we can think of them as a system all connected in terms of their interactions.
- **Networks** offer a halfway point in terms of our level of understanding. These two-dimensional representations are not too simple and yet offer comprehensive systems-level insights into molecular wiring without requiring detailed mechanistic models in bioinformatics.[2]
    - For example, networks often look like two-dimensional “hairballs”. The slide below shows an example from the yeast proteome.[3]

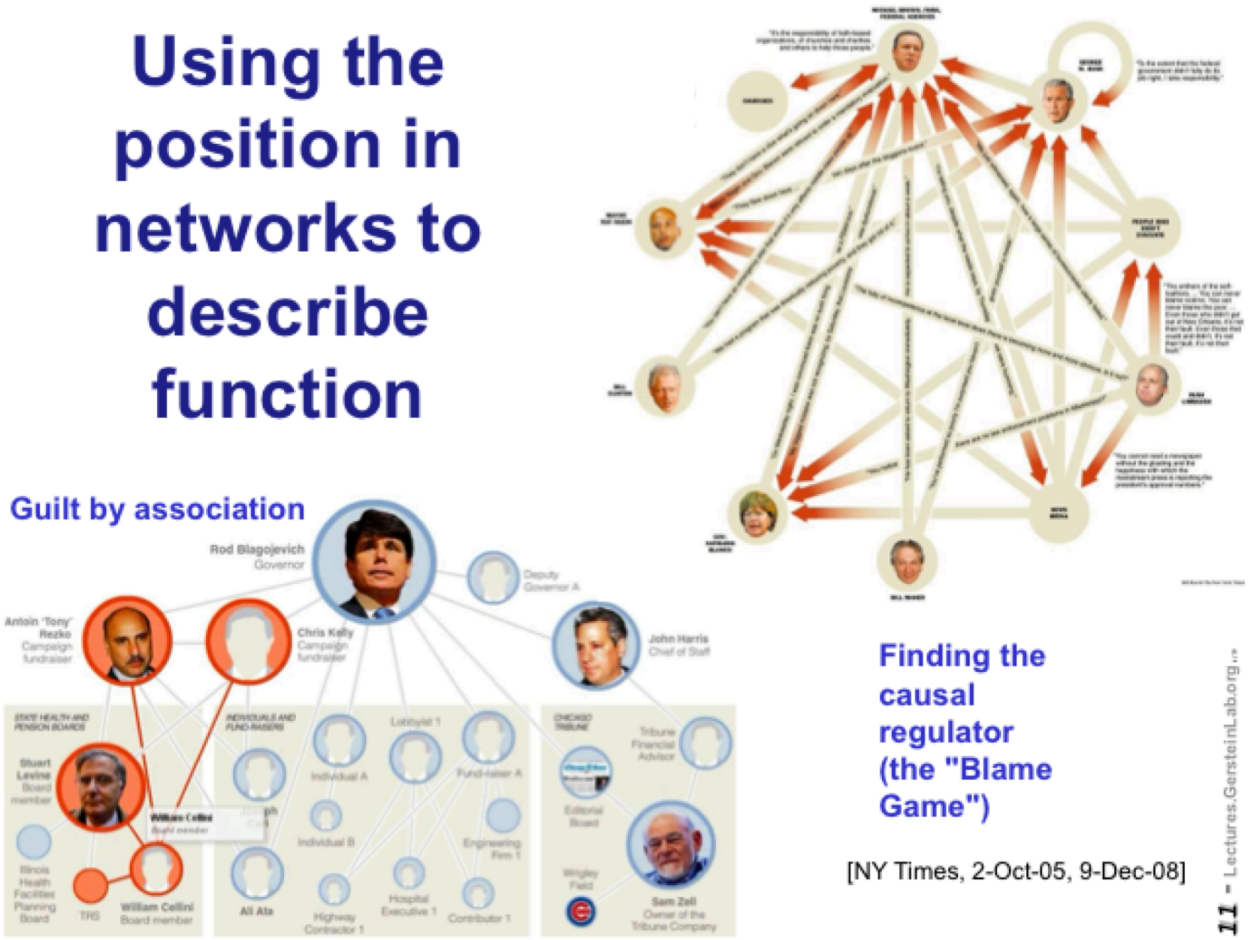


- Networks can also shed light on disease pathologies and visualize larger **interactome** networks and drug-disease associations.[4]
    - For example, there are many pathways and groups of genes and subnetworks (red nodes and black lines) that are deregulated in diseases like breast cancer, Alzheimer’s disease, Parkinson’s disease, and Multiple Sclerosis.
    - Interactomes have been a popular approach for visualizing many cancer pathways.[5]

- We can model medications that are acting on the whole molecular pathway or interfering with specific parts of the pathway (purple lines)



- Networks and more broadly network science can be thought of as a universal language that has been applied across many disciplines including molecular biology, epidemiology, computer science, environmental science, engineering, neuroscience, and even social media.
    - For example, networks can be used to model protein interactions, disease spread, the internet, food social, electric circuits, neural networks, and social networks
- Beyond just simple visual representations, networks can also be used for prediction. The **guilt-by-association principle** is widely used to predict unknown functions of genes by analyzing network proximity to known functional elements.[6]
    - Outside of biological data, this principle was used in social network analysis. Rod Blagojevich, former Governor of Illinois, was linked to corruption through his acquaintances. In other words, since he was connected to some shady people, it meant that he must be shady too!
    - We can also use this principle to identify causality. If we represent politicians blaming each other as a network, we can figure out the cause by following the blame (red arrows).



- Networks allow us to integrate lots of diverse biological data, including: metabolic pathways (e.g., TCA cycle), transcription regulatory networks, protein-protein interactions, co-expression relationships, and gene interactions (synthetic lethality)

Discussion/Comment

- Networks are not just simple two-dimensional models, but offer a universal language for us to explicitly communicate the mechanism of disease or shed light on a gene pathway
- Networks give us a systems-level understanding and allow us to identify patterns or predict key functions (e.g, **guilt-by-association principle**) and causal regulators that would have been missed if we used the **simple linear track** view

Suggest Readings and Other Helpful References of Key Concepts

Are the readings for the class useful?

- Yes, there is only one suggested reading from the lecture (Reference #2 in red). Section 1 provides useful background on networks, their real-world applications, and key network quantities. For more details on network quantities, refer to the lecture summary for 2510b.

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

- The other references are example figures used in the lecture slides (References #1, 3, 5) and explanations of key terms like the Guilt-by-association principle (References #4 and #6)

1. Chiu, W., Baker, M. L. & Almo, S. C. Structural biology of cellular machines. *Trends Cell Biol.* **16**, 144–150 (2006). **[Figure 1, High-dimensional mechanistic model of ribosome]**
2. McGillivray, P. *et al.* Network Analysis as a Grand Unifier in Biomedical Data Science. *Annu. Rev. Biomed. Data Sci.* **1**, 153–180 (2018). **[Suggested Reading from Lecture: Section 1]**
3. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001). **[Figure 1, Yeast proteome]**
4. Bang, D., Lim, S., Lee, S. & Kim, S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nat. Commun.* **14**, 3570 (2023). **[Figure 1, Guilt-by association applied to drug repurposing through knowledge graphs]**
5. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008). [**Figure 5a, Glioblastoma network]**
6. Gillis, J. & Pavlidis, P. The Impact of Multifunctional Genes on 'Guilt by Association' Analysis. *PLOS ONE* **6**, e17258 (2011). **[Another real-world example of how the Guilt-by-Association can be used to predict gene function]**

References ISL/ESL (if any)

- No relevant references
- Neural networks discussed in a separate lecture summary (please see 25t3)