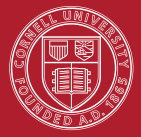


Cornell University

Lecture 15 - Biological Networks

Models for real networks

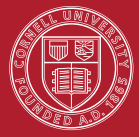
ECE 5260 – ORIE 5735

Chenyue Zhang

Cornell Tech

March 18

Content

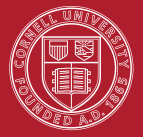


Cornell University

Lecture goal

The goal of this lecture is not to study biology in detail, but to understand how **biological systems can be represented and analyzed as networks**.

Outline

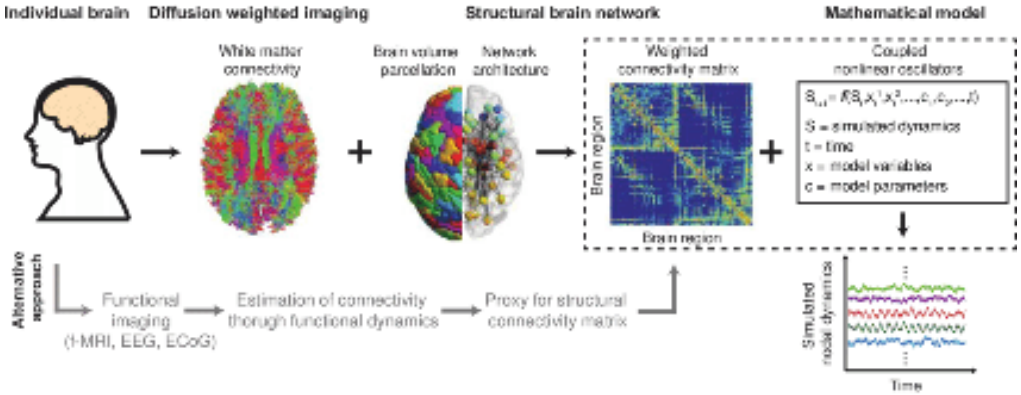


Cornell University

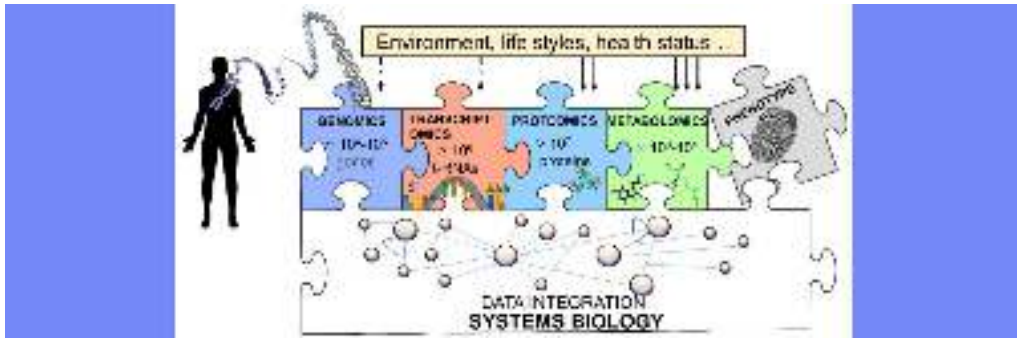
Relational networks: networks in biology

Examples

- Brain networks



- Biochemical networks: metabolic, protein or gene interactions



- Ecological networks: herding, swarm intelligence, predator-prey

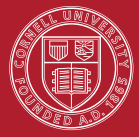
Takeaway: Many biological systems can be modeled as networks; the main modeling choice is how to define **nodes, edges, and interactions**.

Three biological systems, three network problems

- **Ecological networks**: how do we model multiple interaction types? $\rightarrow$ multilayer graphs
- **Brain networks**: how do we infer edges from activity data? $\rightarrow$ time-series network inference
- **Gene regulatory networks**: what local structures appear in directed biological networks? $\rightarrow$ motifs in directed graphs

Takeaway: This lecture uses biology to introduce **three important network-science questions**.

Network abstraction across biology



System	Nodes	Edges
Ecosystem	species	ecological interactions
Brain	brain regions	neural connections
Gene regulation	genes	regulatory influence

Common idea

The biological details are different, but the network abstraction is the same.

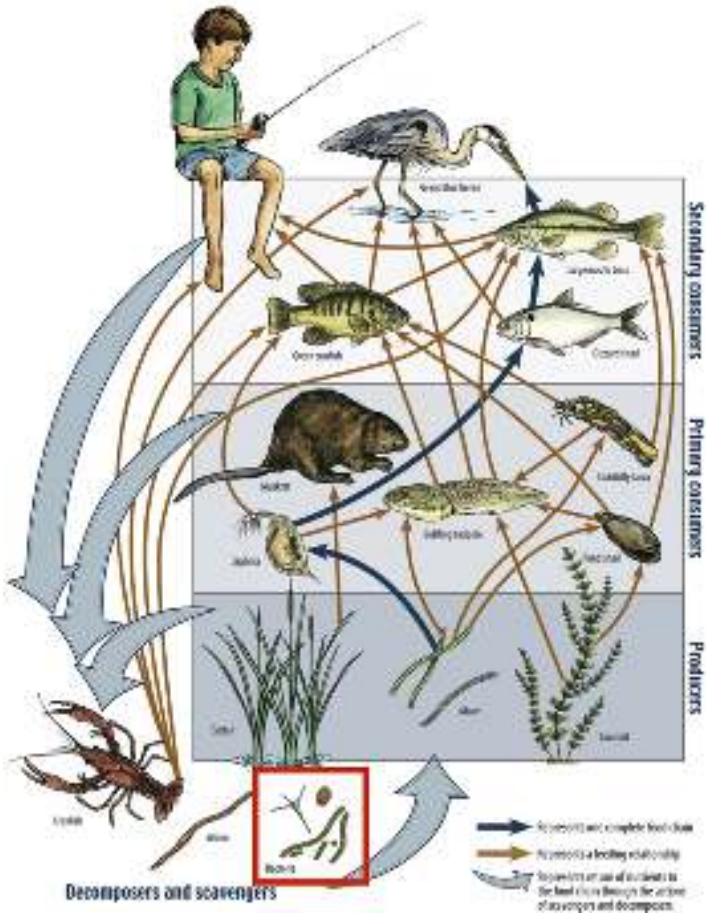
Takeaway: Very different biological systems can be represented using the same **network framework**.

Ecological networks

Interpretation: **Nodes** are species, and **edges** are interactions such as predation, competition, or cooperation.

- Food webs are a classic example: they show who eats whom
- Network structure helps us study stability and cascading effects
- They connect biological questions to graph structure

Takeaway: Ecological systems can be viewed as networks of interacting species.



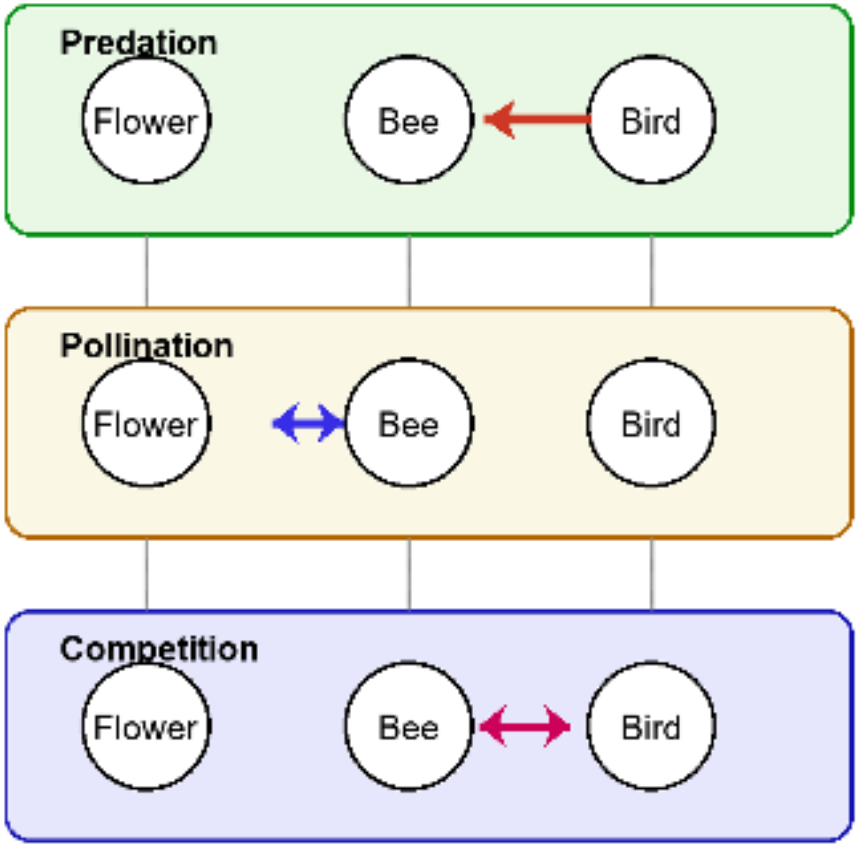
Ecological systems are often multilayer

Why multilayer? Species often interact in several ways at once, so a **multilayer network** keeps those interaction types separate.

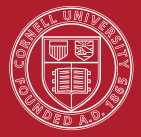
- The same species can appear in several layers
- Examples: predation, pollination, host-parasite
- A single adjacency matrix may mix different mechanisms

$$G = \{V, E^{(1)}, E^{(2)}, \ldots, E^{(L)}\}$$

Takeaway: Ecological systems are often better modeled as **multilayer networks** than as one single graph.

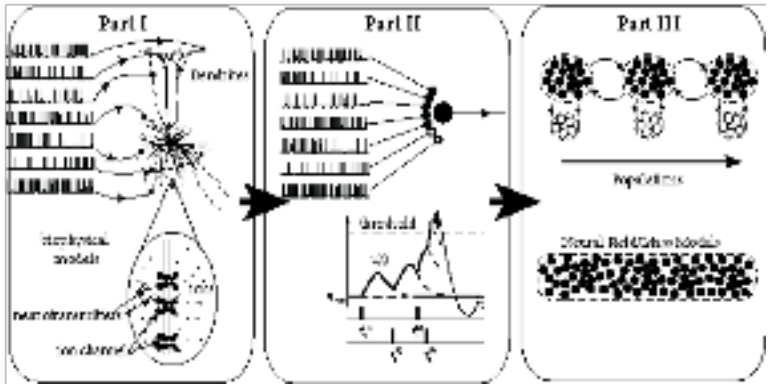


Outline



Cornell University

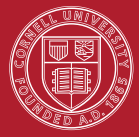
Nodes: groups of neurons



- A node represents a population of neurons
- We study interactions between groups, not detailed internal wiring
- We do not model single neurons one by one
- Each group has its own structure, connectivity, and activity

Takeaway: The graph keeps population-level interactions and hides microscopic detail.

Defining the nodes: brain parcellation



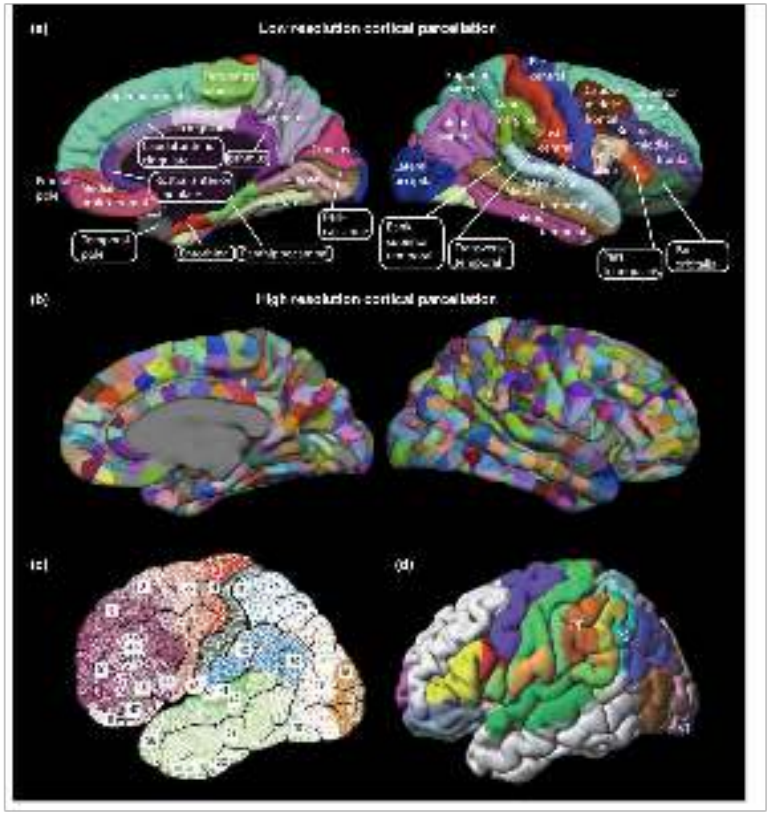
- Brain nodes are defined by dividing the brain into regions
- Two common choices are:
  1. brain anatomy (**spatial / structural connectivity**)
  2. areas that become active together during a task or behavior
- This division into regions is called **parcellation**

Interpretation

Parcellation means **dividing the brain into regions** that will become nodes in the network.

Takeaway: Parcellation determines **what counts as a node**, so it directly changes the graph we analyze.

Parcellations of the brain

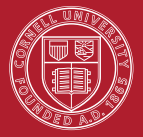


Identification of large-scale structural network nodes in the human brain by four methods currently in use.
(a)Automated parcellation of a single subject's structural MR image into nodes based on the geometry of large sulcal landmarks.
(b)High-resolution parcellation with arbitrary granularity.
(c)Classical Brodmann atlas based on cytoarchitectonic features.
(d) The Julich–Dusseldorf cytoarchitectonic probabilistic brain atlas, based on observer independent mapping of cortical areas in ten post-mortem brains.

- Coarse and fine parcellations give different graph resolutions

Takeaway: Changing the parcellation changes node count and network scale.

Edges: pathways between brain areas



- Edges are long white-matter pathways connecting brain areas
- These pathways are naturally **directed** because signals have direction
- If communication goes both ways, we often model the link as **undirected**
- Key question: how do we identify these edges from real data?

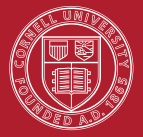
Takeaway: Once nodes are fixed, the next modeling step is to define edges between brain regions.

Structural connectivity

- Structural connectivity describes the **physical wiring** between brain regions
- It tells us which interactions are anatomically possible
- Evidence can come from tract tracing, diffusion MRI, or anatomical measurements
- The key modeling point is that structure defines a network of possible communication pathways

Takeaway: Structural connectivity aims to recover the **physical pathways** that make communication possible.

Large-scale functional networks



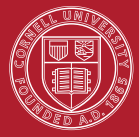
Key distinction

Structural connectivity describes **physical wiring**, while functional connectivity describes **coordinated activity**.

- A functional network is a set of brain areas that work together
- Structural connectivity tells us which interactions are *possible*
- Functional connectivity tells us which areas actually *work together*
- It supports perception, memory, and action

Takeaway: Structural connectivity describes **possible links**; functional connectivity describes **coordinated activity** observed in data.

Identifying functional nodes

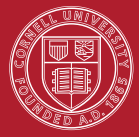


Areas with high task-related activity are candidate functional nodes:

- PET (positron emission tomography): measures metabolic activity
- fMRI (functional MRI): measures blood-oxygen changes (BOLD)
- EEG / LFP: shows synchronized electrical activity

These modalities give different measurements, but all help define activity-based network nodes. **Takeaway:** PET, fMRI, and EEG do not measure the same signal, but they often highlight related functional structure.

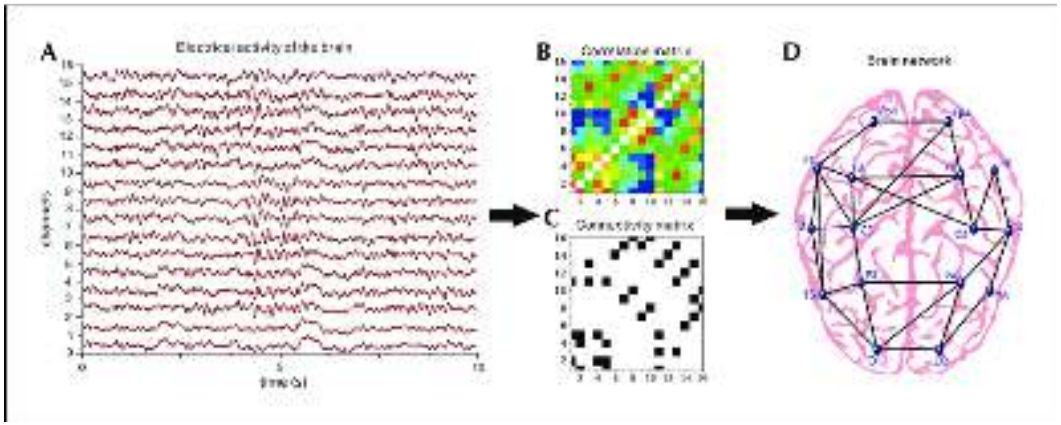
Functional edges: measuring interactions



- Edges measure how strongly activity in two nodes is related over time
- **Undirected** edges from symmetric measures:
  - Cross-correlation, spectral coherence, phase synchrony
- **Directed** edges from methods that aim to capture causal influence:
  - Granger causality
  - Dynamic causal modeling
- Different time scales reveal different aspects of communication

Takeaway: Functional edges are inferred from **time-series relationships**, either symmetric or directed.

Correlation-based edges



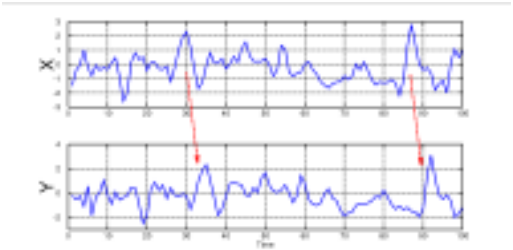
- Let $x_i(t)$ be the signal at node i
- The cross-correlation at lag τ is

$$[\boldsymbol{R}(\tau)]_{ij} = \frac{\sum_{t=1}^{T}(x_i(t) - \hat{\mu}_i)(x_j(t+\tau) - \hat{\mu}_j)}{T\hat{\sigma}_i\hat{\sigma}_j}$$

- This defines a symmetric weighted adjacency matrix

Interpretation: Correlation measures whether two brain regions show **similar activity patterns over time**. **Takeaway:** Correlation turns co-fluctuation into an undirected weighted edge.

Granger causality: the key idea



- If the past of one signal improves prediction of another, it may influence it
- Formally, $X(t)$ is a Granger cause of $Y(t)$ if

$$P(Y(t) \in A \mid \mathcal{H}(t)) \neq P(Y(t) \in A \mid \mathcal{H}_Y(t))$$

where $\mathcal{H}(t)$ is the full history and $\mathcal{H}_Y(t)$ is the history of Y alone

Caution in brain imaging: In fMRI, Granger-causal interpretation should be treated carefully because the observed BOLD signal is **slow and indirect** relative to neural activity.

Takeaway: Granger causality asks whether one past signal improves prediction of another.

A toy example of Granger causality

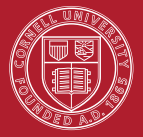
Consider two signals:

$$x_1(t) = 0.8x_1(t-1) + \epsilon_1(t),$$
$$x_2(t) = 0.5x_2(t-1) + 0.6x_1(t-1) + \epsilon_2(t)$$

- The past of x_1 helps predict x_2
- So x_1 Granger-causes x_2
- But x_2 does not Granger-cause x_1 in this model

Takeaway: Granger causality asks whether adding the past of one signal improves prediction of another.

Granger causality: multivariable model



- For recordings from many brain areas, fit a vector auto-regressive (VAR) model:

$$\boldsymbol{x}(t) = \sum_{\tau=1}^{L} \boldsymbol{A}(\tau)\, \boldsymbol{x}(t-\tau) + \boldsymbol{\epsilon}(t)$$

- The past of node j Granger-causes node i if $[\boldsymbol{A}(\tau)]_{ij}$ is nonzero for some lag τ
- The matrices $\{\boldsymbol{A}(\tau)\}$ define a **directed, weighted adjacency matrix**
- Significance can be tested statistically from the fitted coefficients

Takeaway: In a VAR model, nonzero lag coefficients define directed functional edges.

A VAR example with 3 regions and 2 lags

Let $\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) & x_3(t) \end{bmatrix}^T$ and consider the VAR(2) model

$$\boldsymbol{x}(t) = \boldsymbol{A}(1)\boldsymbol{x}(t-1) + \boldsymbol{A}(2)\boldsymbol{x}(t-2) + \boldsymbol{\epsilon}(t)$$

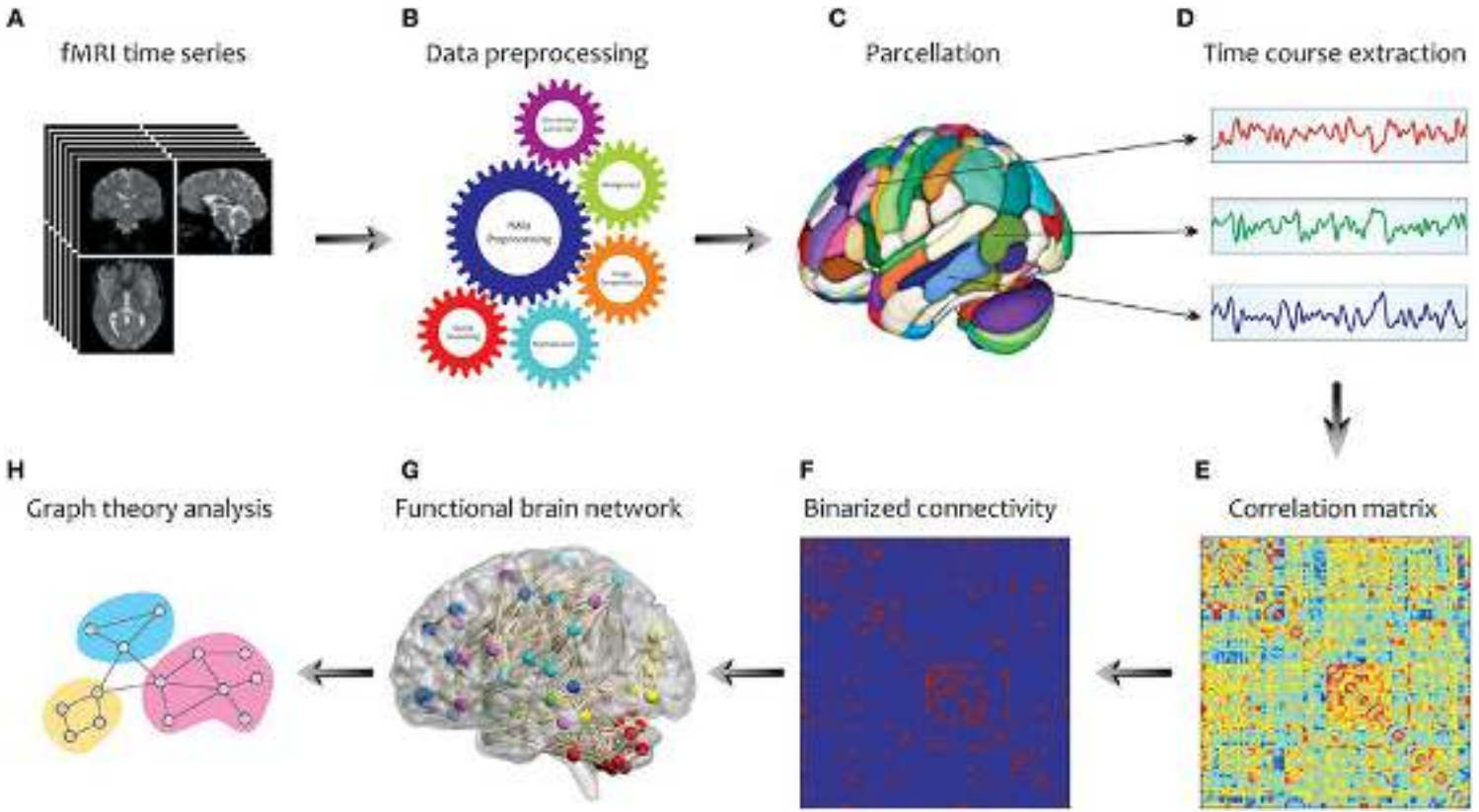
with

$$\boldsymbol{A}(1) = \begin{bmatrix} 0.6 & 0.2 & 0 \\ 0 & 0.5 & 0.3 \\ 0.1 & 0 & 0.4 \end{bmatrix}, \qquad \boldsymbol{A}(2) = \begin{bmatrix} 0 & 0 & 0.1 \\ 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \end{bmatrix}.$$

- Nonzero entries in $\boldsymbol{A}(1)$ define lag-1 directed influences
- Nonzero entries in $\boldsymbol{A}(2)$ define lag-2 directed influences
- Example: $x_1(t)$ depends on $x_2(t-1)$ and $x_3(t-2)$

Takeaway: A VAR model converts multiregion time-series data into directed, weighted, lag-dependent edges.

From brain data to network analysis



- Once the graph is built, we can apply the same network-analysis tools from Module 1

Takeaway: After graph construction, standard network tools can be applied.

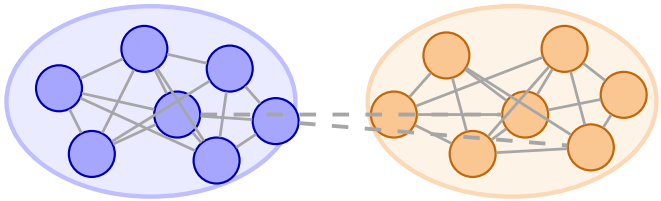
Small-world and community structure

Small-world structure

- strong local clustering; short paths between most node pairs
- a few long-range links support efficient global communication

Community structure

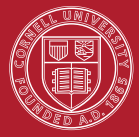
- dense internal connections within groups
- relatively sparse connections between groups
- groups often align with functional or anatomical modules



Why this matters: Brain networks must support both specialized local processing and efficient long-range communication.

Takeaway: Brain networks often combine local clustering, global efficiency, and modular organization.

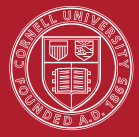
Hubs and centrality



- **Hub structure**: some brain areas have many more connections than others
  - in some datasets, the degree distribution appears heavy-tailed
- **Degree centrality**: identifies highly connected regions
- **Betweenness centrality**: identifies bridge nodes between modules
- **Motifs**: recurring small connection patterns

Takeaway: Centrality measures identify the most connected or most influential regions in the brain network.

The metabolic cost of brain networks



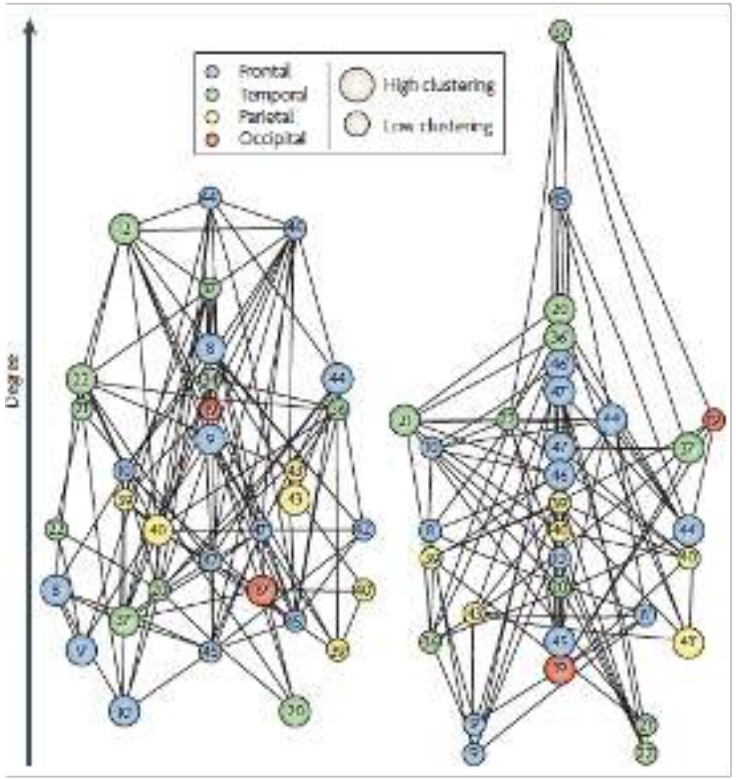
- Brain networks depend on physical distance: the chance of an edge scales as

$$P(\text{edge}) \propto d_{ij}^{-\alpha}$$

  where d_{ij} is physical distance and α controls how expensive long links are
- Two forces compete:
  1. Energy cost of long-distance connections
  2. Functional benefit of connecting useful regions
- Healthy brains usually maintain a good balance between these forces
- When the balance breaks down, it may indicate disease

Takeaway: Brain networks are shaped by a tradeoff between **wiring cost** and **functional benefit**.

Network features and disease



Disease-related disorganization of brain from structural MRI data.
In both parts, the nodes (circles) represent cortical regions and the connections represent high correlation in the gray matter density.
The nodes are arranged vertically by degree and are separated horizontally for clarity of representation. Nodes with high clustering are larger. (a) healthy volunteers has a hierarchical organization characterized by low clustering of high-degree nodes 24. (b) The equivalent network constructed from MRI data on people with schizophrenia shows loss of this hierarchical organization — high-degree nodes are more often highly clustered.

- Changes in clustering, degree, or hubs can help distinguish healthy and diseased brains

Takeaway: Graph statistics can serve as markers of neurological disease.

Recent directions in brain-network research

- **More detail**: connect large-scale activity recordings with detailed structural maps
- **More personalized models**: build brain-network maps for each individual, not only group averages
- **More complex interactions**: move beyond static pairwise edges to time-varying and multi-region interactions
- **More data integration**: combine connectivity with gene-expression, cellular, and molecular data

Main message

Brain-network research is becoming more detailed, more personalized, and more multimodal.

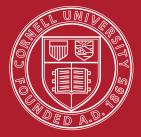
Takeaway: Recent work moves from simple average graphs to richer, individualized, and multimodal brain-network models.

What did brain networks add to our toolbox?

- Ecological networks introduced **multilayer interactions**
- Brain networks introduced **edge inference from time-series data**

Takeaway: Different biological systems motivate different graph problems.

Outline



Cornell University

Gene Regulatory Networks: Graph–Dynamics

Key idea: A gene regulatory network is a *dynamical system on a graph*.

State:

$$x_i(t) \;=\; \text{expression level of gene } i$$

Dynamics:

$$\frac{dx_i}{dt} = f_i(x)$$

Interpretation:

- Graph: who interacts with whom
- Function $f_i(\cdot)$: how interactions happen

Activation vs Repression:

- Activation: f_i increases with x_j; repression: f_i decreases with x_j

A Simple Dynamical Model

Example model:

$$\frac{dx_i}{dt} = -x_i + \sum_j w_{ij}\,\sigma(x_j)$$

Interpretation:

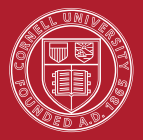
- $-x_i$: natural decay
- w_{ij}: interaction strength
- $\sigma(\cdot)$: nonlinear response

Sign of w_{ij}:

- $w_{ij} > 0$: activation
- $w_{ij} < 0$: repression

Remark: Similar structure to neural networks and graph models.

Nonlinear Regulation: Hill-Type Models



Activation (Hill function):

$$f(x_j) = \frac{x_j^k}{K^k + x_j^k}$$

Repression:

$$f(x_j) = \frac{1}{1 + (x_j/K)^k}$$

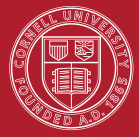
Parameters:

- K: activation threshold
- k: cooperativity (steepness)

Key properties:

- Saturation
- Nonlinear switching
- Biochemical realism

Dynamics and Emergent Behavior



Different dynamics lead to different behaviors:

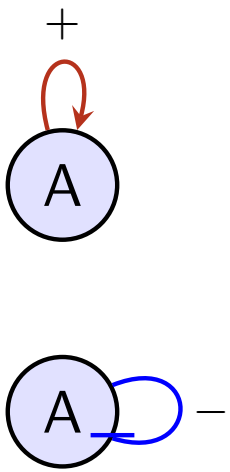
- Stable equilibrium (cell type)
- Oscillations (biological rhythms)
- Multistability (cell differentiation)

Key idea:

- Network structure constrains behavior
- Dynamics determines outcome

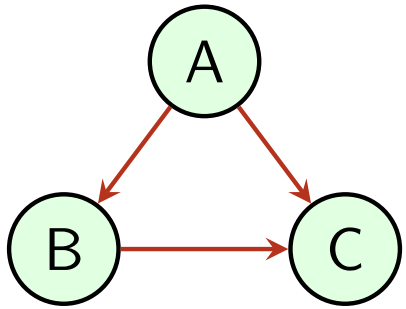
Motifs as local dynamical circuits

Feedback



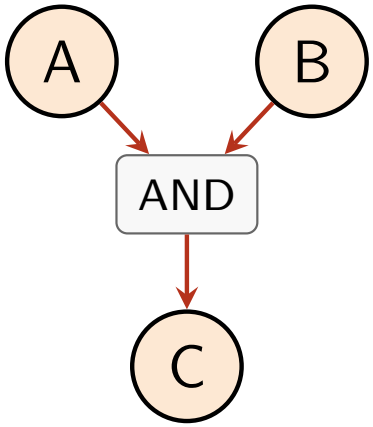
- positive feedback → switch / memory
- negative feedback → stability / oscillation

Feed-forward loop



- direct + indirect regulation
- filters noise
- delays response

Cooperation



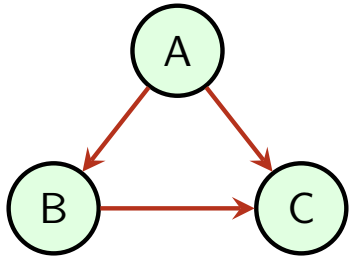
- multiple TFs act together
- nonlinear combination
- logic-gate behavior

Interpretation: Motifs are recurring regulatory circuits, and each motif supports a characteristic dynamical function.

Takeaway: Motifs are reusable local circuits for switching, stabilization, timing, and combinatorial control.

Why motifs matter: a classic feed-forward example

- Motif analysis asks which small directed patterns appear more often than expected by chance
- A classic example is the **feed-forward loop**
- This motif is overrepresented in transcriptional networks such as *E. coli*



Why this matters: Overrepresentation suggests that the motif performs a function the system repeatedly needs.

Takeaway: Motif analysis links local directed graph patterns to biological function.

GRN vs Opinion Dynamics: A Comparison

DeGroot model:

$$x_i(t+1) = \sum_j W_{ij} x_j(t)$$

GRN dynamics:

$$\frac{dx_i}{dt} = -x_i + \sum_j w_{ij}\, \sigma(x_j)$$

Similarities:

- Network-driven interactions
- State evolves based on neighbors

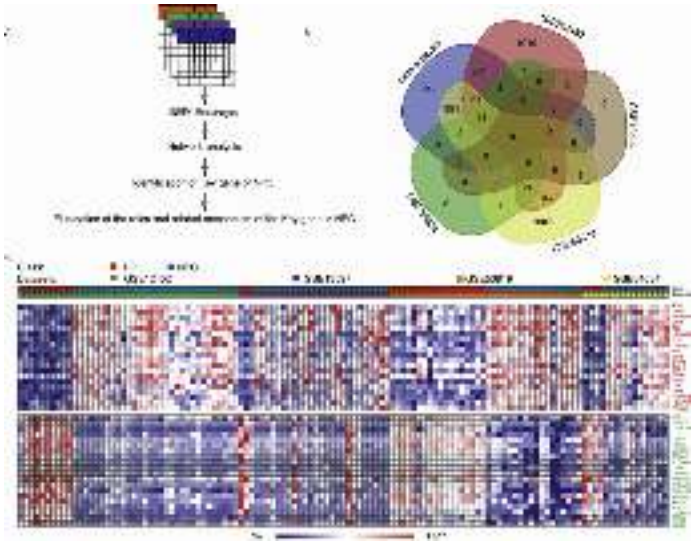
Key differences:

- Linear vs nonlinear
- Averaging vs activation/repression
- Consensus vs richer behaviors

Inferring GRNs from data

Data sources

- **RNA-seq / microarrays**: measure gene-expression levels
- **Perturbation data**: how the system responds when genes are modified

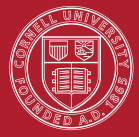


Inference problem

- estimate directed edges from biological measurements
- combine dependence, causality, and dynamical modeling

Takeaway: GRN inference is a directed network-reconstruction problem.

How GRN research is changing



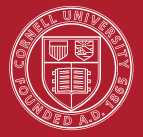
- **More detailed**: combine gene expression with chromatin accessibility and related molecular measurements
- **More causal**: use perturbation data, such as CRISPR experiments, to support directed regulatory edges
- **More spatial**: study regulation together with cell and tissue location
- **More data-driven**: use larger models and richer representations, while still requiring careful benchmarking

Main message

GRN research is moving from static average descriptions toward richer, more causal, and more context-aware regulatory models.

Takeaway: Recent GRN research is becoming more detailed, more causal, more spatial, and more data-rich.

What did GRNs add to our toolbox?

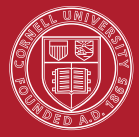


Cornell University

- Ecological networks: multilayer graphs
- Brain networks: inferring edges from time series
- GRNs: motifs in directed networks

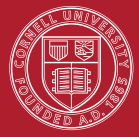
Takeaway: The same biological-network theme leads to three different network-science questions.

Outline



Cornell University

Network abstraction across biology



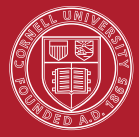
Common principle

Although ecosystems, brains, and gene regulation are very different biological systems, they can all be studied using the same **network abstraction**.

System	Nodes	Main network problem
Ecosystem	species	multilayer interactions
Brain	brain regions	edge inference from time series
Gene regulation	genes	directed motifs

Takeaway: The same network abstraction can motivate **different graph problems** in different biological systems.

Summary



- Ecological networks highlight the need for **multilayer graph models**
- Brain networks highlight **edge inference from time-series data**
- Structural connectivity describes **possible links**; functional connectivity describes **observed coordination**
- Gene regulatory networks highlight **motifs in directed graphs**
- Biology provides motivating examples, but the main lesson is about **network problems**