

# Connectome Coding

Joshua Vogelstein

Dept of BME, Johns Hopkins university

jovo@jhu.edu / <http://neurodata.io>

Please interrupt and ask questions!

- what is connectome coding
- applications of connectome coding

**what is connectome coding?**

**Neural coding:** characterizing the relationship between the **ongoing environment** and **neural activity**

**Ongoing environment:** stimulus, movements, rewards, etc.

**Connectome Coding:** characterizing the relationship between the **past environment** and the **neural connectivity**

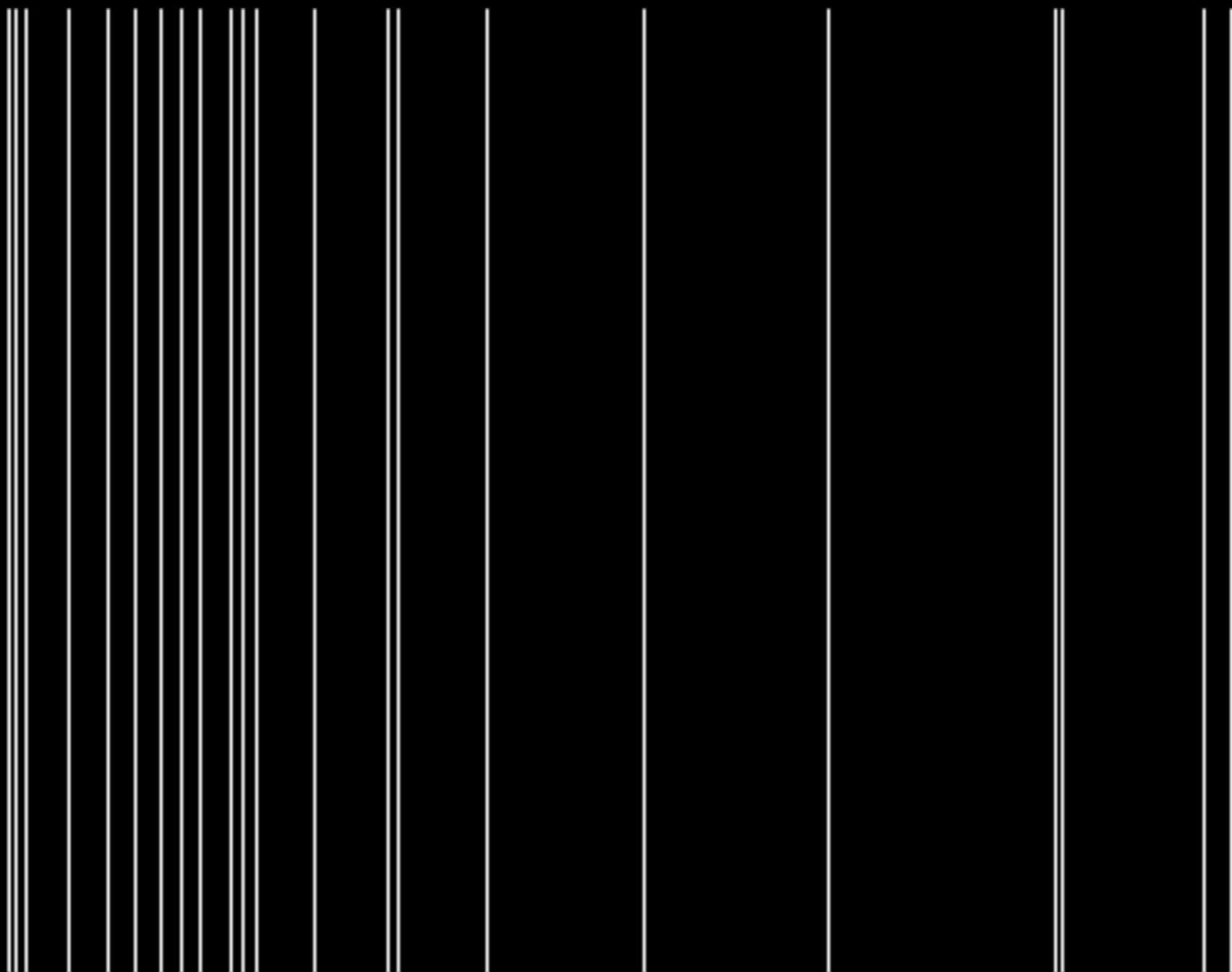
**Past environment:** genome, psychiatric condition, memory, location, etc.

# Principles of Data Science

- Look at it
- Keep it simple

Let's do it for **neural coding**

# Look at it



# Keep it Simple

Neural Encoding:  $P[r | s]$

Neural Decoding:  $P[s | r]$

Neural Code:  $P[s, r]$

# Keep it Simple

We need the joint distribution of brain stuff & external stuff:

- Joint distribution:  $P[r,s] = P[r | s] P[s] = P[s | r] P[r]$
- So, we need  $P[r]$ ,  $P[s]$ ,  $P[r|s]$ ,  $P[s|r]$
- Let's start with  $P[r]$

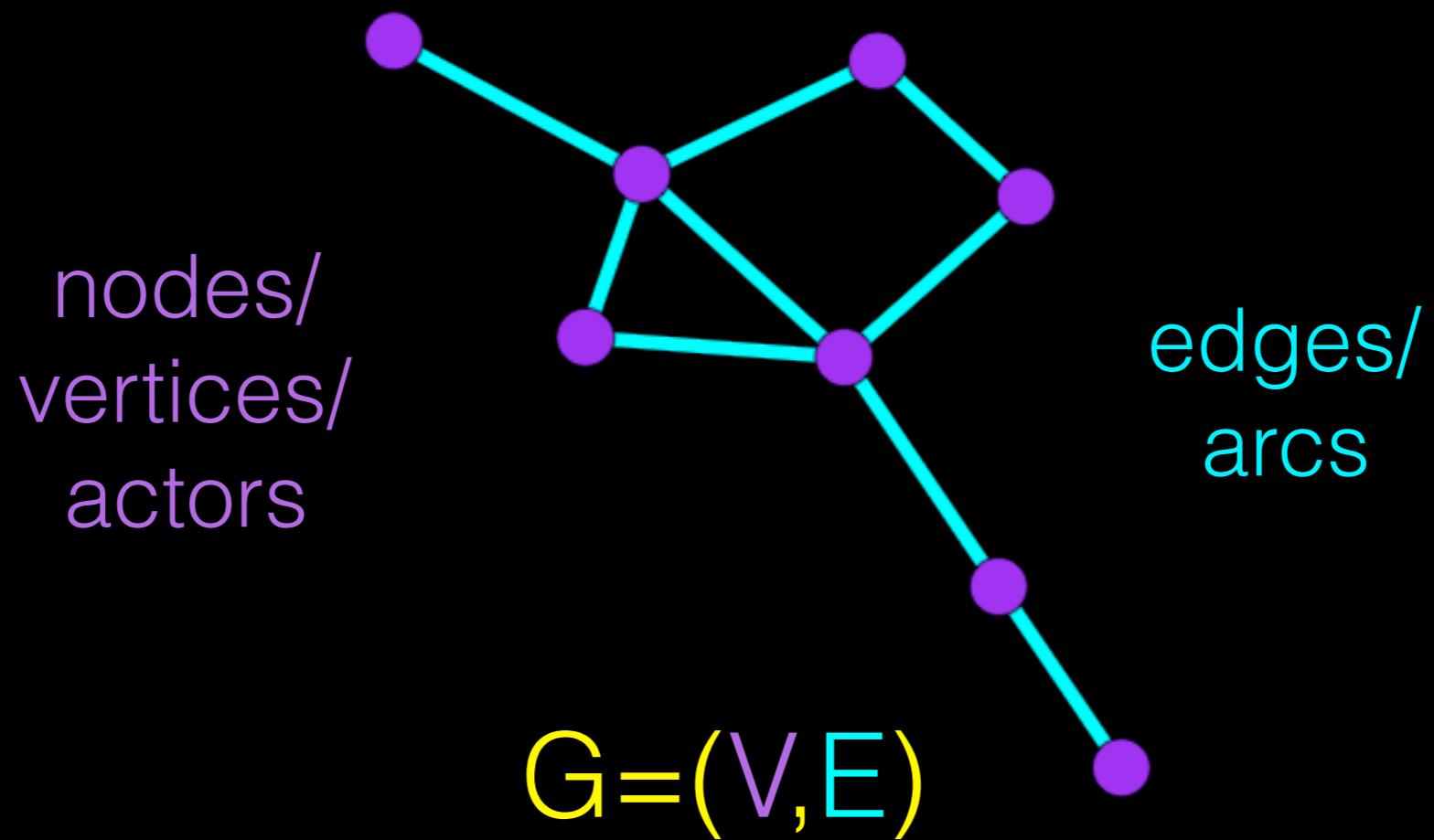
# Simplest Neural Code

- Each spike is independent
- Probability of a spike at any time is  $\lambda$
- $P[r] = \text{Poisson}(\lambda)$

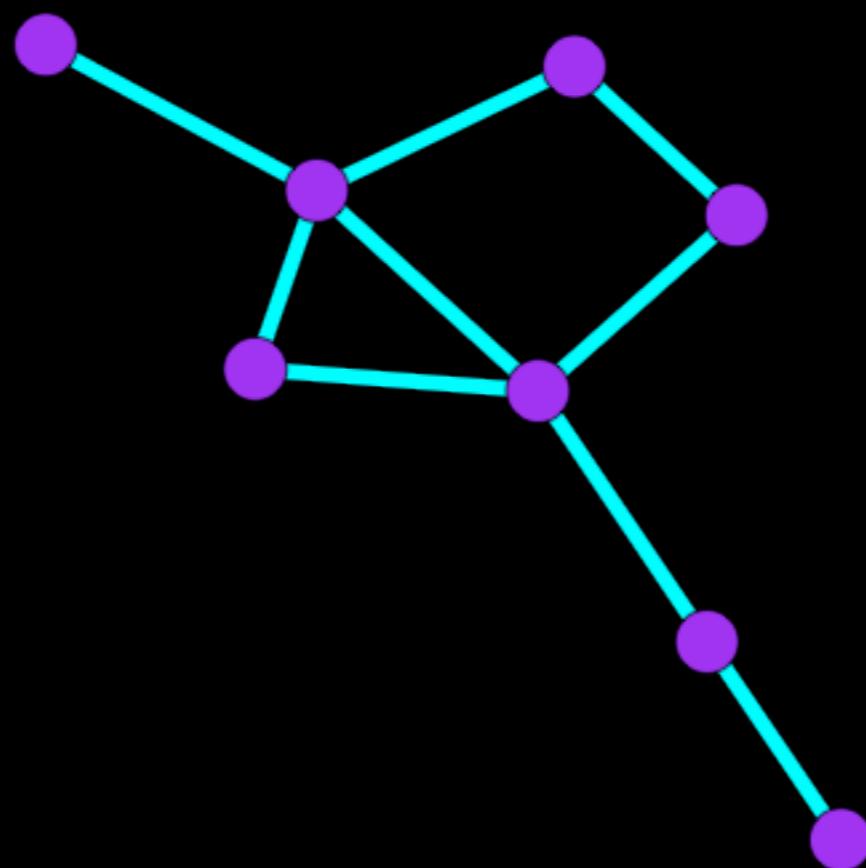
# P[g]

Let's do it for networks, P[g]

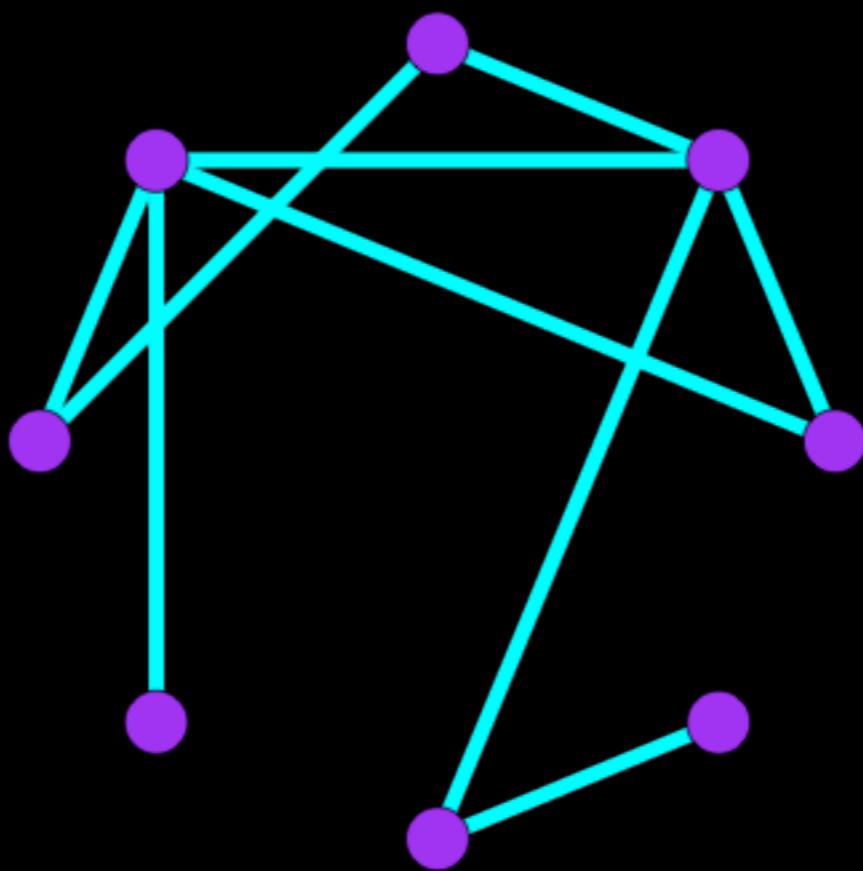
# a simple graph



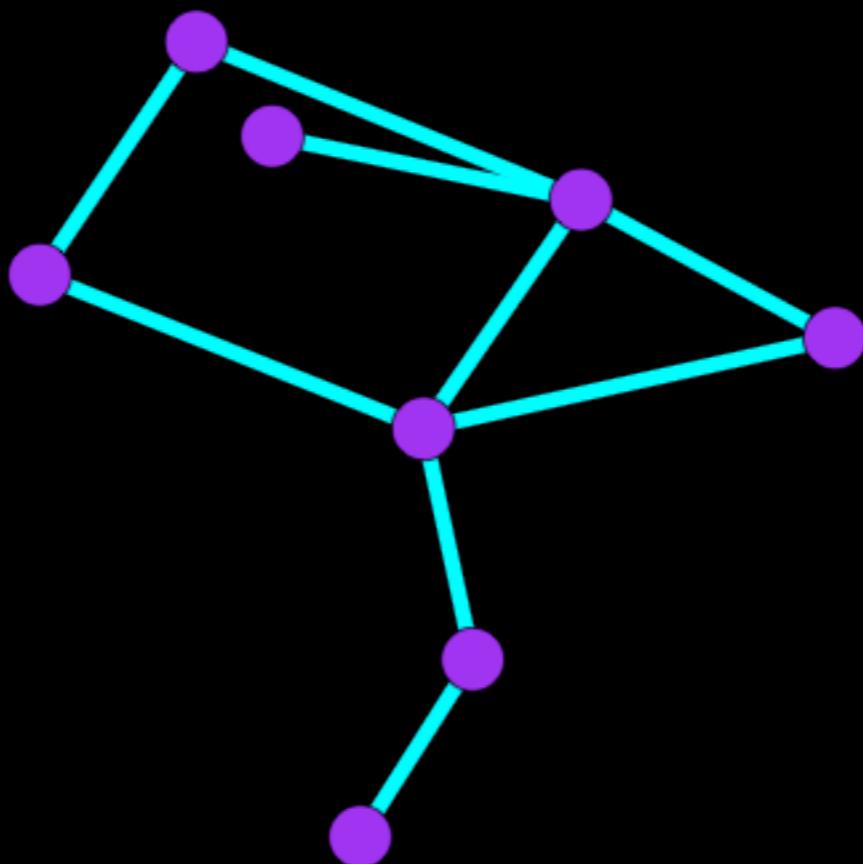
the same graph



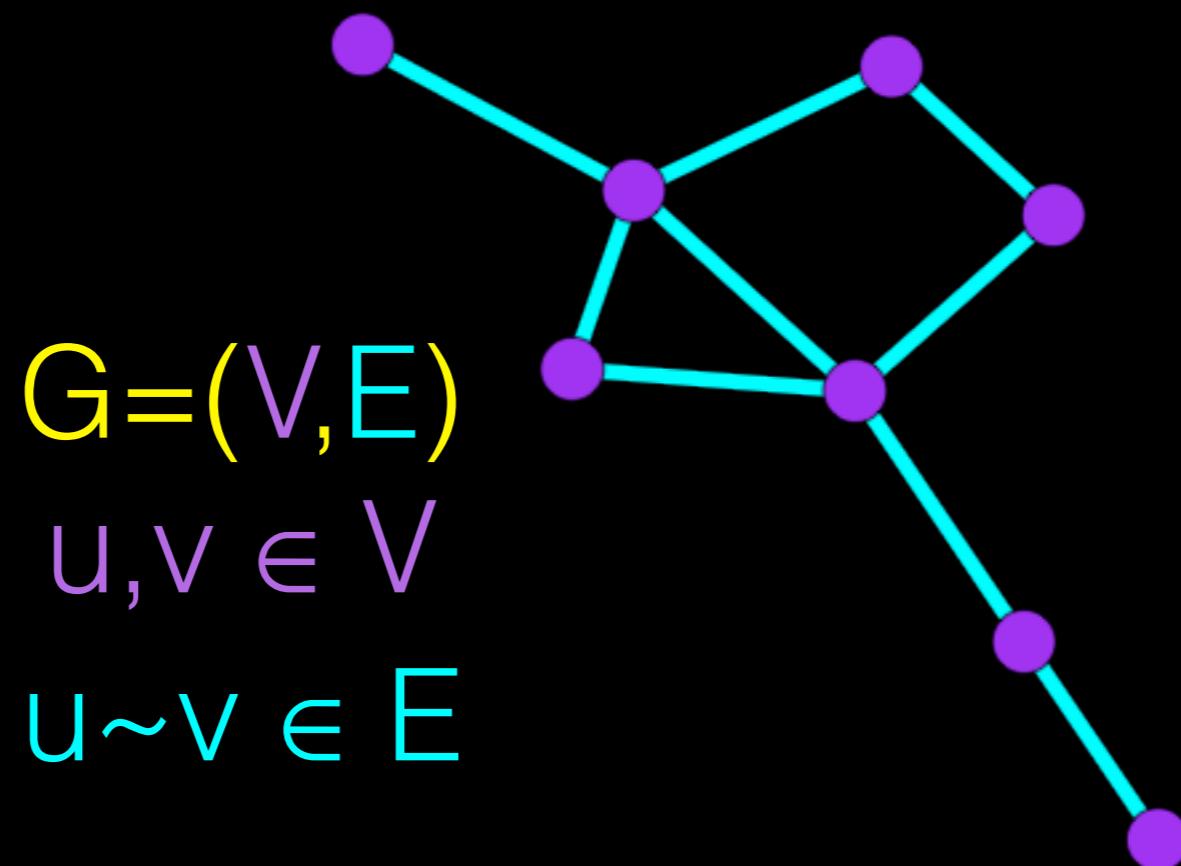
the same graph



the same graph

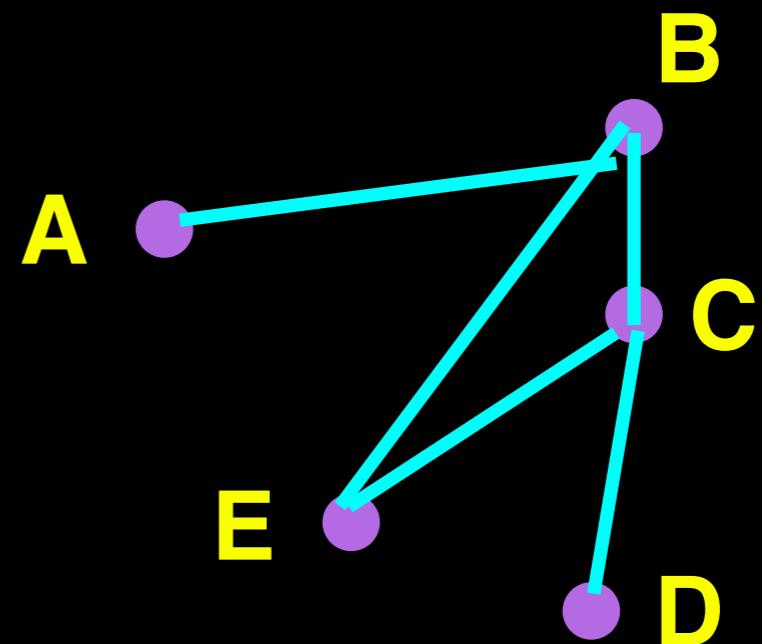


the same graph

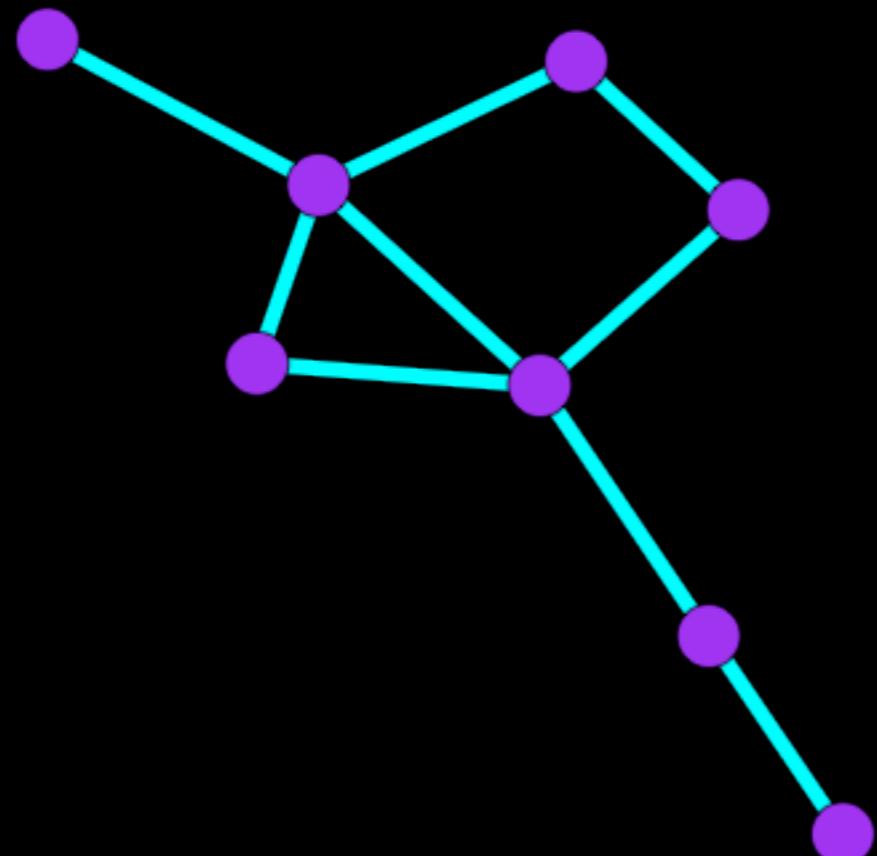


# labeled graph

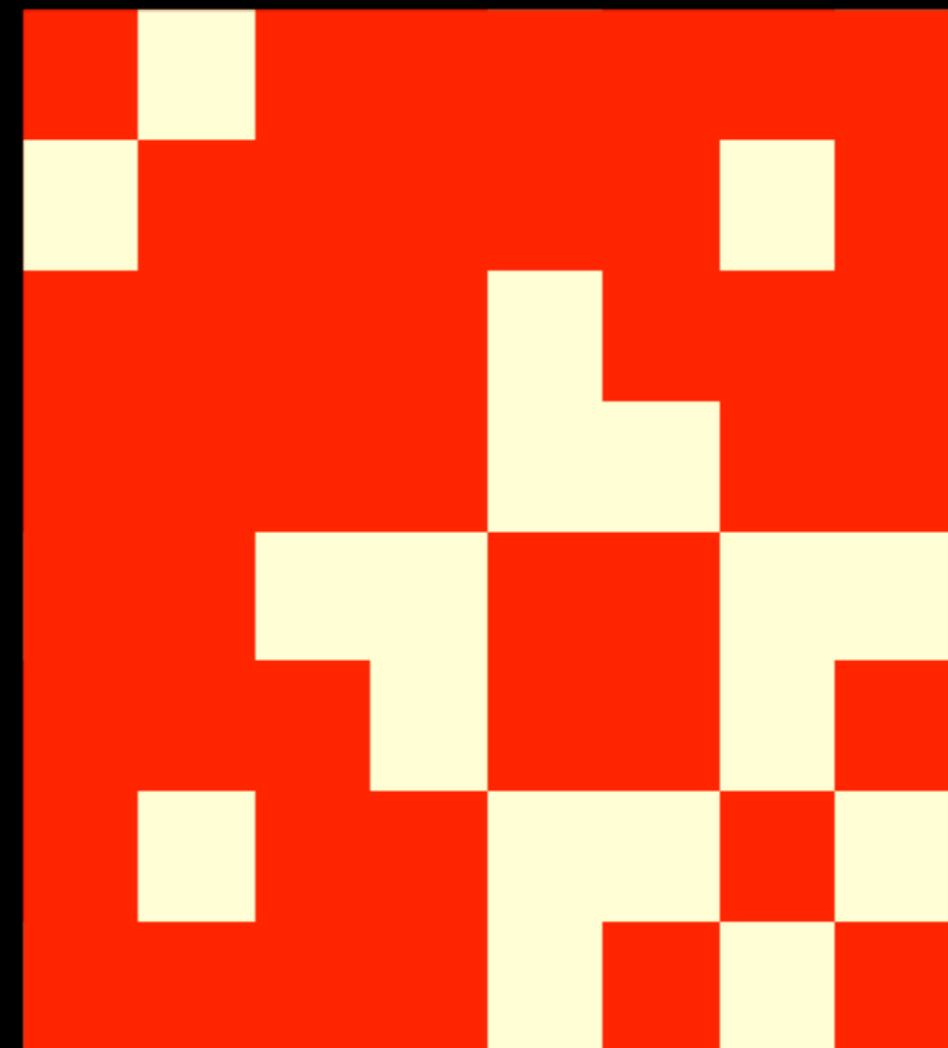
- $uv$  means there is an edge between  $u$  &  $v$
- e.g., in this graph  $AB$ , but not  $AE$



same graph  
(2D layout)

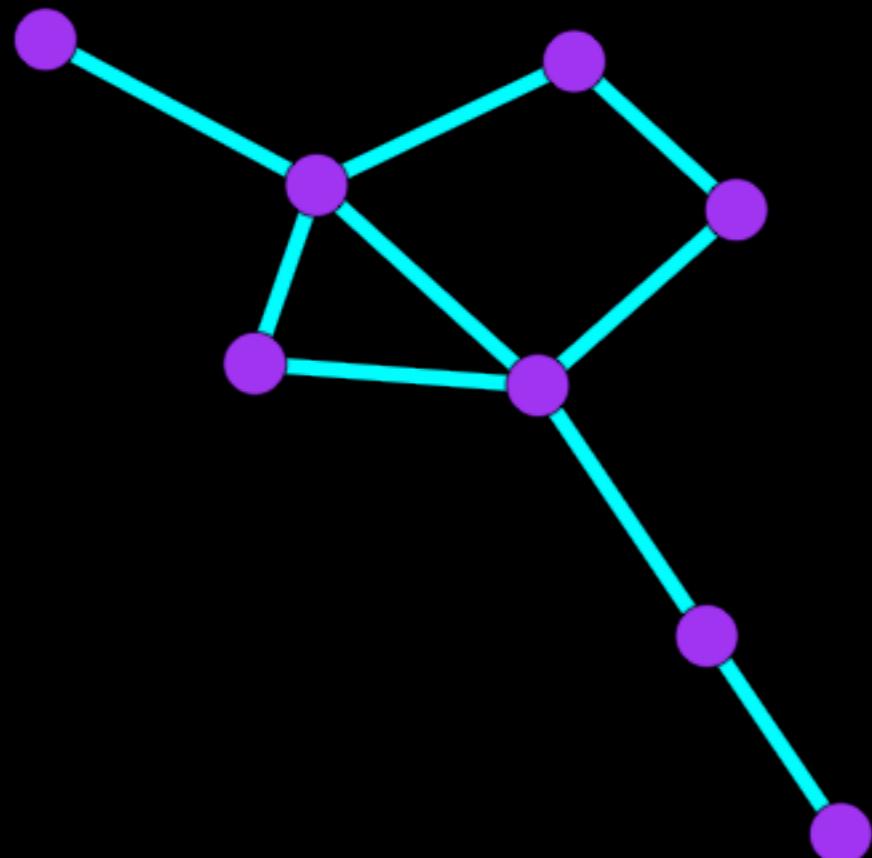


same graph  
(adjacency matrix)

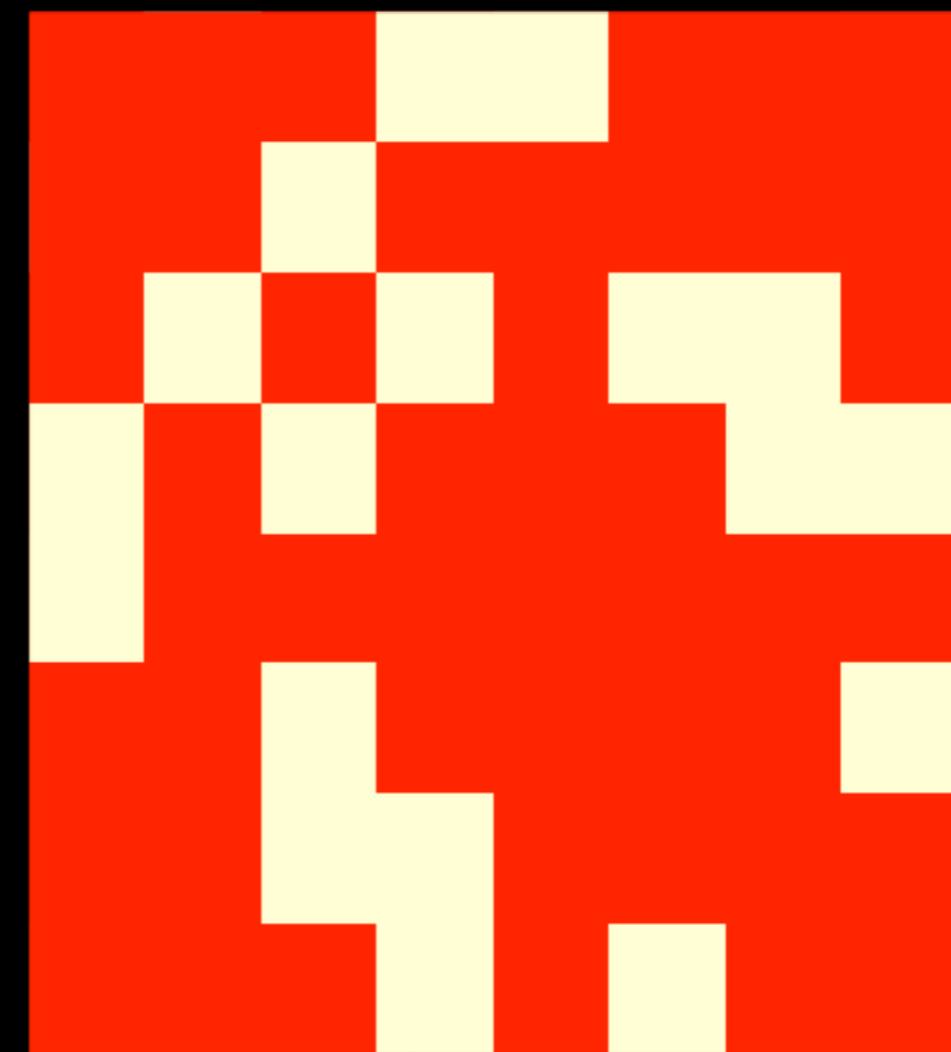


- $A$  is the adjacency matrix
- $A(u,v) = 1$  iff  $u \sim v$

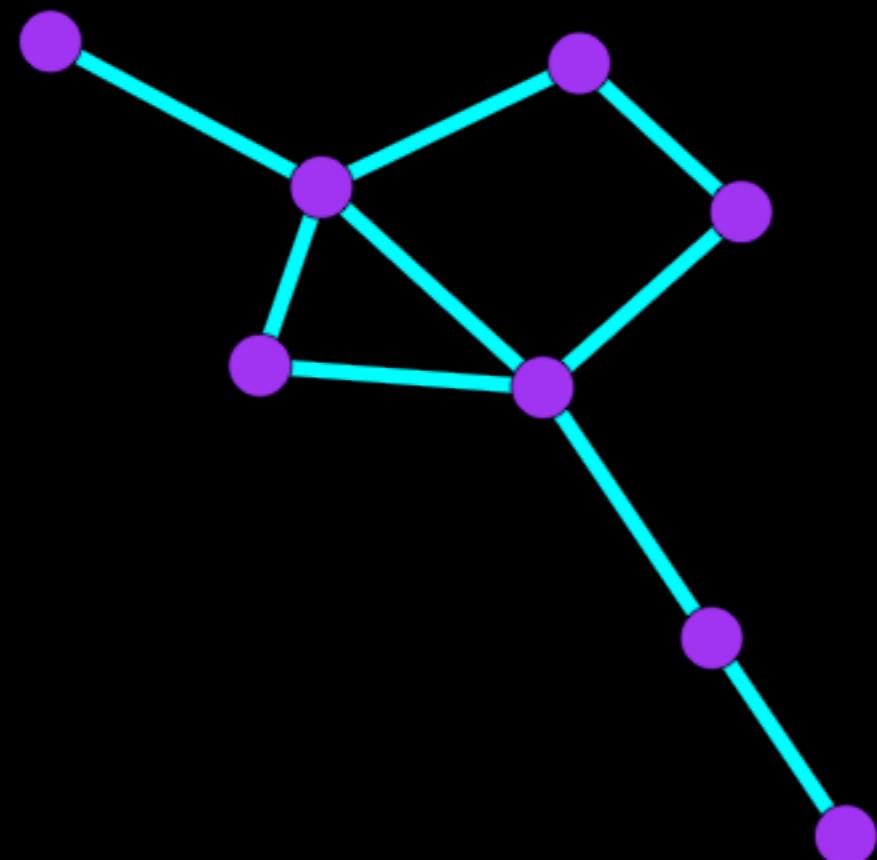
same graph  
(2D layout)



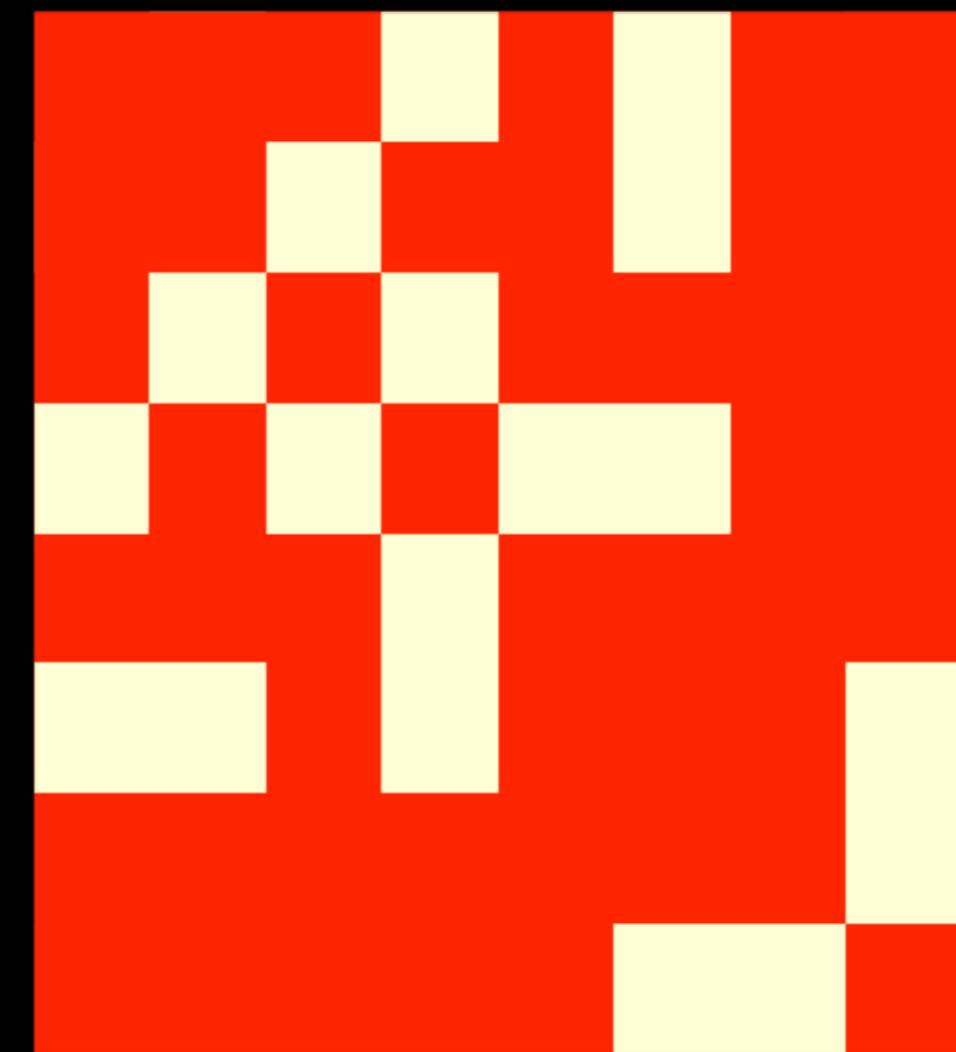
same graph  
(adjacency matrix)



same graph  
(2D layout)



same graph  
(adjacency matrix)



# Look at it?

- *c. elegans*:  $\sim 10^2$
- larval drosophila:  $\sim 10^4$
- larval zebrafish, adult drosophila:  $\sim 10^5$
- DTI-derived human connectome:  $\sim 10^6$
- adult zebrafish, mouse:  $\sim 10^7$
- monkey:  $\sim 10^9$
- human:  $\sim 10^{11}$

# Simplest Network Code

- Each **edge** is independent
- Probability of a edge between any pair of vertices is p
- $P[a_{uv}] = \text{Bernoulli}(p)$

# Keep it Simple

Neural Encoding:  $P[r | s]$

Neural Decoding:  $P[s | r]$

Neural Code:  $P[s,r]$

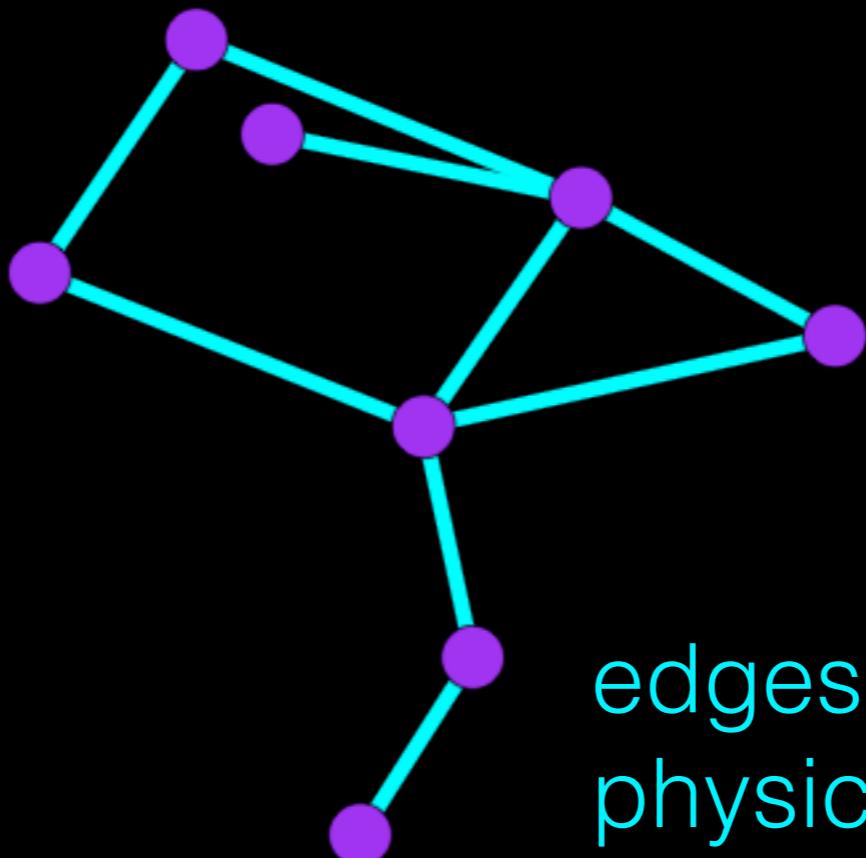
Connectome Encoding:  $P[g | s]$

Connectome Decoding:  $P[s | g]$

Connectome Code:  $P[s,g]$

**What's a connectome?**

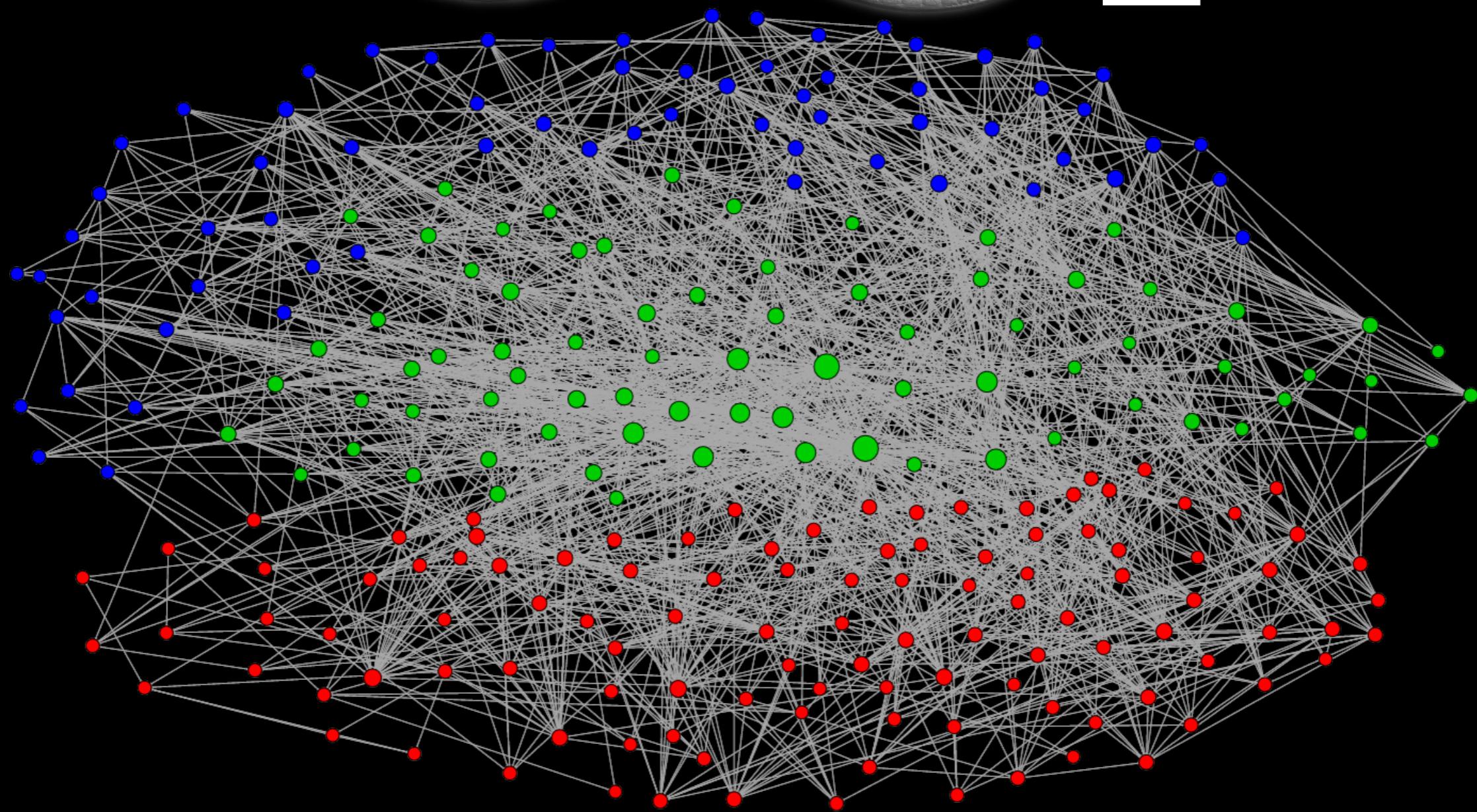
# connectome = brain-graph



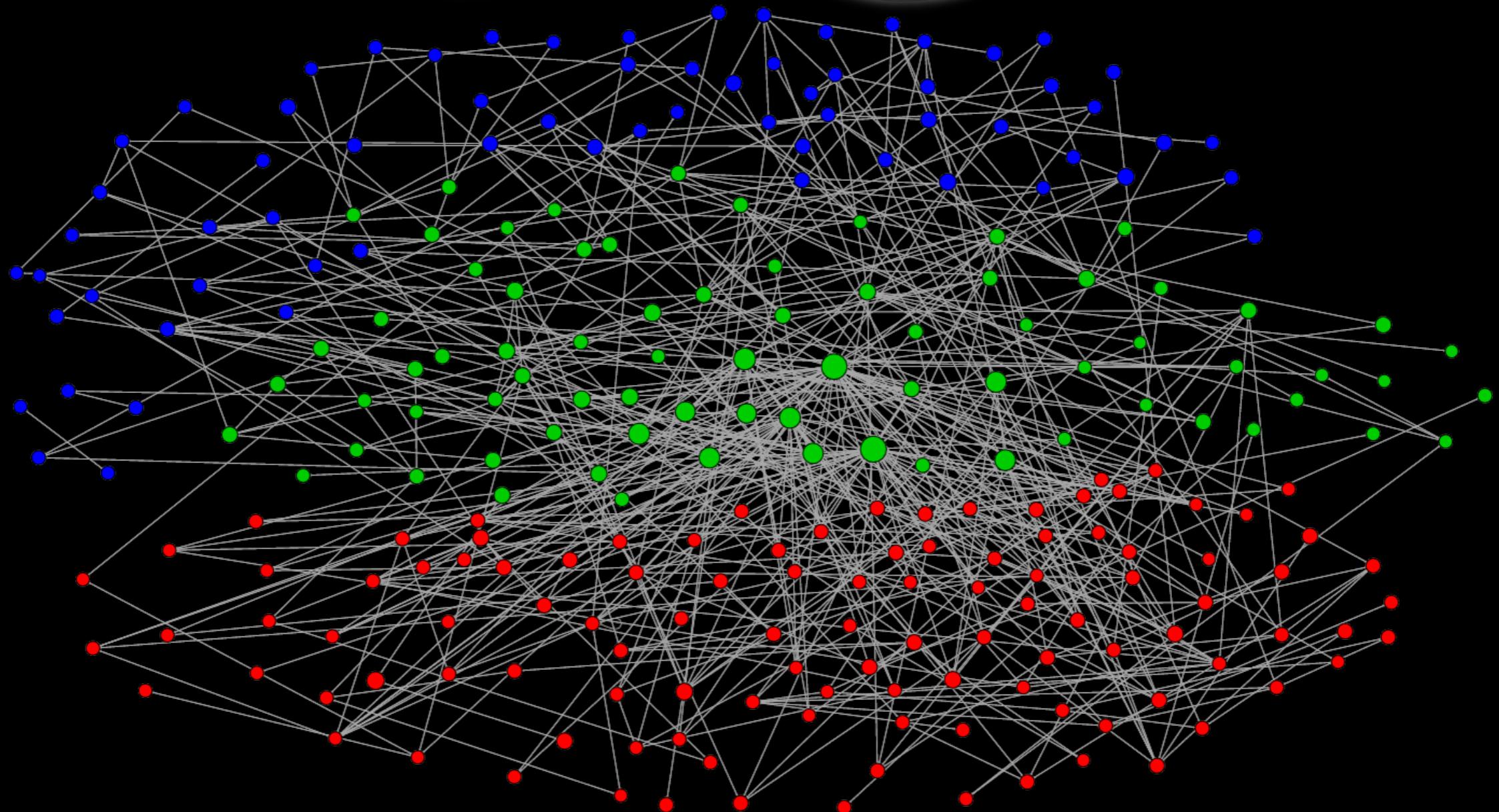
vertices:  
brain regions  
or voxels  
or electrodes  
or neurons  
or even compartments

edges:  
physical connections  
or functional connections  
or correlations

*c. elegans*



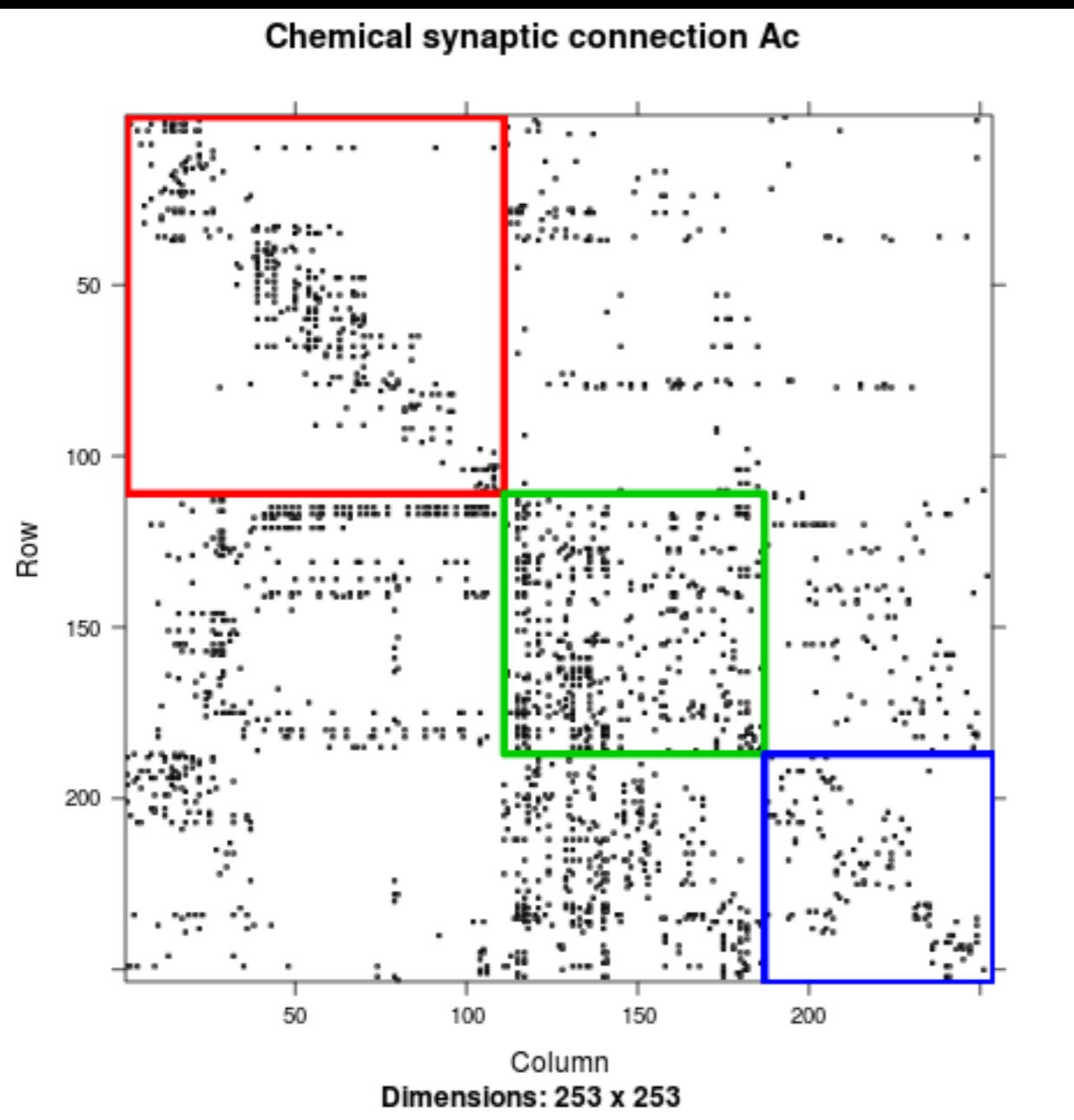
*c. elegans*



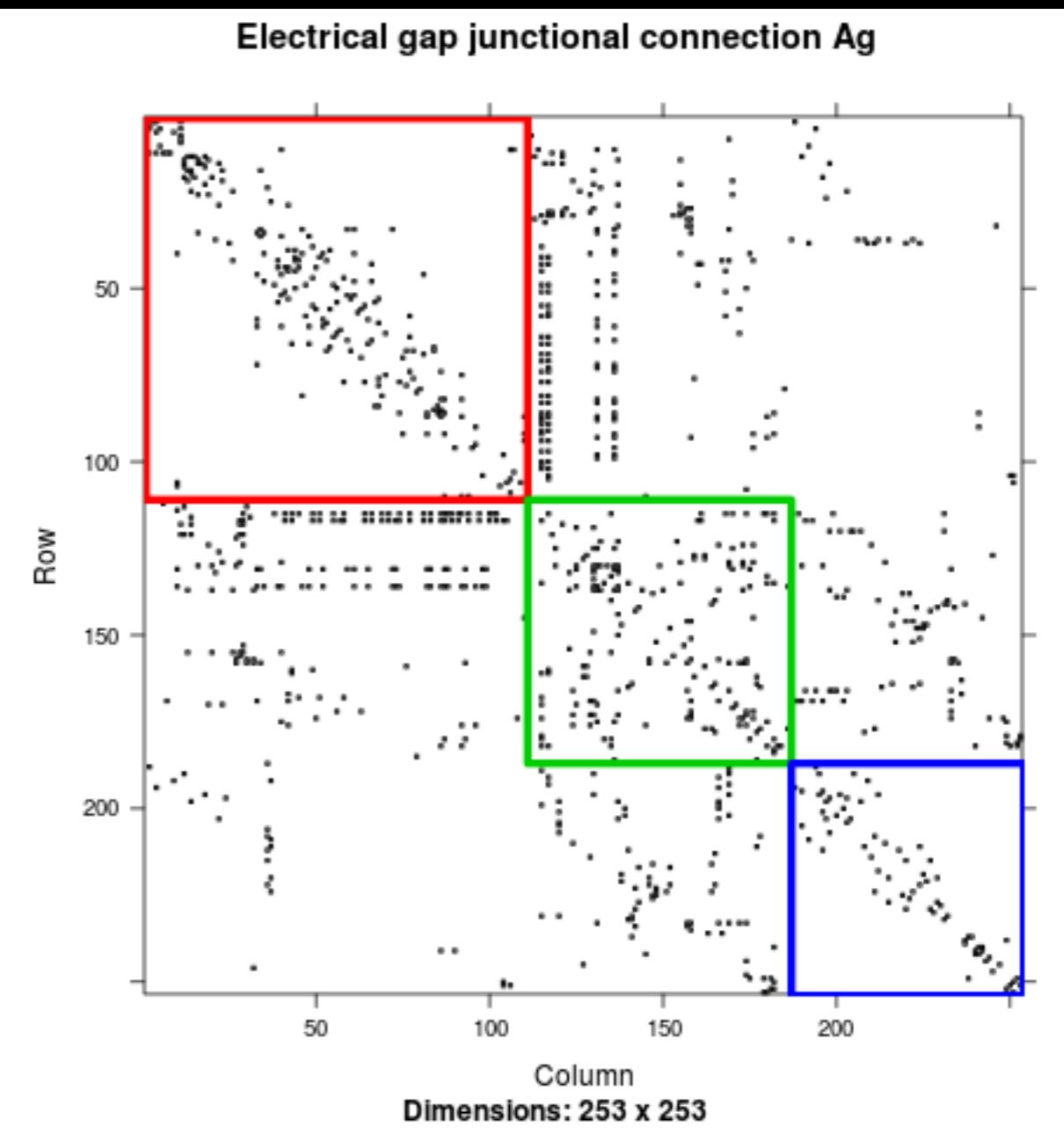


*c. elegans*

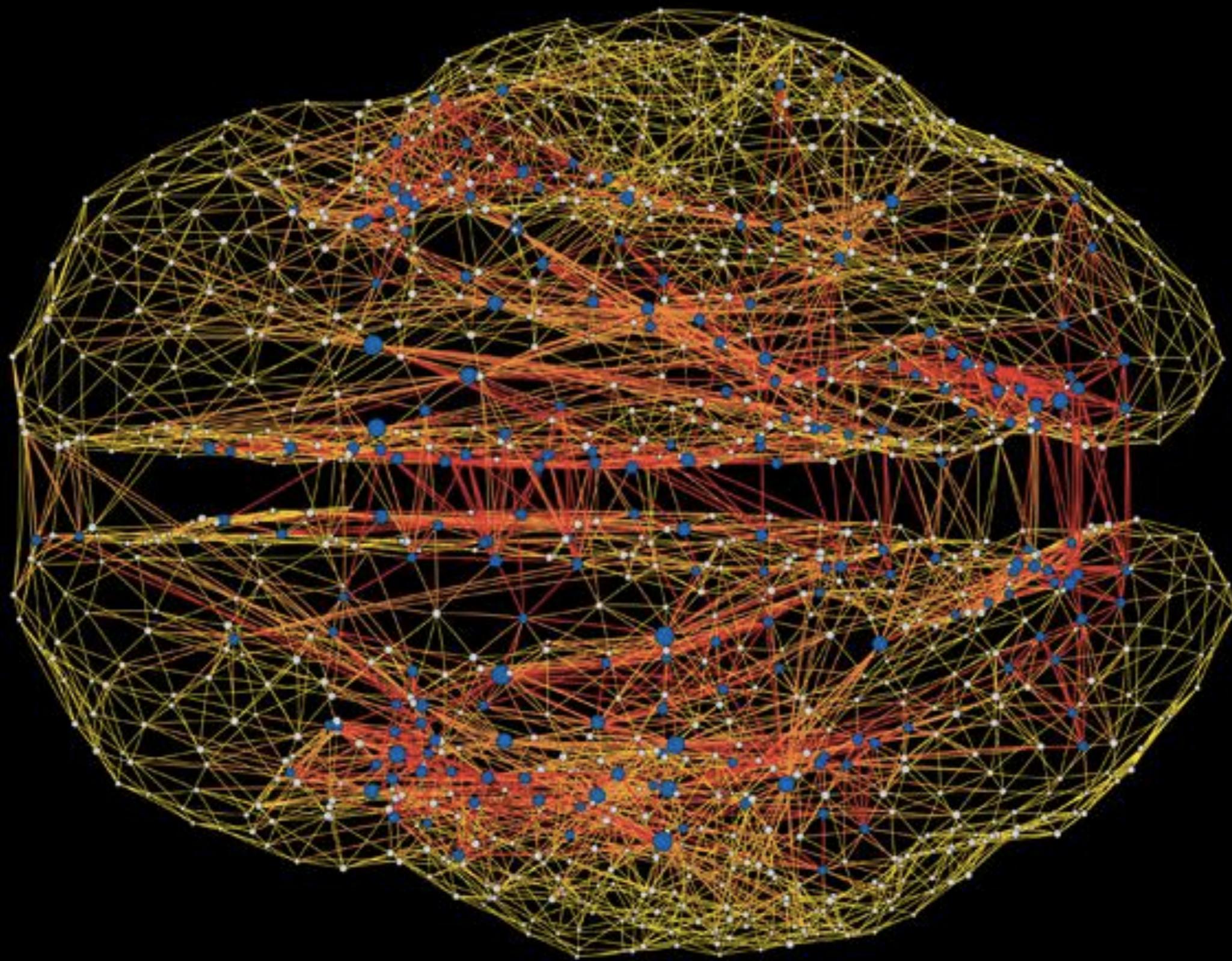
Chemical synaptic connection Ac



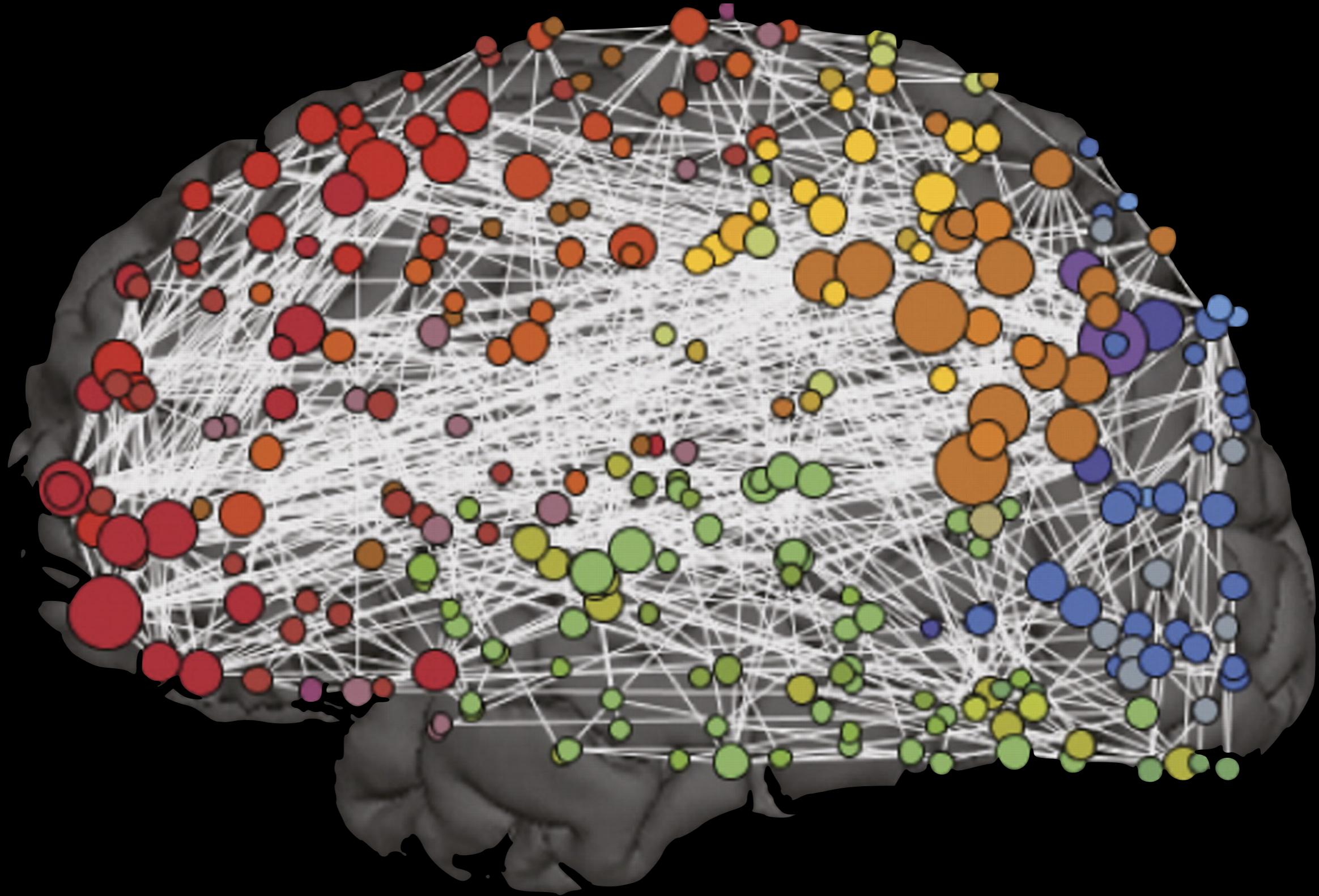
Electrical gap junctional connection Ag



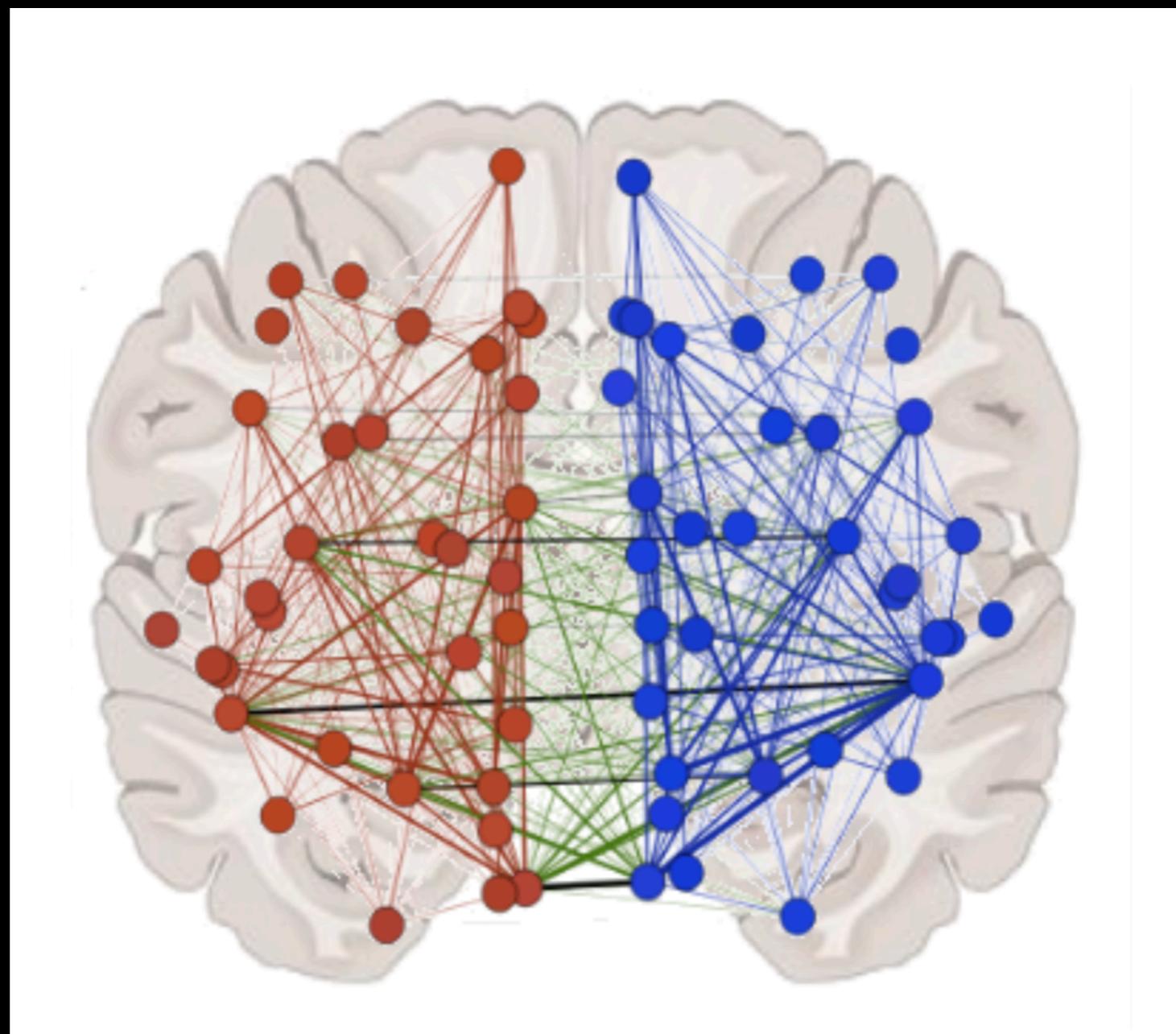
# human connectome



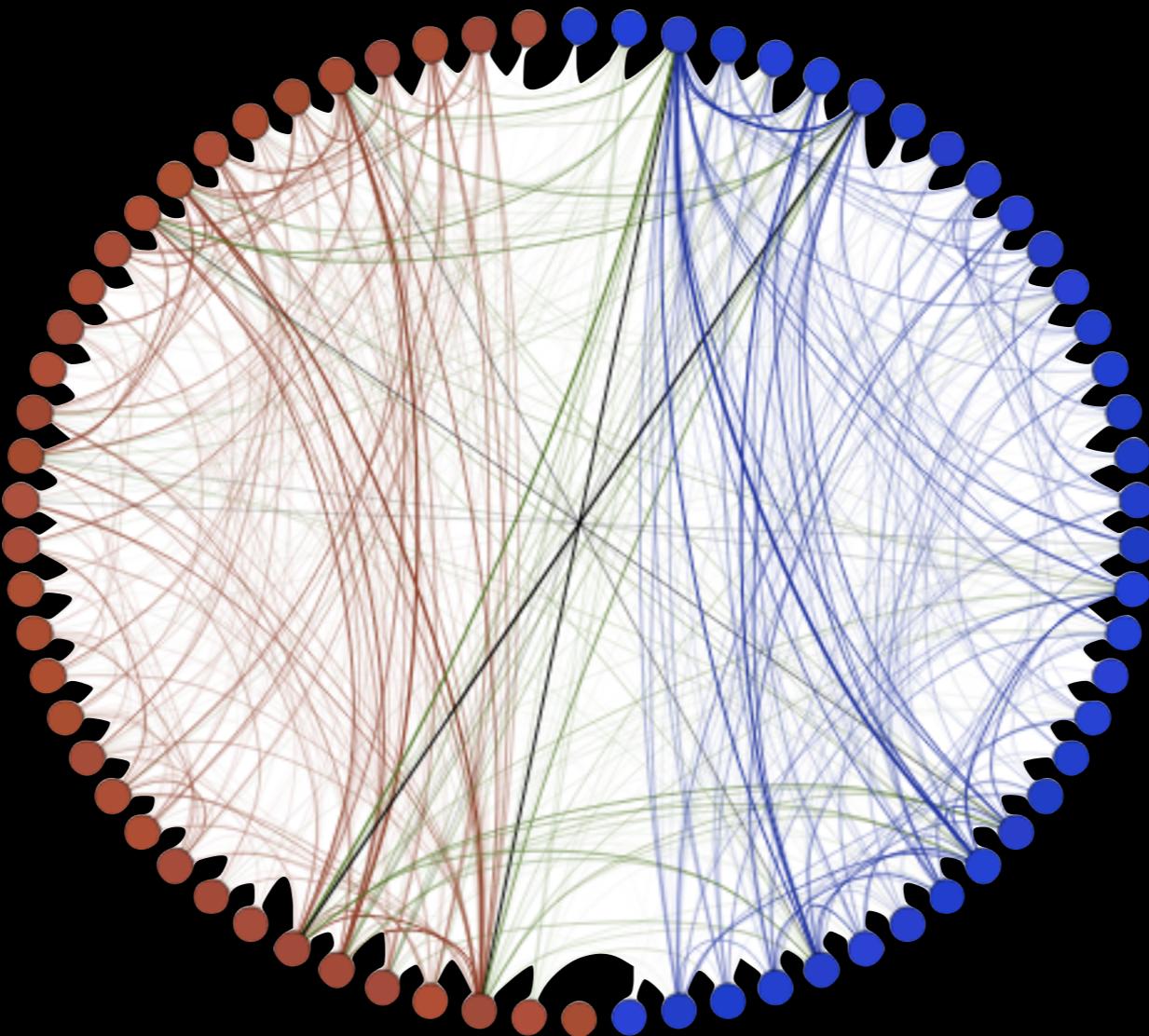
# human connectome



# human connectome



# human connectome



# Simplest Connectome Code

- Each **edge** is independent
- Probability of a edge between any pair of vertices is  $p$
- $P[a_{uv}] = \text{Bernoulli}(p)$

$$P[r|s]$$

Let's do it for the conditional response

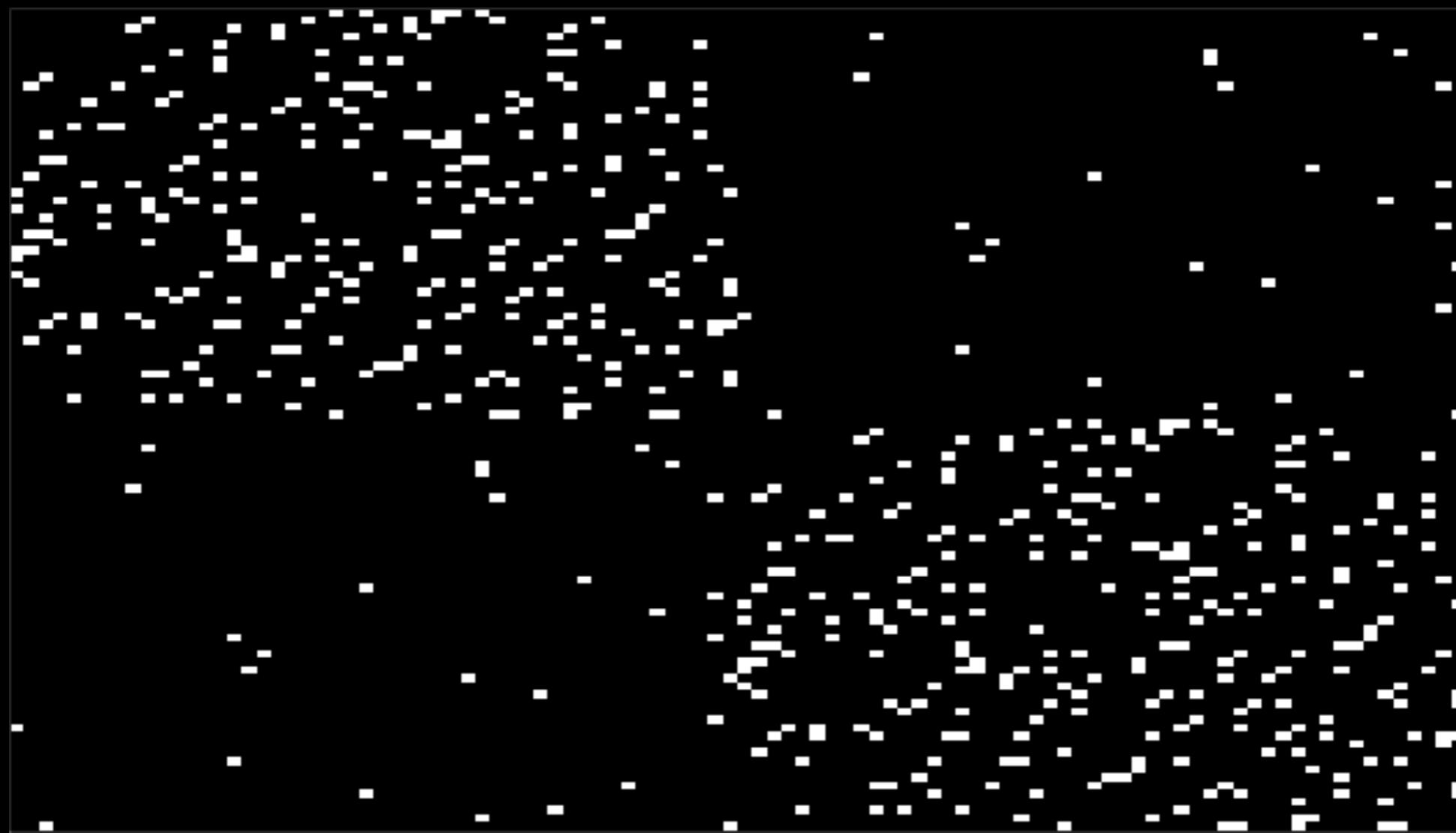


# Simplest Conditional Neural Code

- Each spike is independent
- Probability of a spike at any time is  $\lambda_i$
- $P[r|s] = \text{Poisson}(\lambda_s)$

$$P[g|s]$$

Let's do it for the conditional connectome

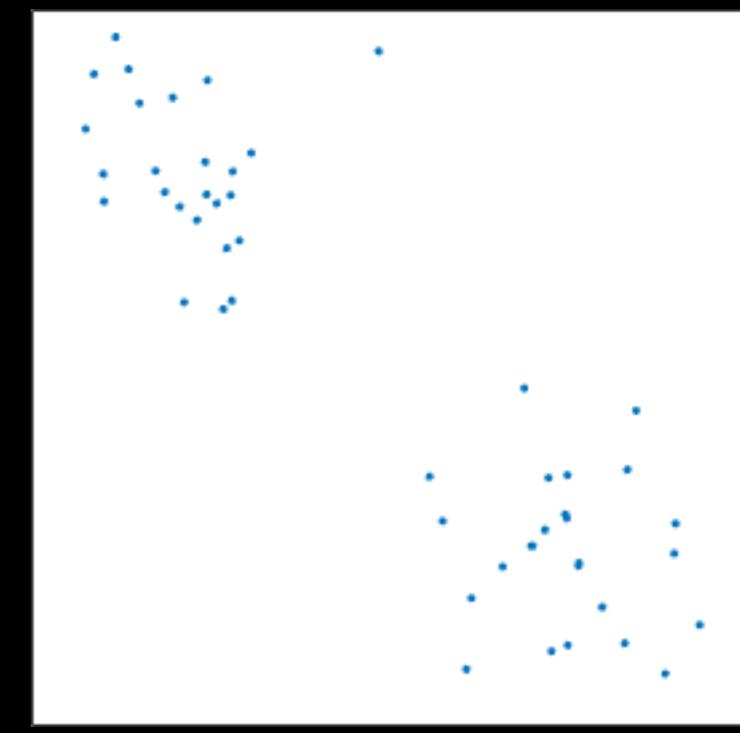
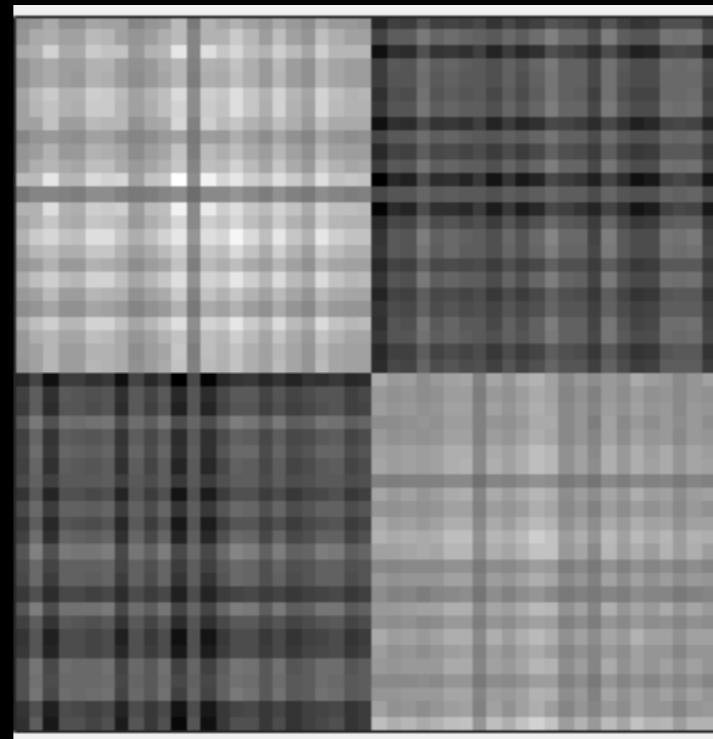


# Simplest Conditional Connectome Code

- Each edge is independent
- $\Pr[\text{edge} \mid \text{within a hemisphere}] = p_w$
- $\Pr[\text{edge} \mid \text{between hemisphere}] = p_b$
- $B = [p_w \ p_b; p_b \ p_w]$
- $P[g|s] = \text{SBM}(B \mid \tau)$

# Simple Conditional Connectome Code

- Each edge is independent
- $\Pr[ u \sim v \mid x_u, x_v ] = \langle x_u, x_v \rangle$
- $P(u,v) = \langle x_u, x_v \rangle$
- $x_u \mid u \text{ in class } c = m_c$
- $P[g|s] = \text{SBM}(B \mid \tau)$



# Simple Conditional Connectome Code

- Each edge is independent
- $\Pr[ u \sim v \mid x_u, x_v ] = \langle x_u, x_v \rangle$
- $P(u,v) = \langle x_u, x_v \rangle$
- $x_u \mid u \text{ in class } c \sim N(m_c, I)$
- $P[g|s] = LPM(P)$

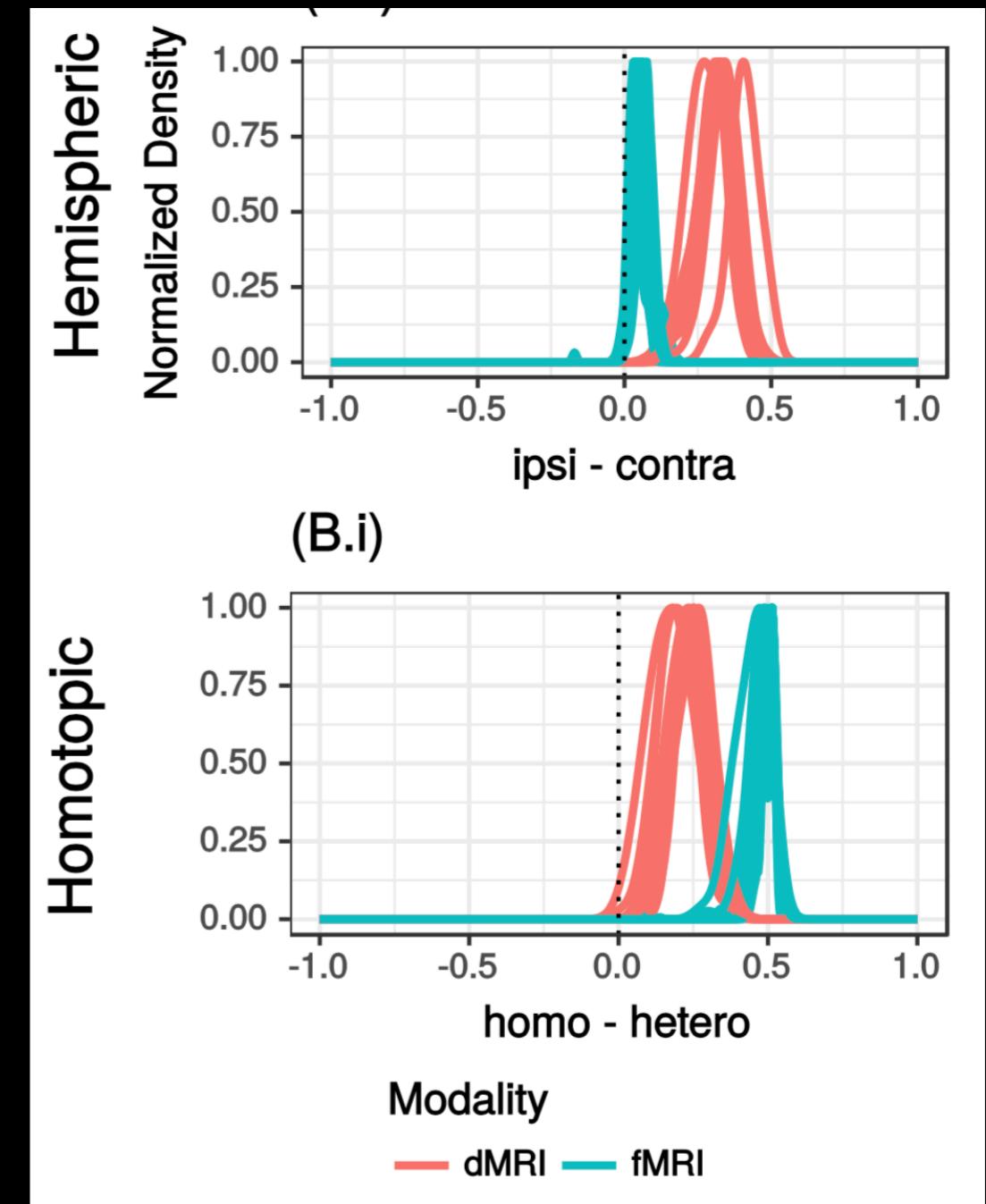
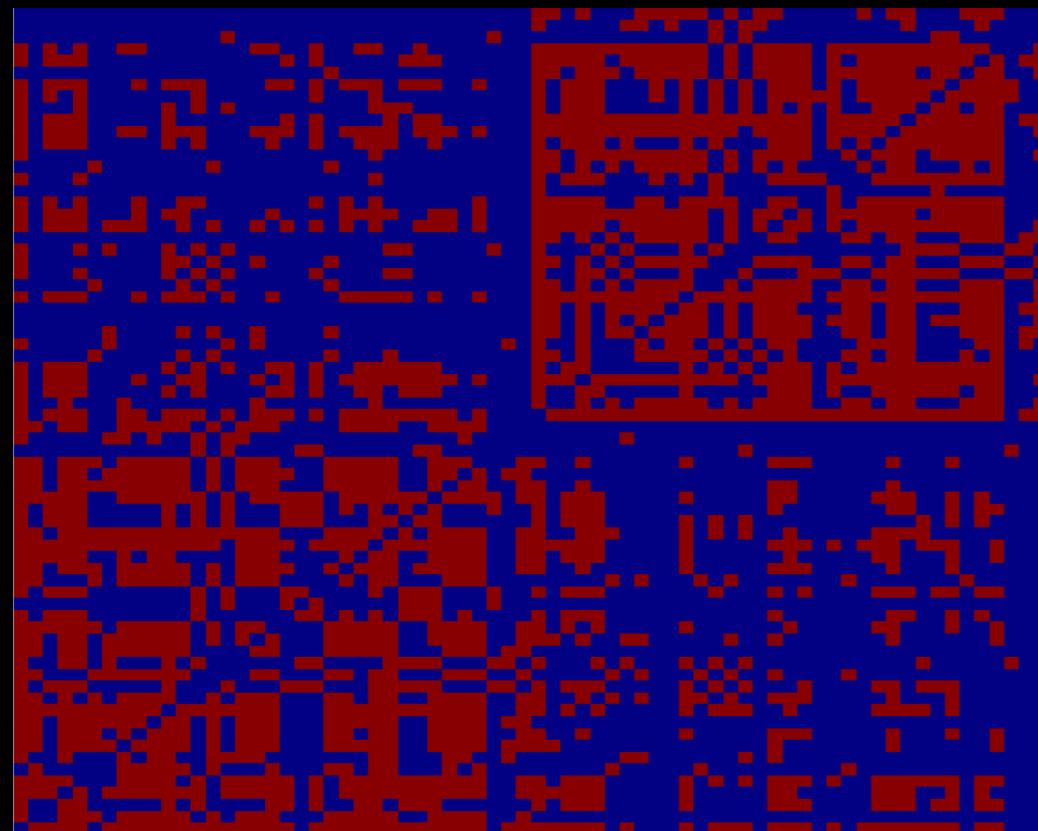
# Conditional Connectome Code

- Each edge is conditionally independent
- $\Pr[ u \sim v \mid \mathbf{x}_u, \mathbf{x}_v ] = \langle \mathbf{x}_u, \mathbf{x}_v \rangle$
- $P(u,v) = \langle \mathbf{x}_u, \mathbf{x}_v \rangle$
- $\mathbf{x}_u \mid u \text{ in class } c \sim F_c$
- $F = \{w_1 F_1, w_2 F_2, \dots, w_c F_c\}$
- $P[g|s] = \text{LSM}(F)$

# applications

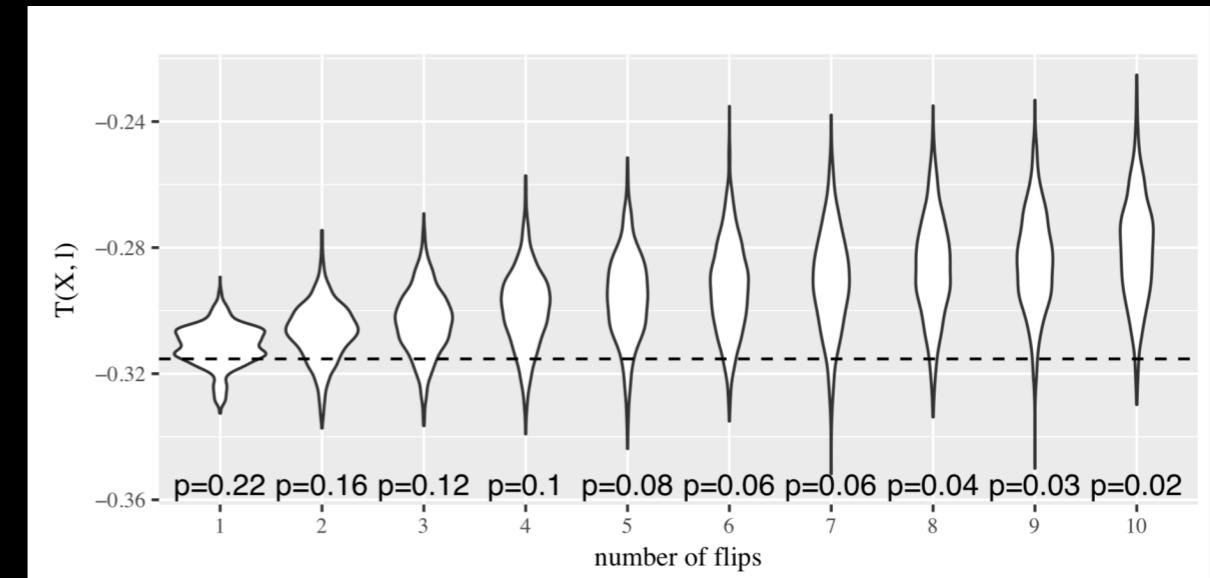
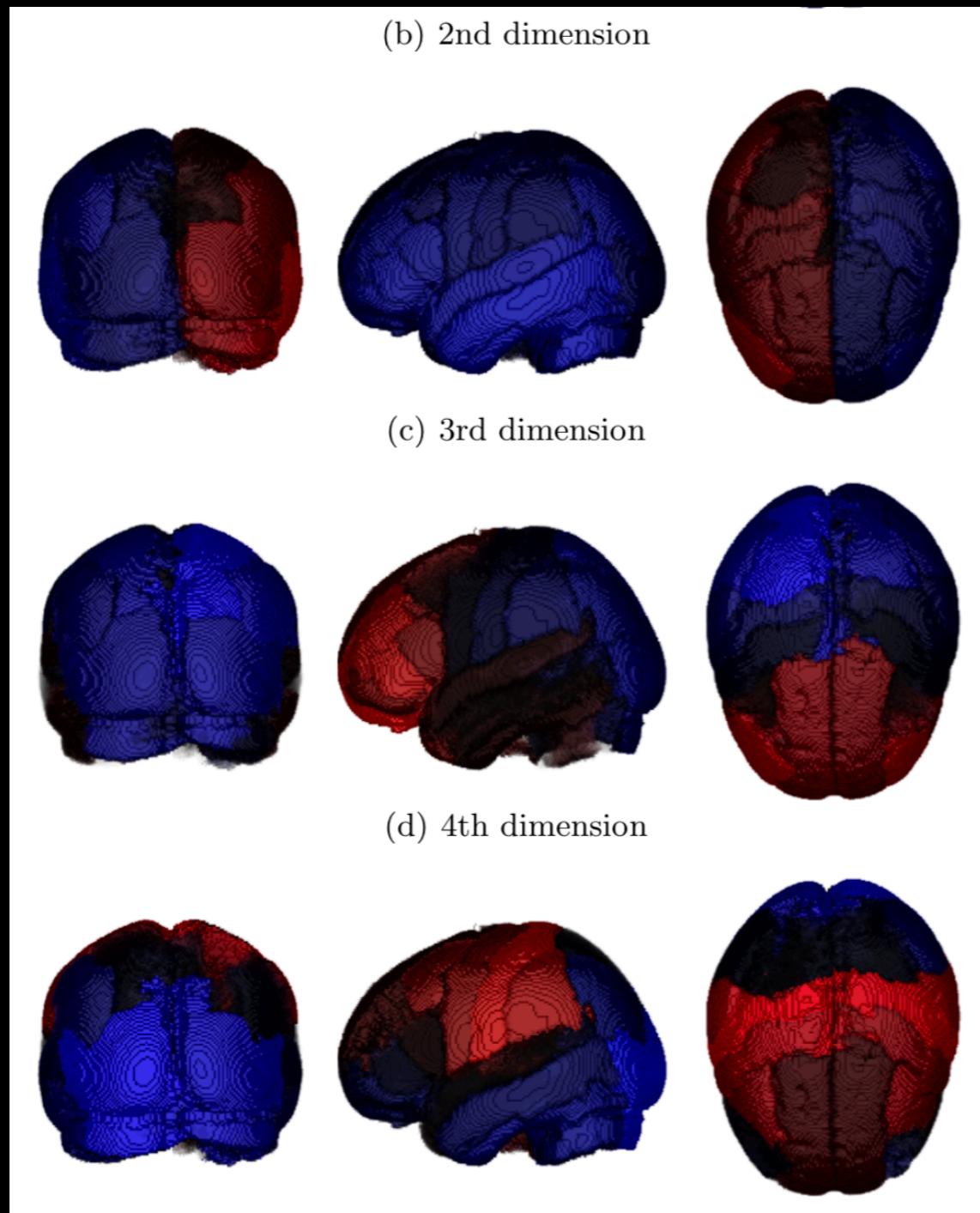
# Human Connectome Code

ipsilateral vs contralateral & homotopic vs heterotopic

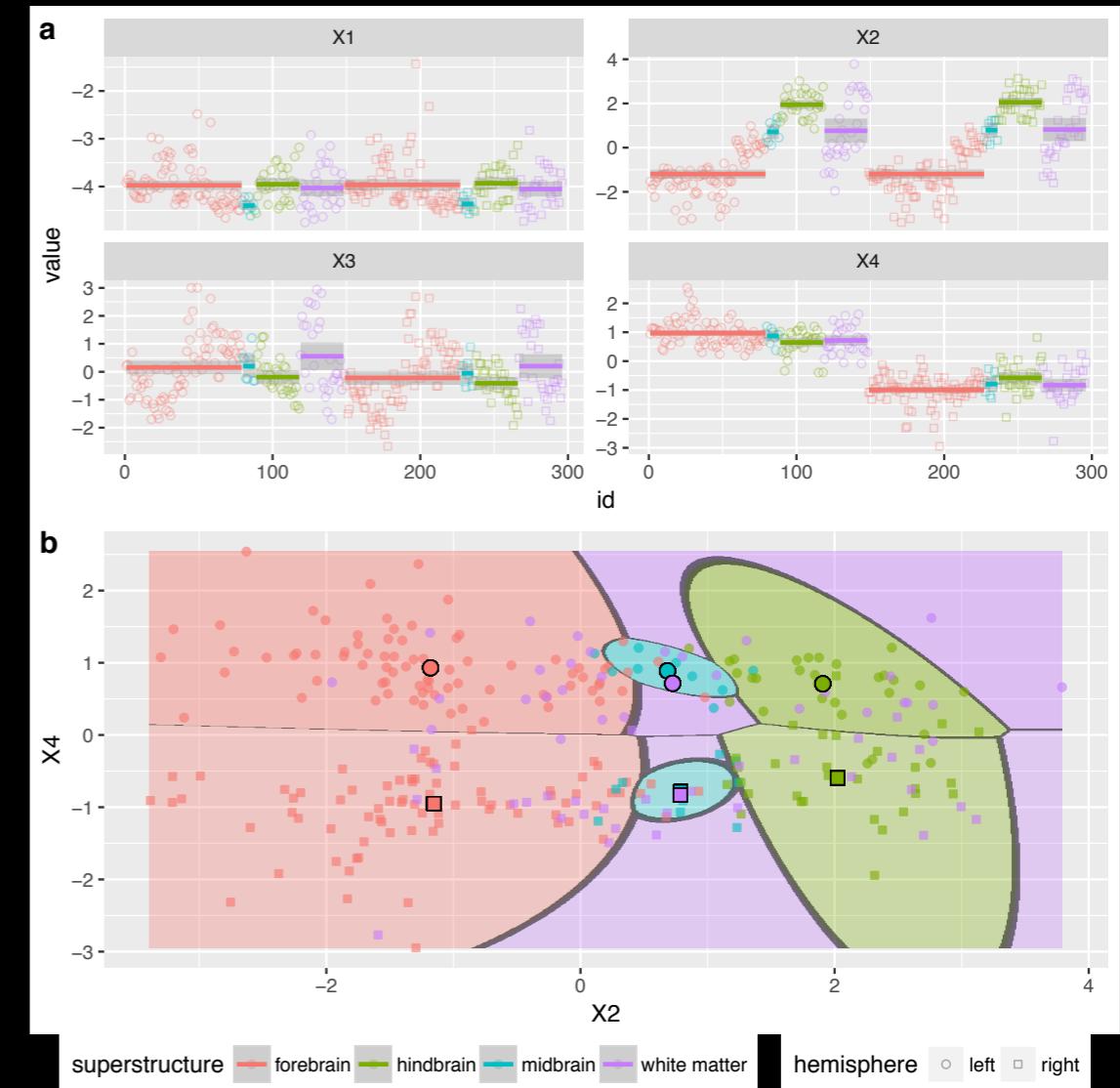
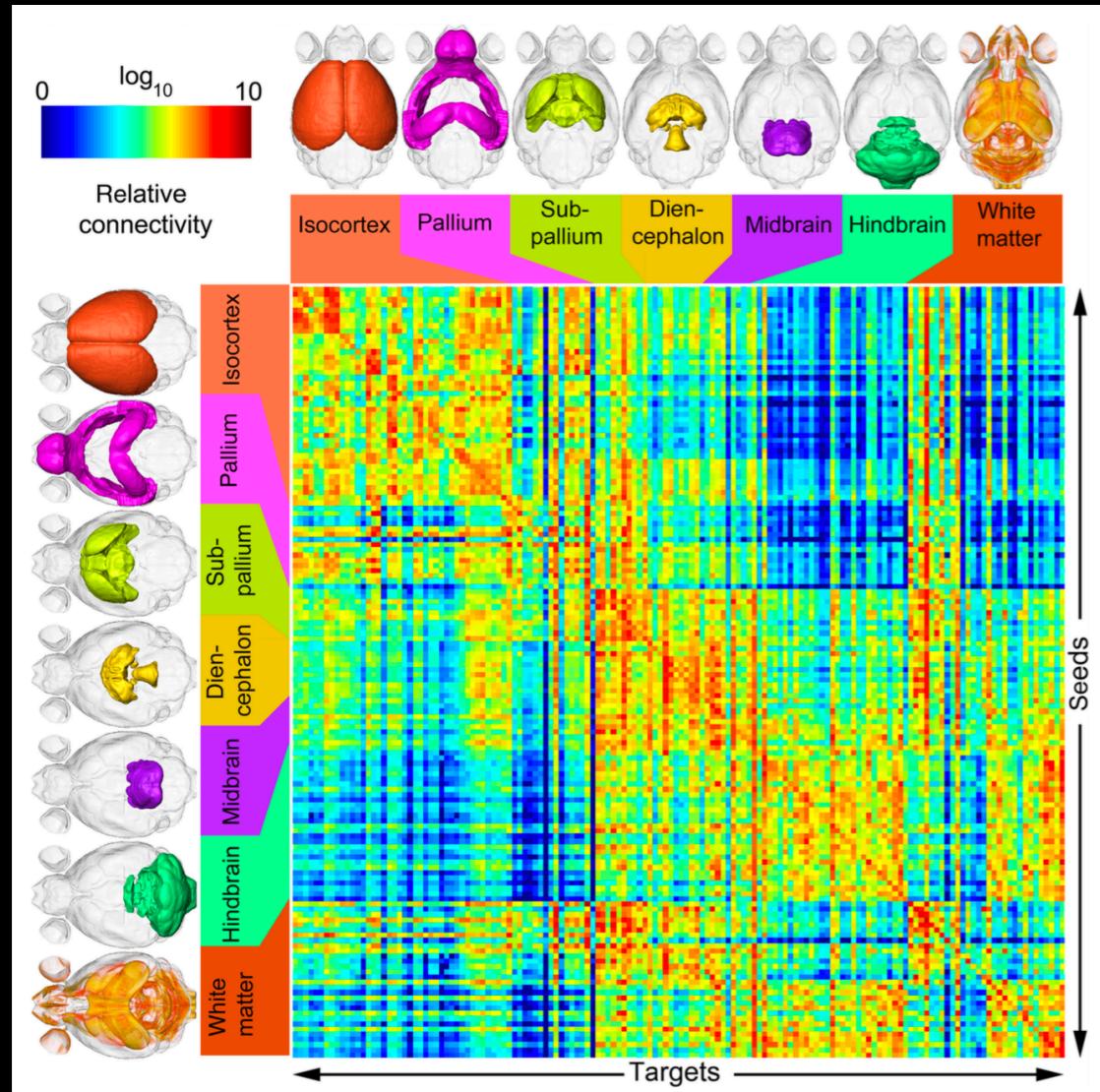


# Human Connectome Code

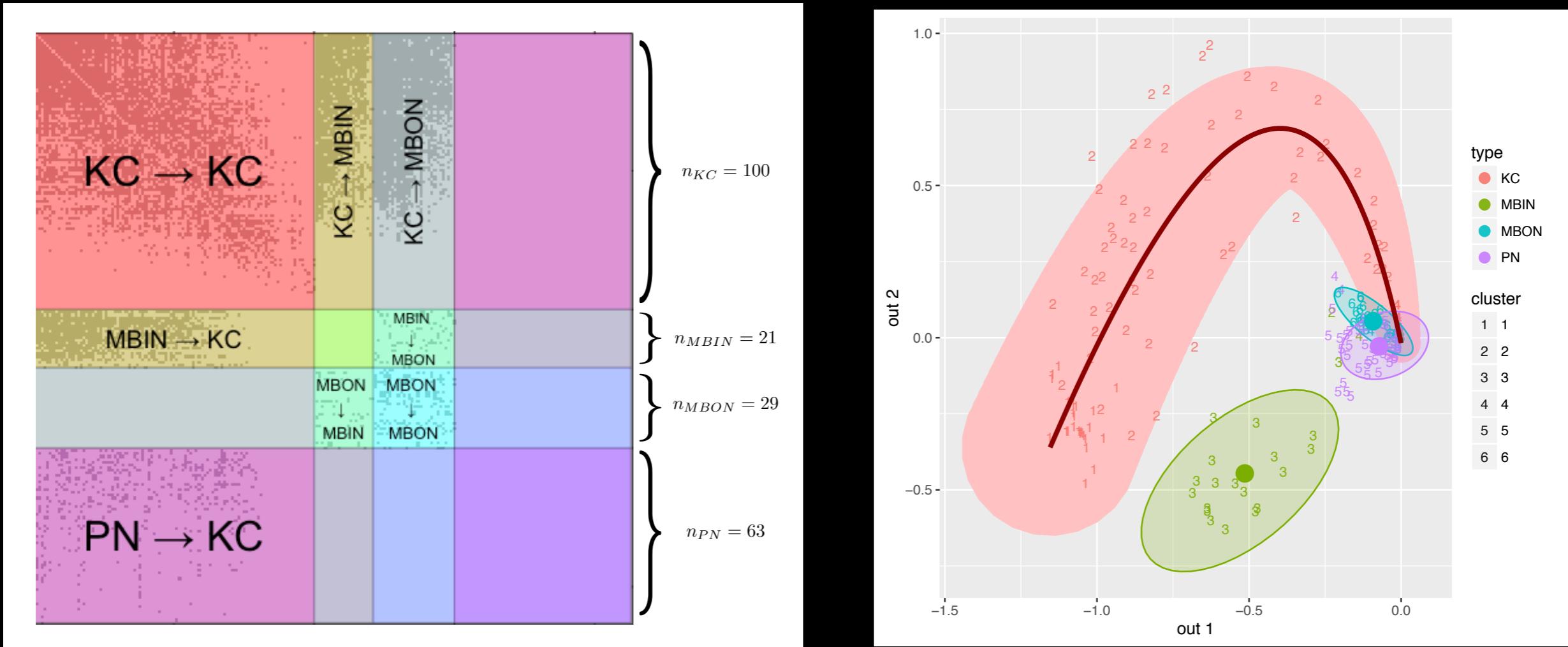
eigenconnectomes = lobes



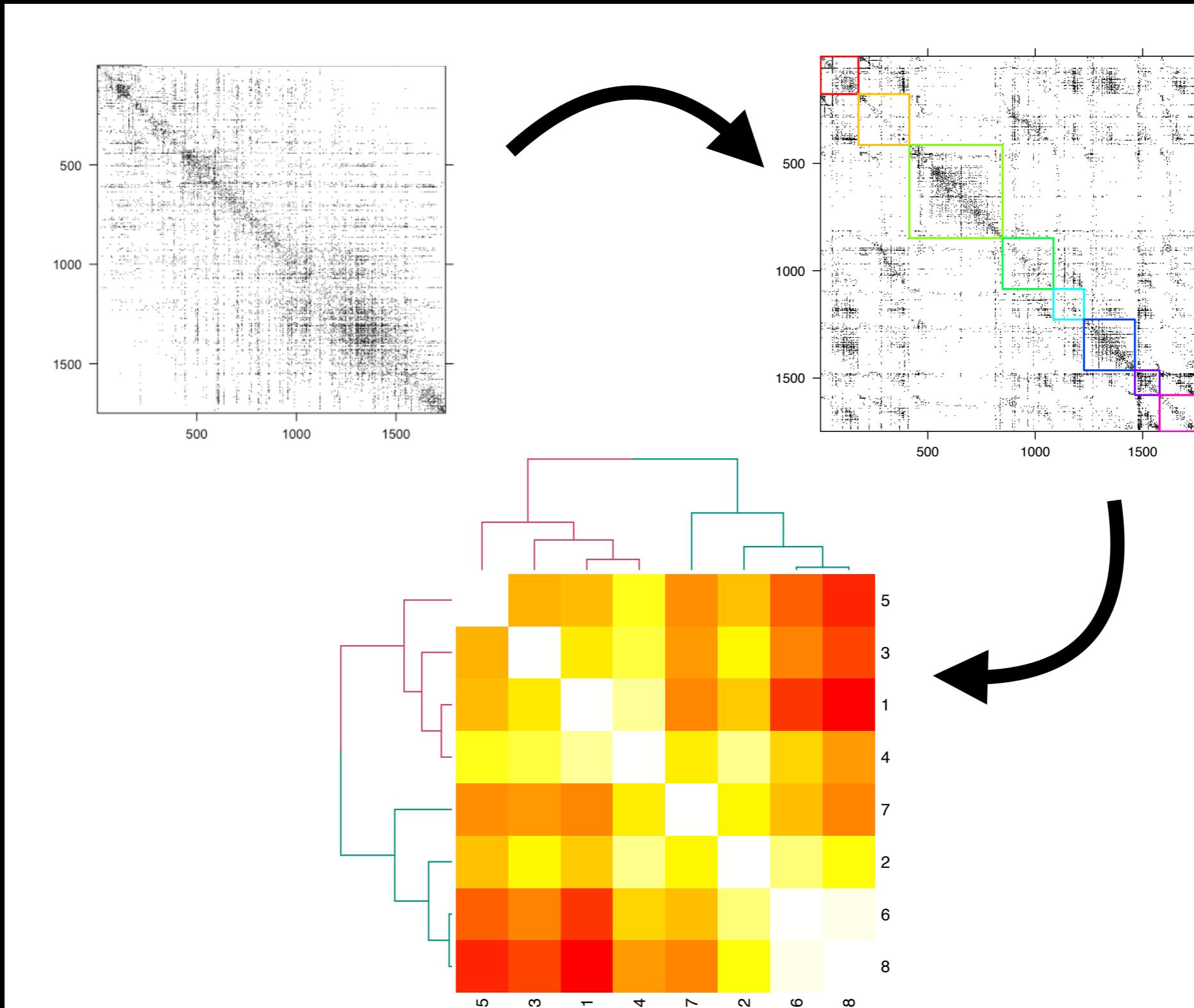
# Mouse Connectome Code



# Larval Drosophila Mushroom Body Connectome Code



# Drosophila Optic Medulla Connectome Code



# References (on arXiv)

1. [Statistical inference on random dot product graphs: a survey](#)
2. [Law of Large Graphs for Statistical Connectomics](#)
3. [A High-Throughput Pipeline Identifies Robust Connectomes But Troublesome Variability](#)
4. [Community Detection and Classification in Hierarchical Stochastic Blockmodels](#)

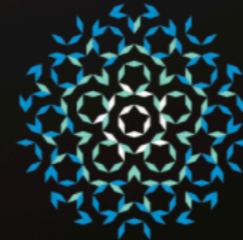
# Questions?

Joshua T. Vogelstein, BME@JHU

Co-founder: NeuroData Lab, Gigantum

Carey Priebe, Randal Burns, Michael Miller, Brian Caffo, Michael Milham, Mitya Chklovskii, R. Jacob Vogelstein, Davi Bock, Cencheng Shen, Minh Tang, Avant Athreya, Youngser Park, Greg Kiar, Eric Bridgeford, Alexandre Badea, Daniel Sussman, Vince Lyzinski, Runze Tang, yummy, lion

## Now Hiring!



Kavli  
NDI



$$\Pr[\mathbf{A}|\mathbf{X}] = \prod_{i < j} (\mathbf{X}_i^\top \mathbf{X}_j)^{\mathbf{A}_{ij}} (1 - \mathbf{X}_i^\top \mathbf{X}_j)^{1 - \mathbf{A}_{ij}}$$

$$\Pr[\mathbf{A}|\mathbf{X}] = \prod_{i < j} \mathbf{P}_{ij}^{\mathbf{A}_{ij}} (1 - \mathbf{P}_{ij})^{1 - \mathbf{A}_{ij}}$$

$$\mathbf{A}\sim\text{SBM}(\tau,\{\mathbf{X}_{(i)}\}_{i=1}^K)$$

$$\mathbf{A}\sim\text{SBM}(\pi,\{X_{(i)}\}).$$

$$\mathbf{P}_{ij}=\mathbf{B}_{\tau(i),\tau(j)},$$

$$\mathbb{P}\left(\|\mathbf{A}-\mathbf{P}\| \leq 2\sqrt{\delta(\mathbf{P})} + C\delta^{1/4}(\mathbf{P})\ln n\right) \geq 1-n^{-c}.$$

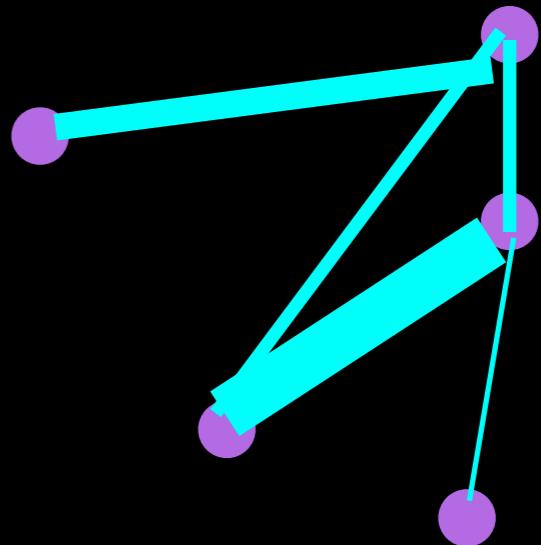
$$\max_i \|(\hat{\mathbf{X}}_n)_i - \mathbf{W}_n(\mathbf{X}_n)_i\| \leq \frac{Cd^{1/2}\log^2 n}{\delta^{1/2}(\mathbf{P}_n)}$$

$$\lim_{n\rightarrow\infty}\Pr\left[n^{1/2}\left(\hat{\mathbf{X}}_n\mathbf{W}_n-\mathbf{X}_n\right)_i\leq z\right]=\int_{\mathrm{supp}\, F}\Phi\left(z,\Sigma(\mathbf{x})\right)dF(\mathbf{x}),$$

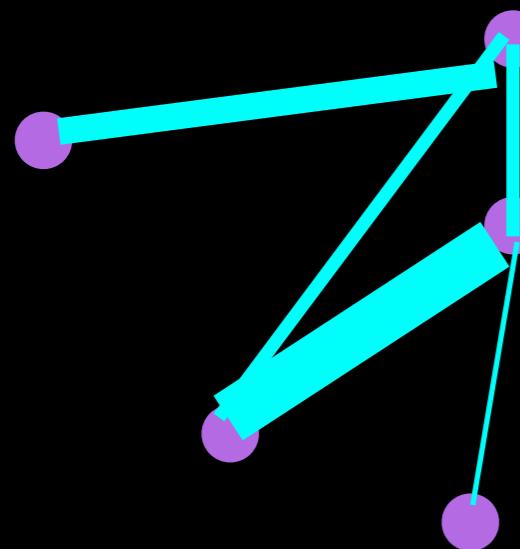
let's do it for a little bit more  
complex models

# weighted graph

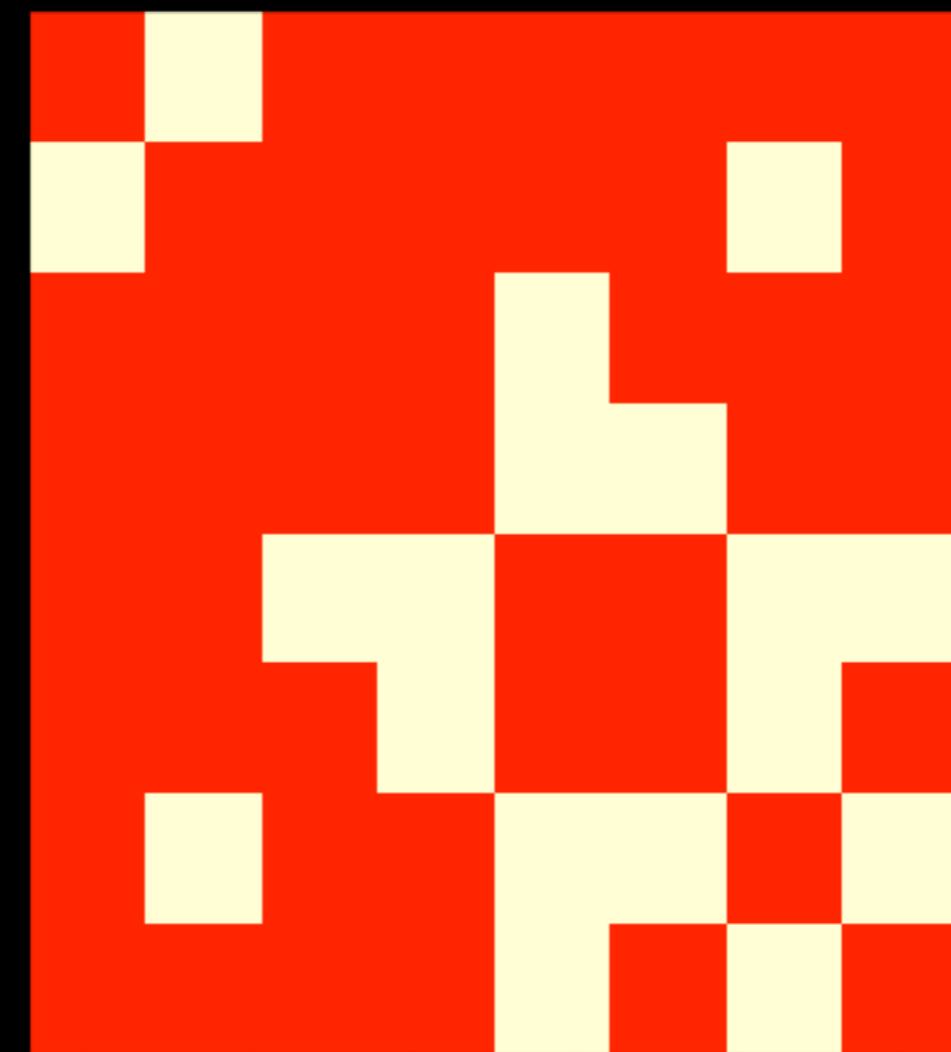
- each edge has a weight
- weights can be positive,  
negative, integer, zero, etc.



weighted graph  
(2D layout)

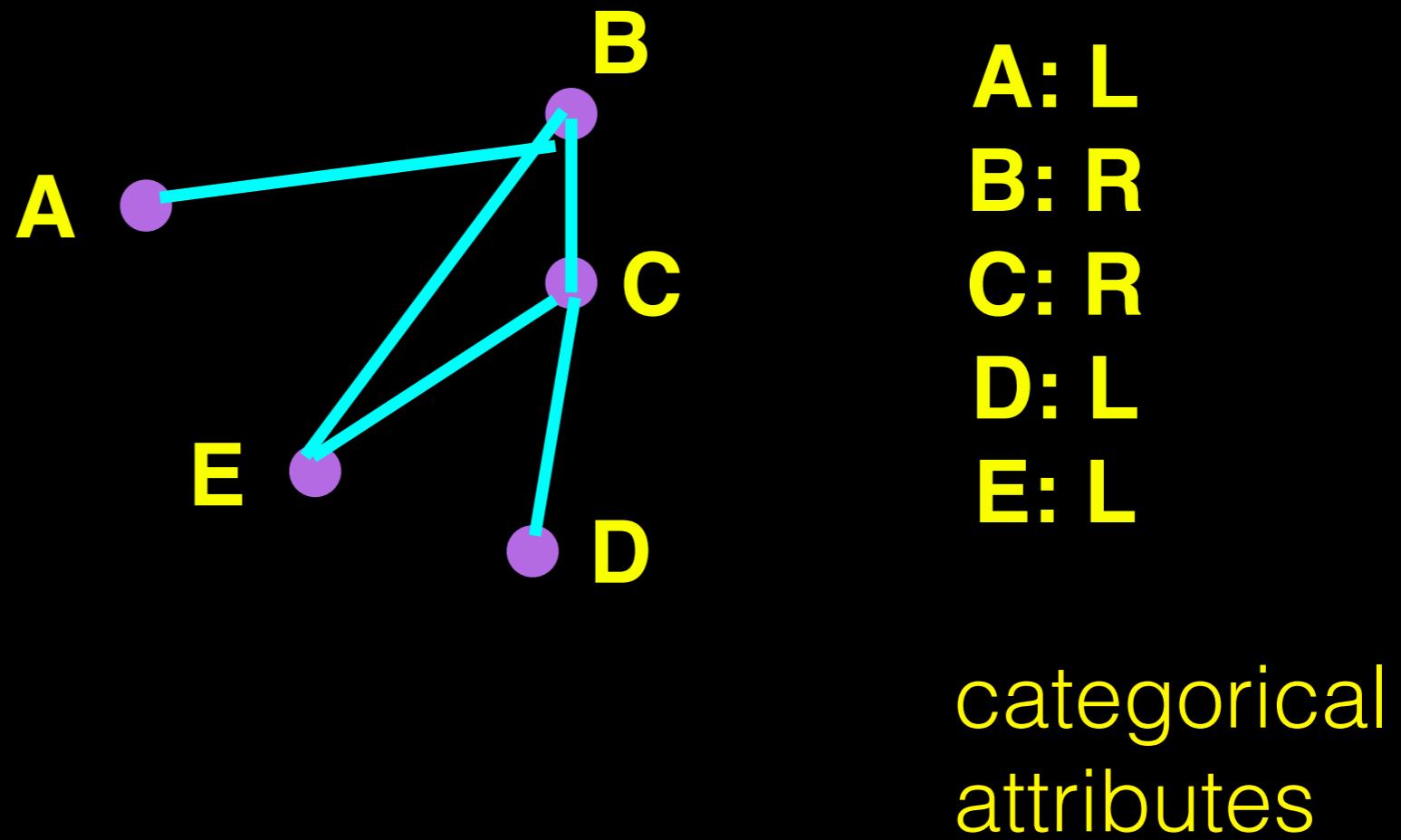


weighted graph  
(adjacency matrix)

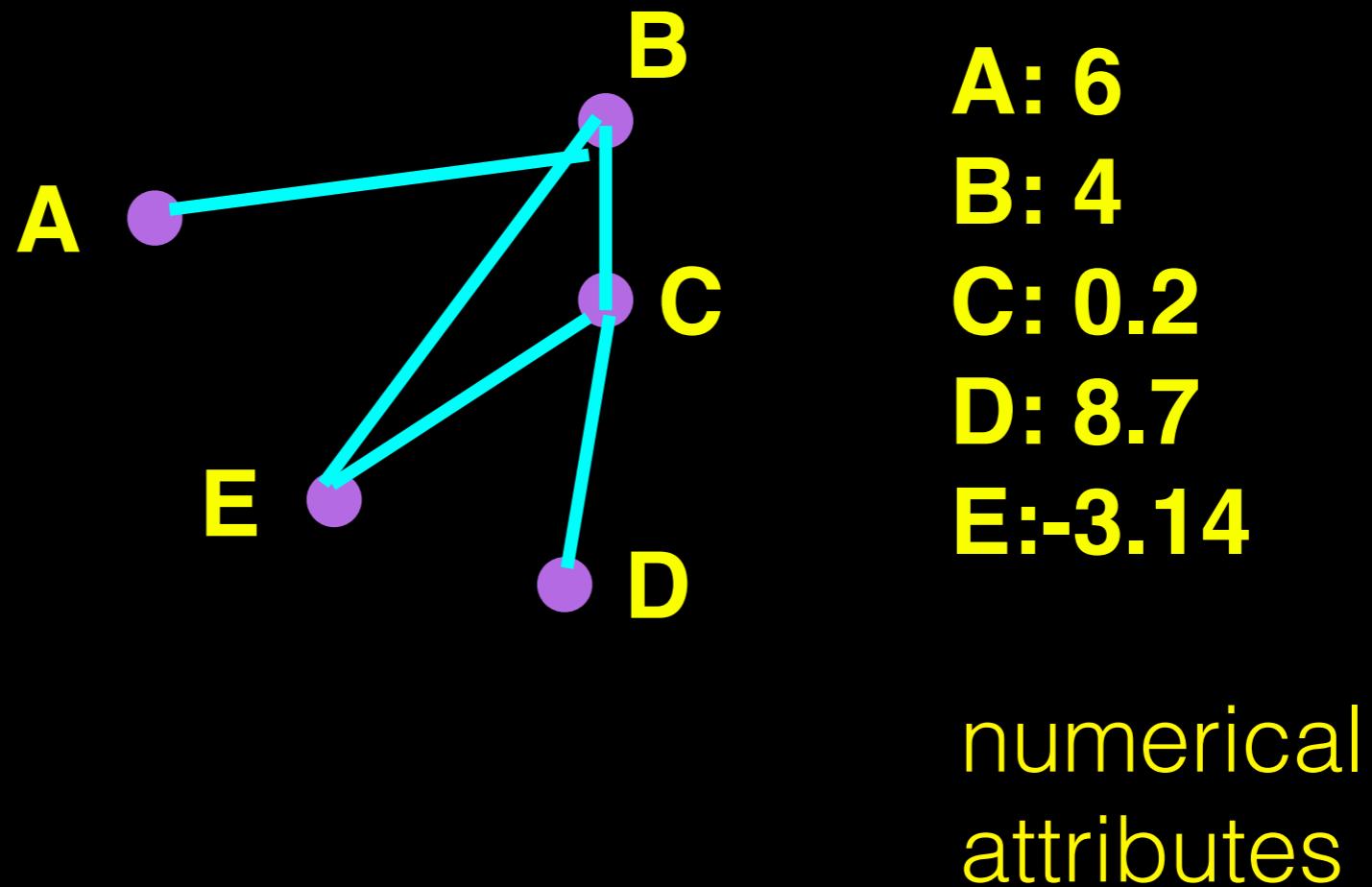


elements of A can take any value

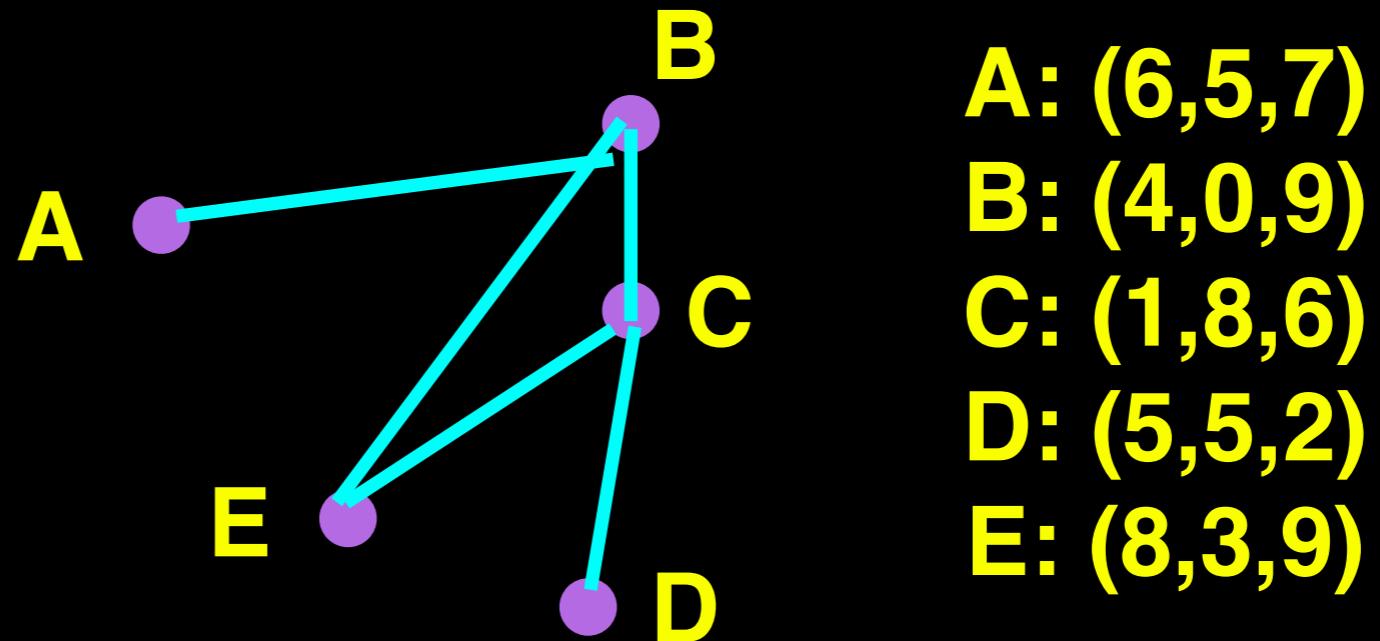
# vertex attributed graph



# vertex attributed graph

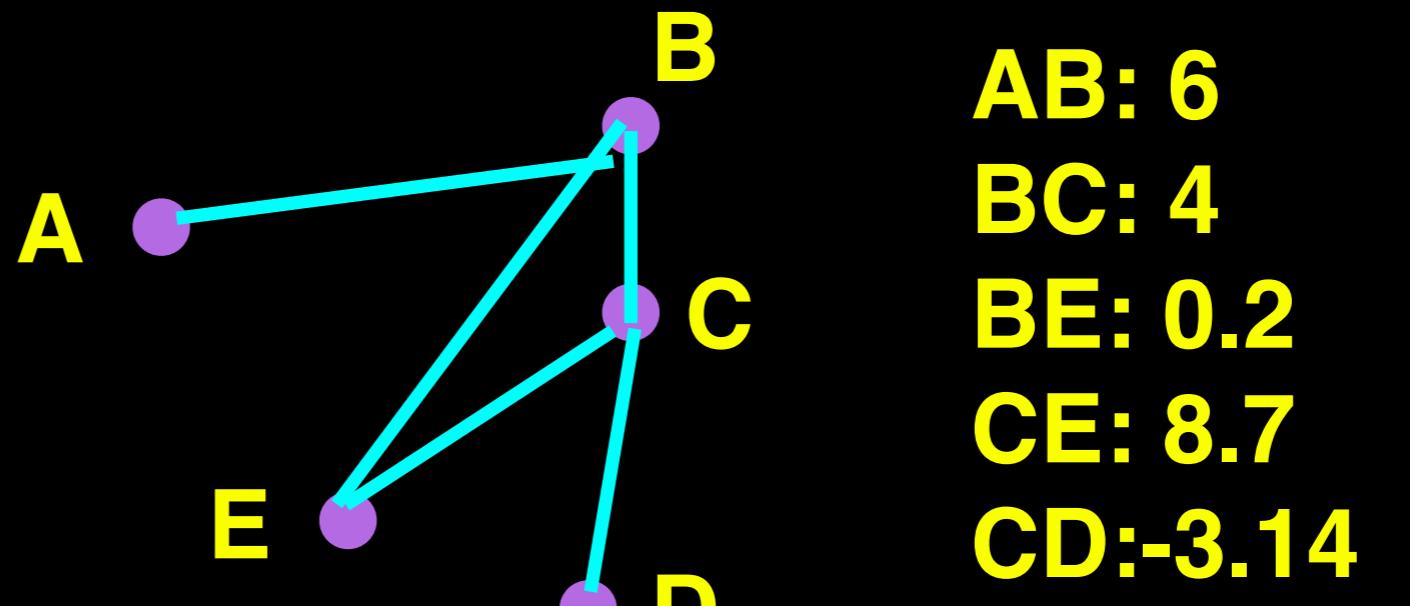


# vertex attributed graph



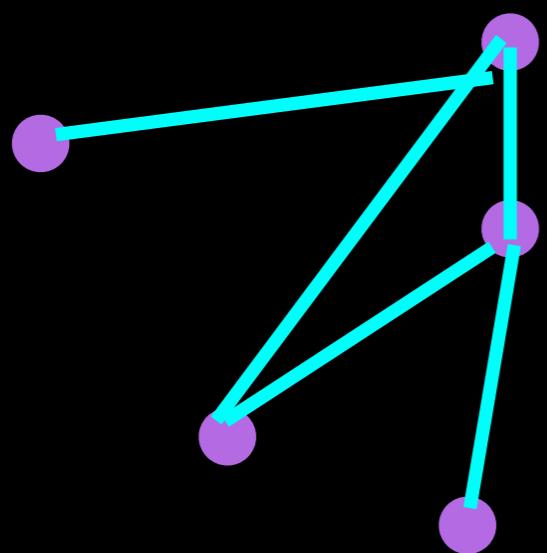
multi-dimensional  
attributes

# edge attributed graph

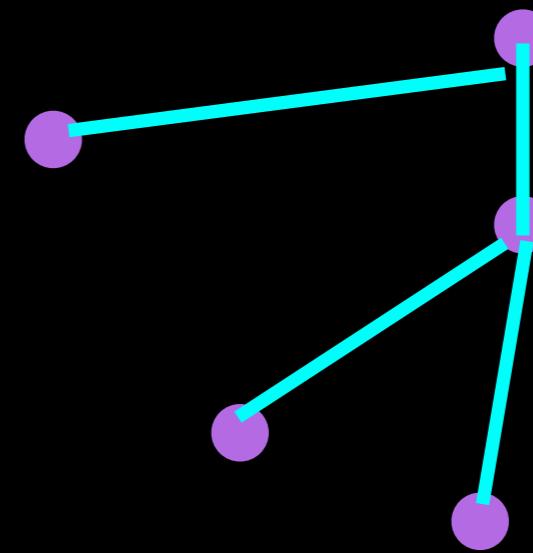


numerical  
attributes

# graph attributed graphs

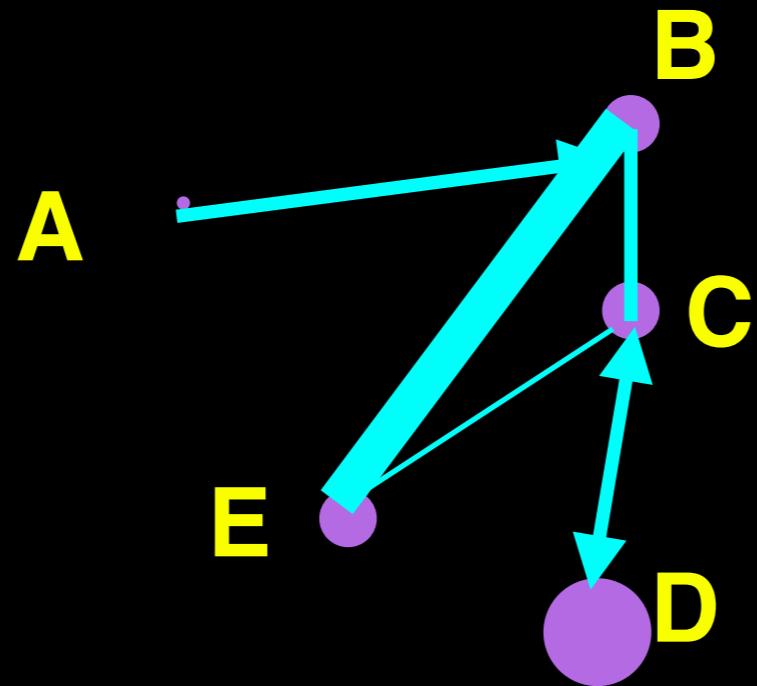


**subject 1**



**scan 2**

# Graphs with Rich attriUTEs (Grutes)



A: 6, L	AB: 6
B: 4, R	BC: 4
C: 1, L	BE: 0.2
D: 0, L	CE: 8.7
E: 8, R	CD:-3.14

# Latent Structure Model

- Each vertex can have **latent** & observed variable
  - can be categorical, numerical, vectors, etc.
- The latent variables have **structure**
  - can be cluster, nonlinear relationships, hierarchy, etc.