



**Brain Reading (MKI43)**

## **Lecture 7: Bayesian Connectomics**

**Marcel van Gerven**  
**Assistant Professor**  
**Distributed Representations Group**  
**Donders Centre for Cognition**

**Radboud University Nijmegen**





Rolf Kötter (1961-2010)

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

**Review**

## The Human Connectome: A Structural Description of the Human Brain

Olaf Sporns\*, Giulio Tononi, Rolf Kötter

**ABSTRACT**

The connection matrix of the human brain (the human “connectome”) represents an indispensable foundation for basic and applied neurobiological research. However, the network of anatomical connections linking the neuronal elements of the human brain is still largely unknown. While some databases or collations of large-scale anatomical connection patterns exist for other mammalian species, there is currently no connection matrix of the human brain, nor is there a coordinated research effort to collect, archive, and disseminate this important information. We propose a research strategy to achieve this goal, and discuss its potential impact.

**Introduction**

To understand the functioning of a network, one must know its elements and their interconnections. The purpose of this article is to discuss research strategies aimed at a comprehensive structural description of the network of elements and connections forming the human brain. We propose to call this dataset the human “connectome,” and we argue that it is fundamentally important in cognitive

Experimental approaches to human cognition have been significantly enhanced by the arrival of functional neuroimaging [5], a set of techniques that can be applied to study a broad range of cognitive functions, with ever-increasing spatial and temporal resolution. But the mechanistic interpretation of neuroimaging data is limited, in part due to a severe lack of information on the structure and dynamics of the networks that generate the observed activation patterns. A potential theoretical framework for conceptualizing cognition as a network phenomenon is based on two main organizational principles found in the cerebral cortex, functional segregation, and functional integration [6,7]. Emerging network theories of cognition emphasize the contextual [8], distributed [9], dynamic [10], and degenerate [11,12] nature of structure–function mappings in the brain. To successfully map structure to function in the human brain, we urgently need a comprehensive, detailed structural model of neuronal units and their connections. Connectional models of the human brain are scarce and poorly defined [13], and they are largely based on extrapolating anatomical information from other primate species such as the macaque monkey, an approach that is problematic [14], in part, because of our incomplete understanding of evolutionary

## Structural description at different scales

- **microscopic: neurons & synapses**
- **mesoscopic: minicolumns and associated connectivity**
- **macroscopic: brain regions and pathways**





## Washington University in Saint Louis - University of Minnesota

Oxford University - Saint Louis University - Advanced MRI Technologies - Indiana University - University d'Annunzio  
- Ernst Strungmann Institute - Warwick University - University of California at Berkeley

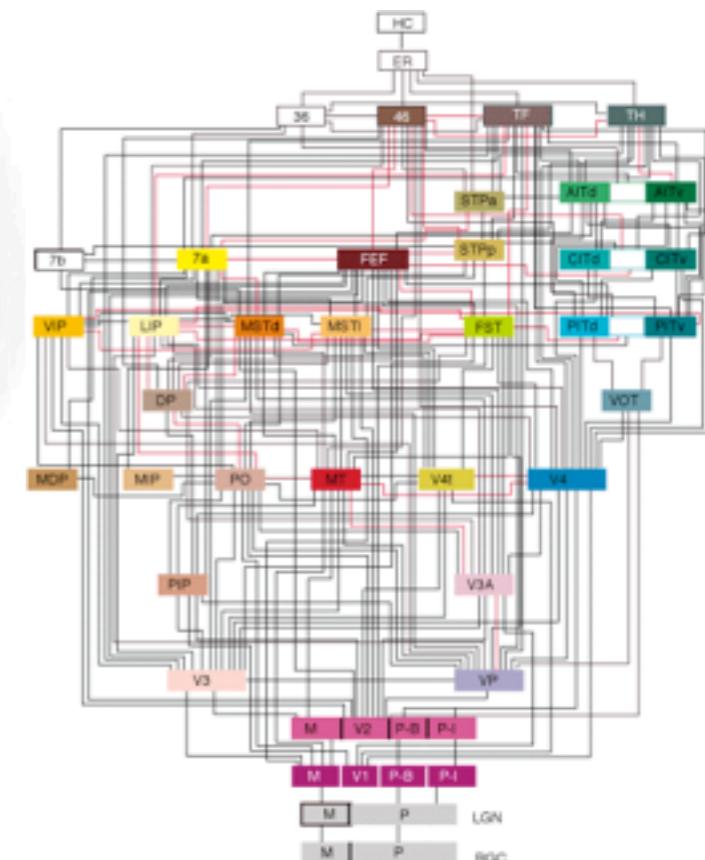
The Human Connectome Project is funded by the National Institutes of Health.

A 9-institution consortium led by Washington University in St. Louis and the University of Minnesota received a 5-year grant to enable development and utilization of advanced Magnetic Resonance Imaging (MRI) methods to chart brain circuitry.

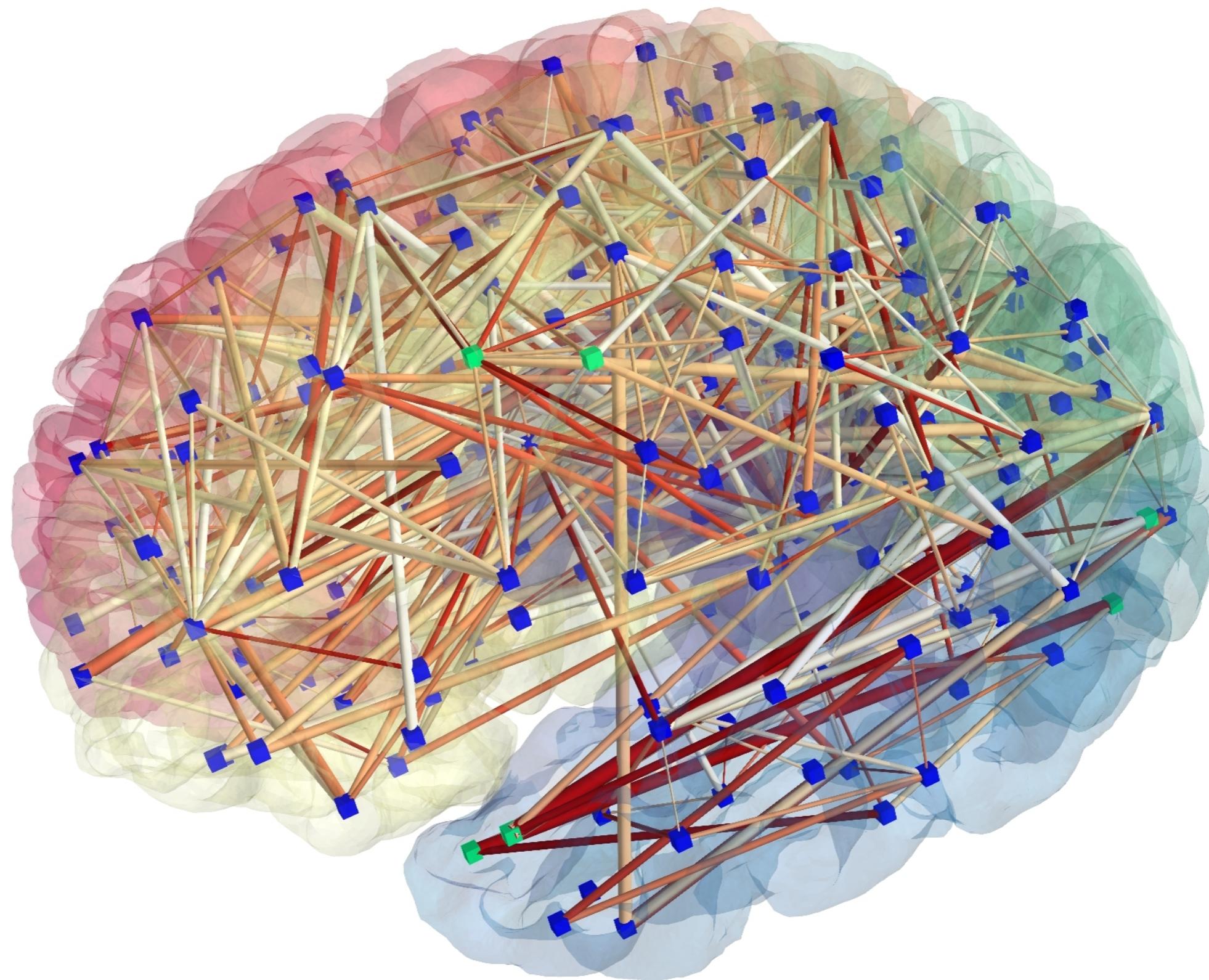
- 1,200 healthy adults (twins and their non-twin siblings) will be scanned on a customized 3T scanner using diffusion imaging, resting-state fMRI, and task-fMRI.
- 200 subjects will also be imaged at ultra-high field strengths (7T and/or 10.5T).
- 100 subjects will be studied using magnetoencephalography (MEG) and EEG.
- Connectivity patterns revealed by MR imaging of brain structure and function will be combined with behavioral testing and genotyping.
- Sophisticated data analysis and visualization capabilities will enable extensive data mining of these freely available datasets.



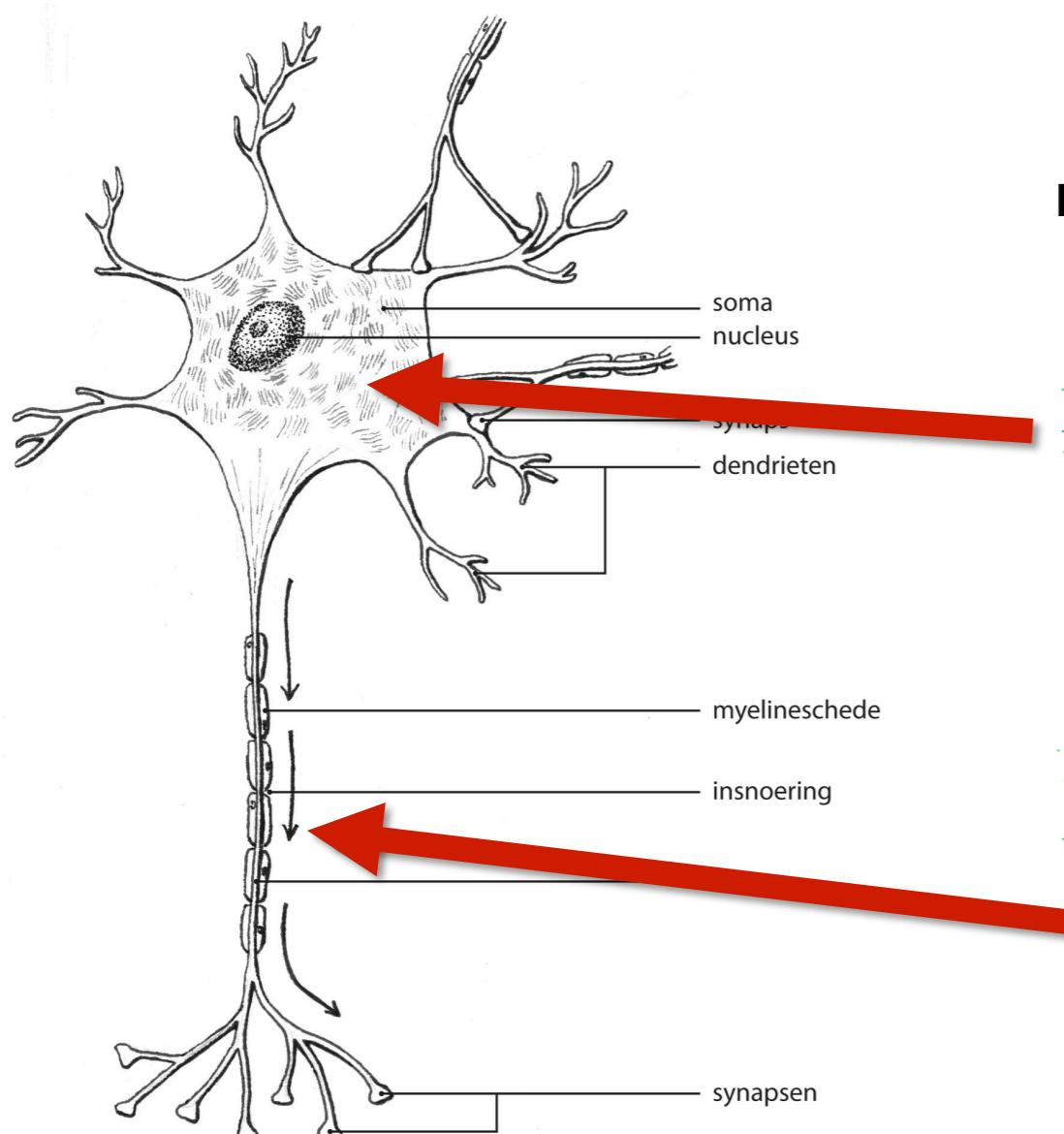
**David van Essen**



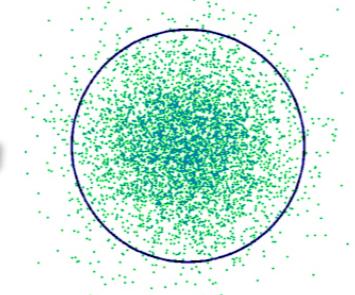
# The structural connectome



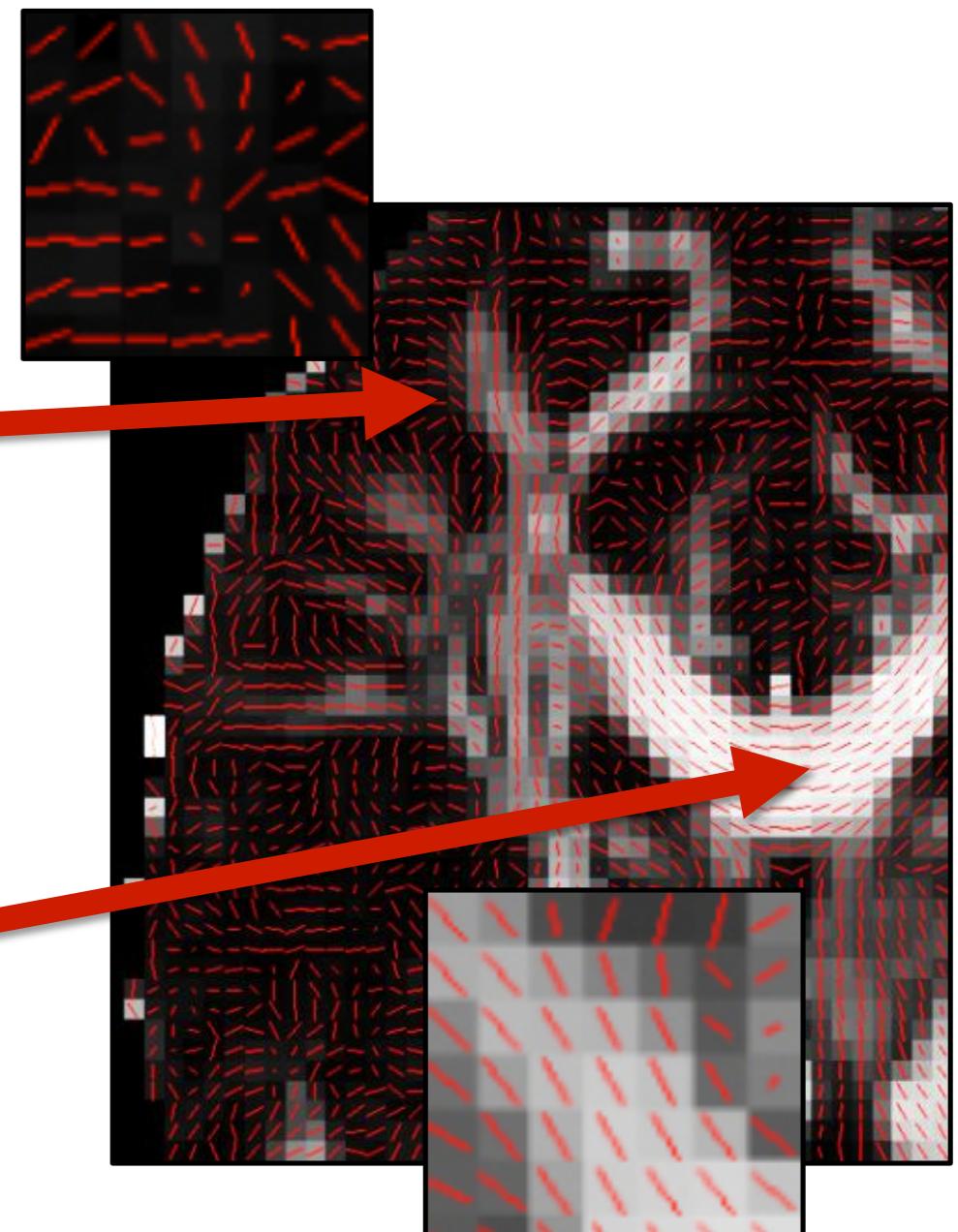
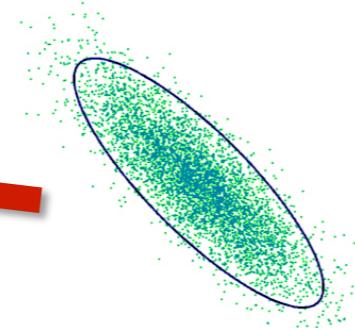
# Diffusion imaging



**Highly isotropic  
(low FA)**



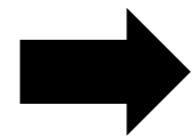
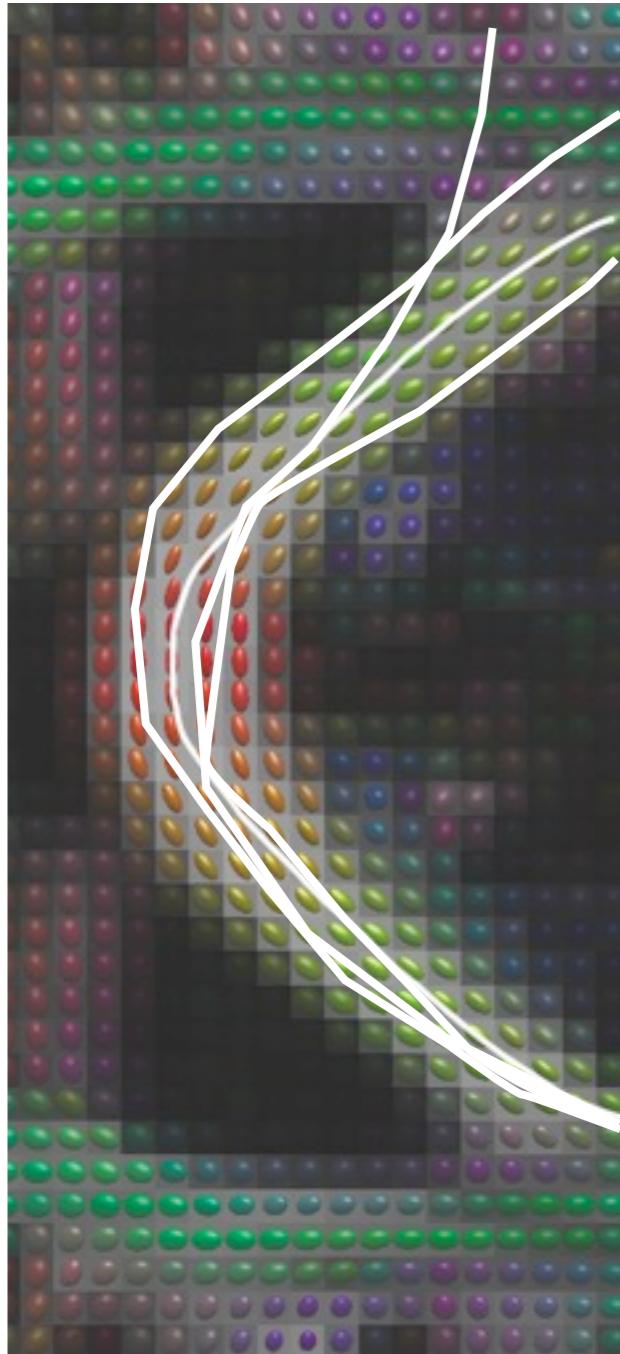
**Highly anisotropic  
(high FA)**



Tensor fitting



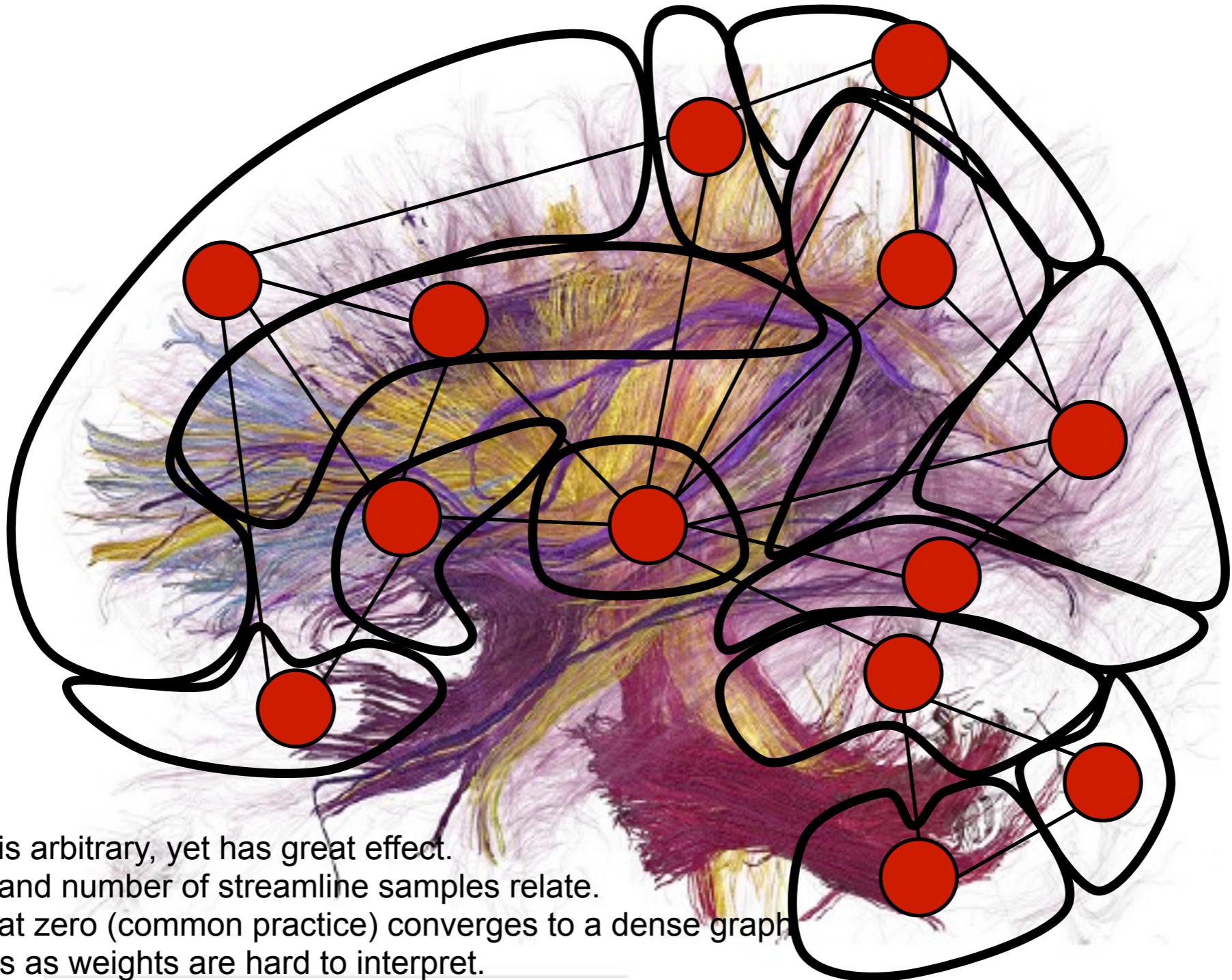
## probabilistic streamlining



# Structural brain networks using a threshold



3: (too) high (or too) low threshold



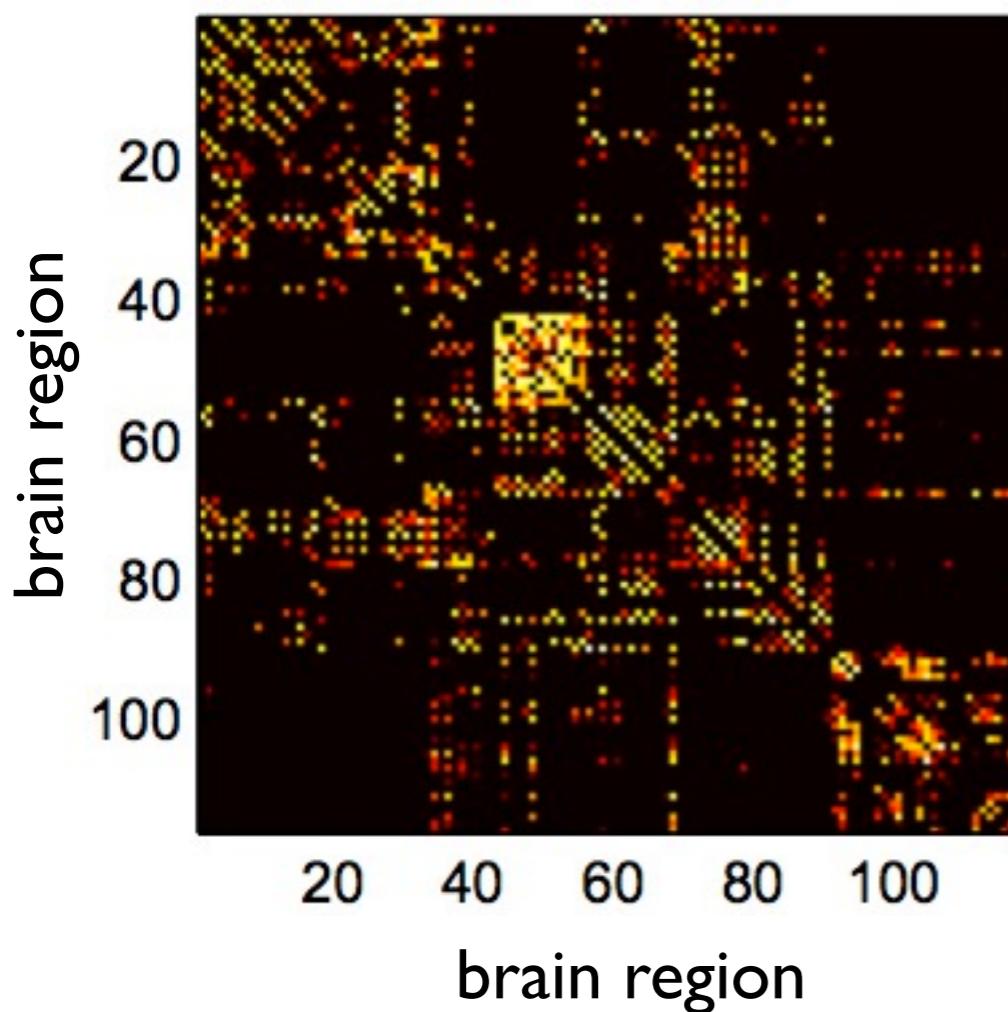
- Threshold is arbitrary, yet has great effect.
- Threshold and number of streamline samples relate.
- Threshold at zero (common practice) converges to a dense graph.
- Streamlines as weights are hard to interpret.

Goal: estimate structural connectivity from streamline counts



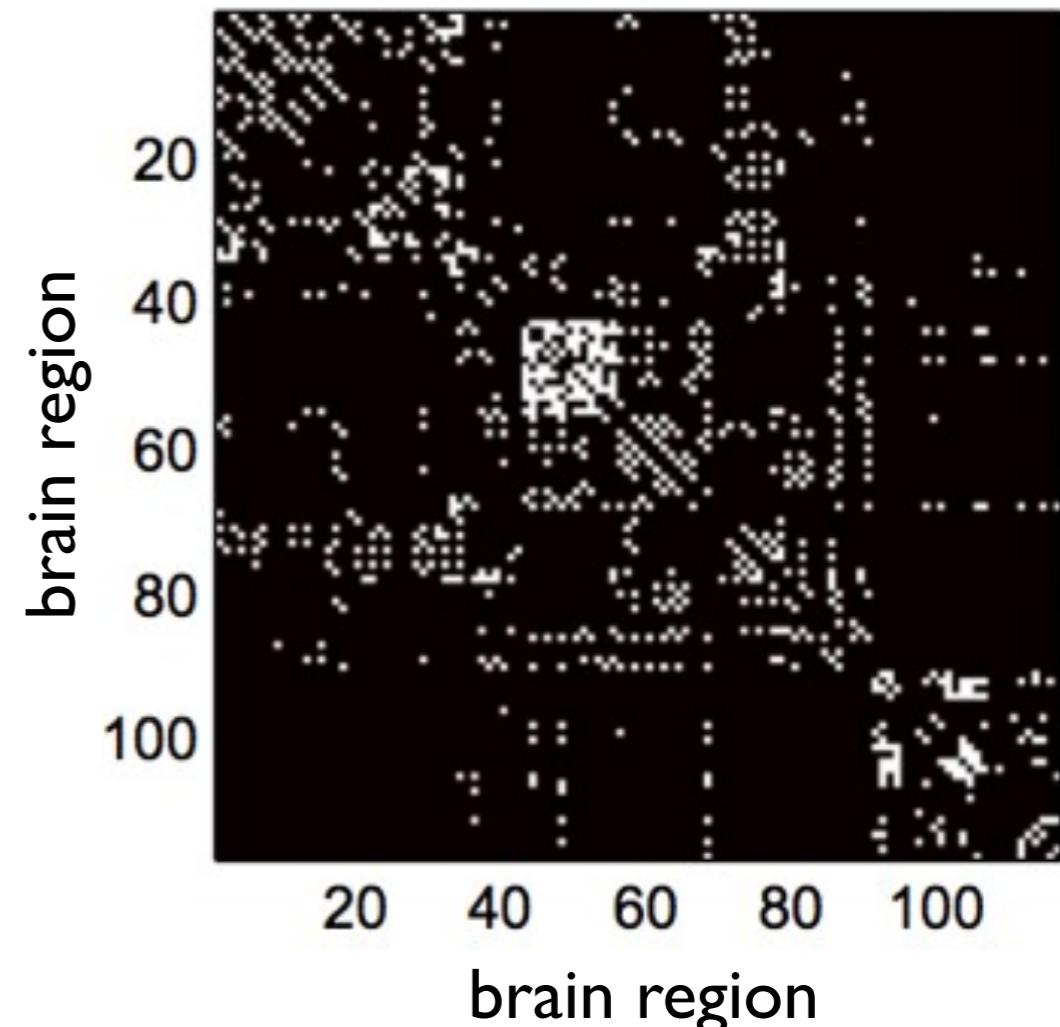
streamline count

$N$



structural connectivity

$A$



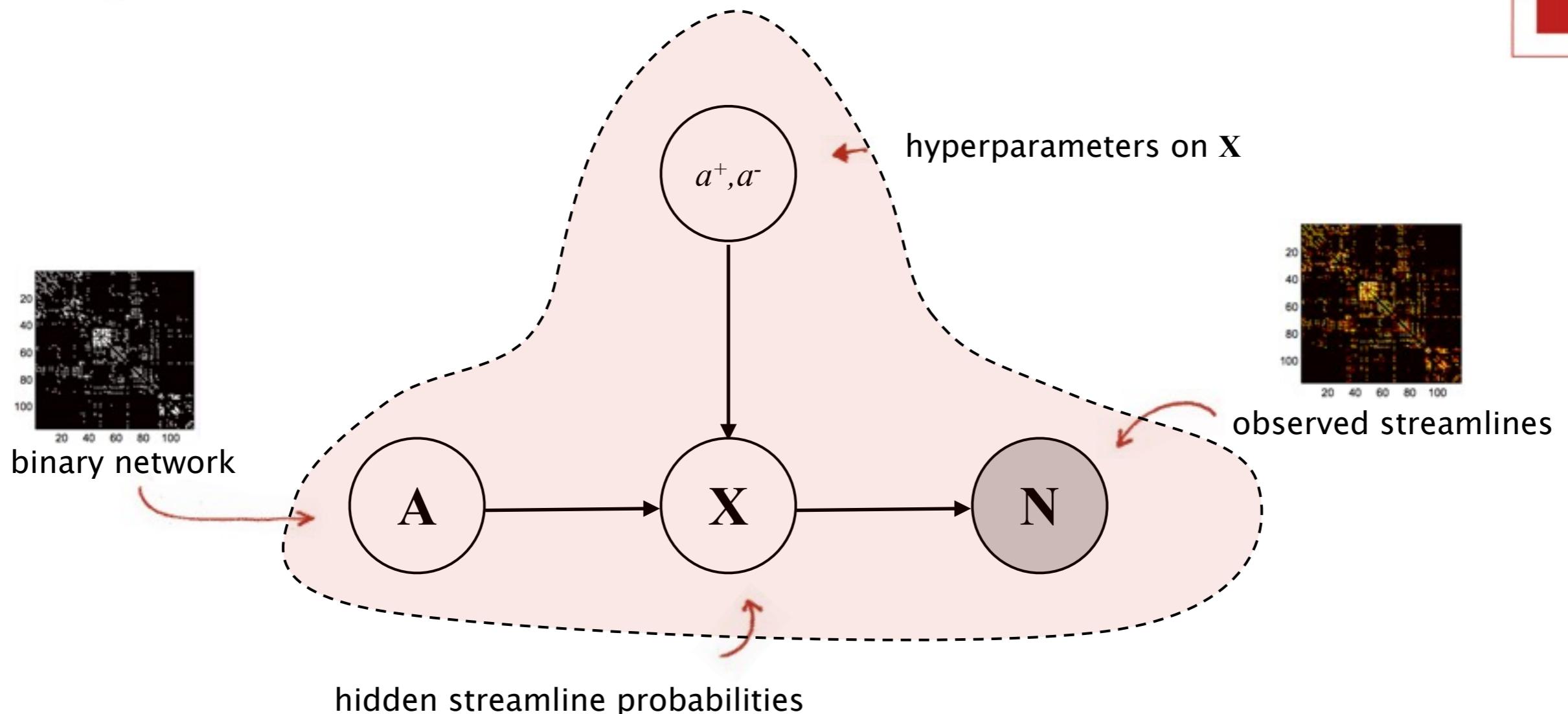


- We'd rather avoid arbitrary thresholds.
- We want to take into account the uncertainty inherent in streamlining
- We want to incorporate prior knowledge:
  - Graph-theoretical properties (e.g., degree distribution, small-worldness)
  - Anatomical constraints (i.e., minimize total edge length),
  - Known connections from other studies (i.e., edges found in tracer studies).
  - Data from other subjects/modalities.
- A Bayesian approach gives all of this (and more)

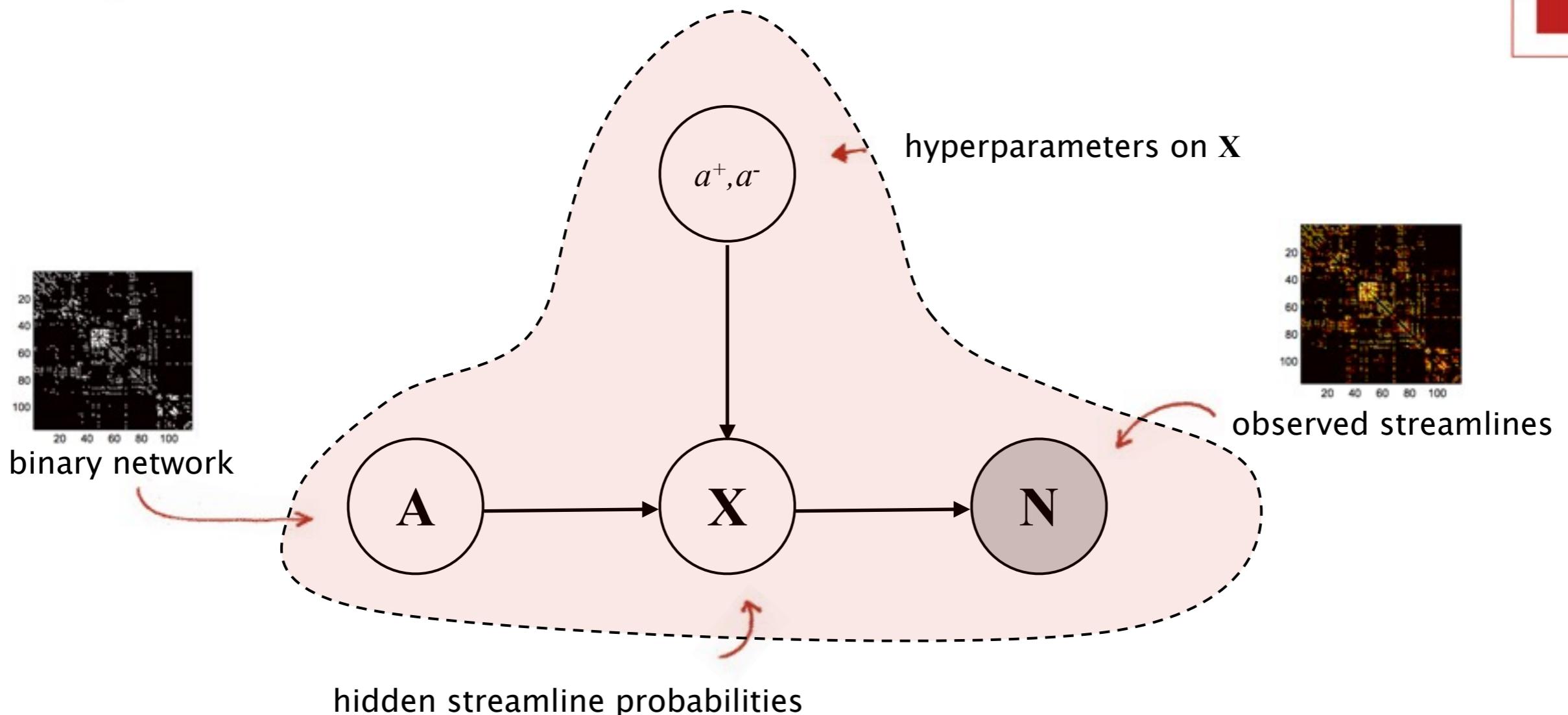
**Hinne, Heskes, Beckmann, van Gerven. 2012. Neuroimage**



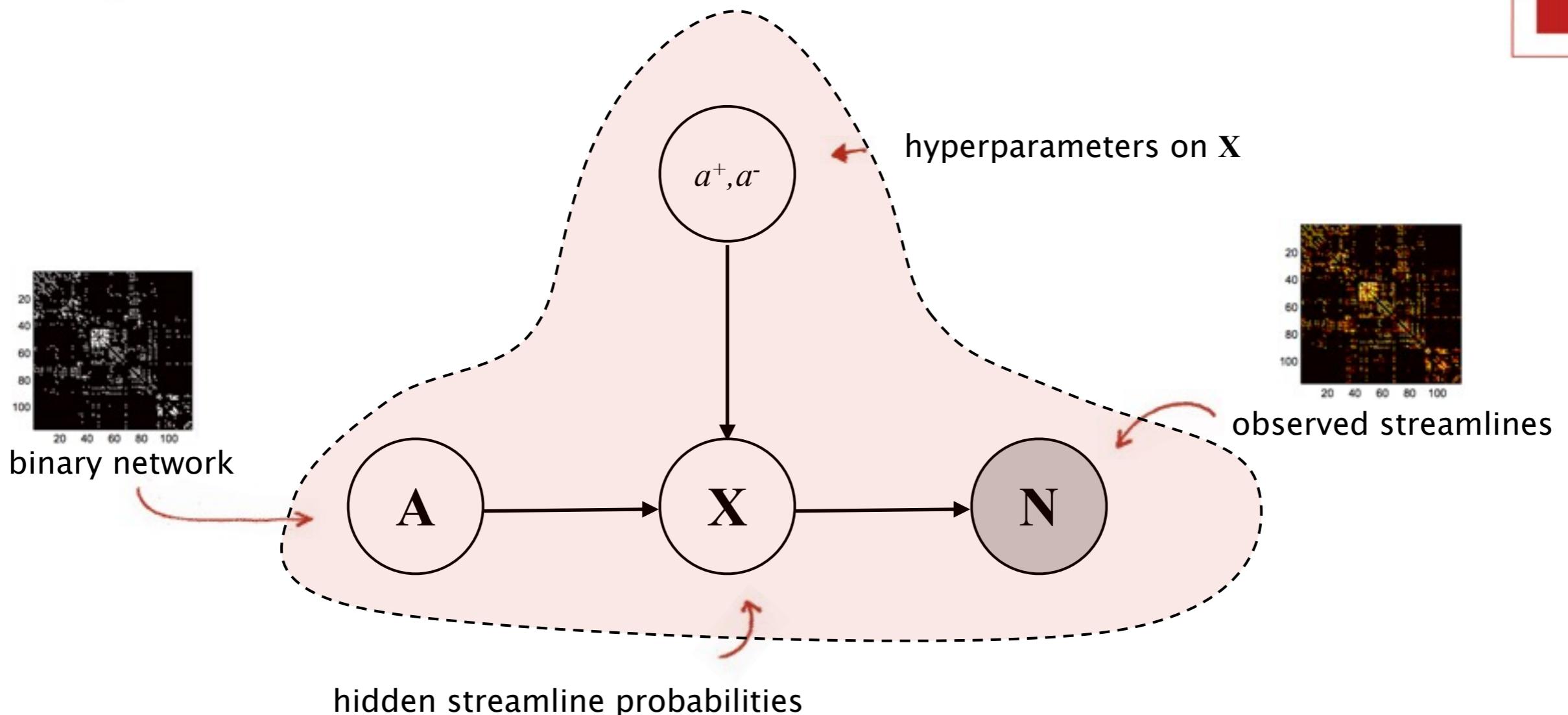
# Forward model for structural connectivity



# Forward model for structural connectivity



# Forward model for structural connectivity



$$\begin{aligned} P(\mathbf{X} \mid \mathbf{A}, a^+, a^-) &= \prod_i P(\mathbf{x}_i \mid \mathbf{a}_i, a^+, a^-) \\ &= \prod_i \text{Dirichlet}(\mathbf{x}_i \mid \boldsymbol{\alpha}_i) \\ &= \prod_i \frac{1}{\text{B}(\boldsymbol{\alpha}_i)} \prod_k x_{ik}^{\boldsymbol{\alpha}_i - 1} \end{aligned}$$

where  $\boldsymbol{\alpha}_i = a^+ \mathbf{a}_i + a^- (1 - \mathbf{a}_i)$

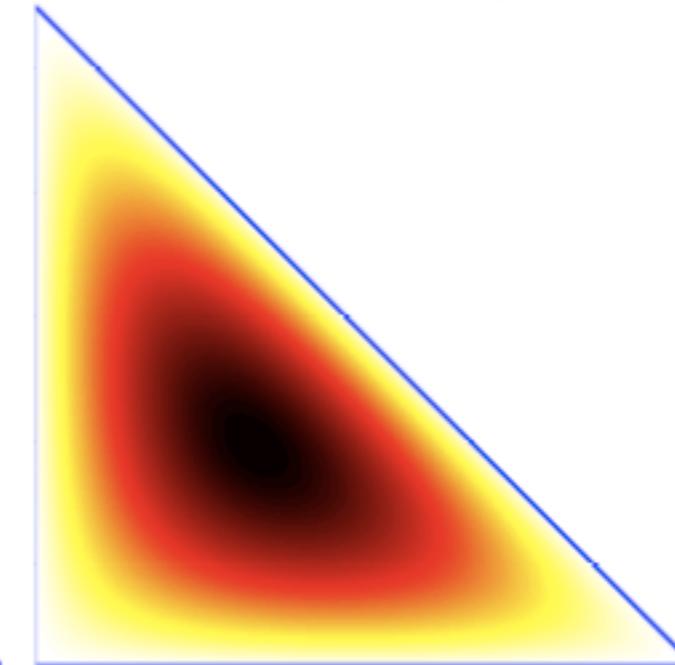
## Interpretation: Dirichlet distribution



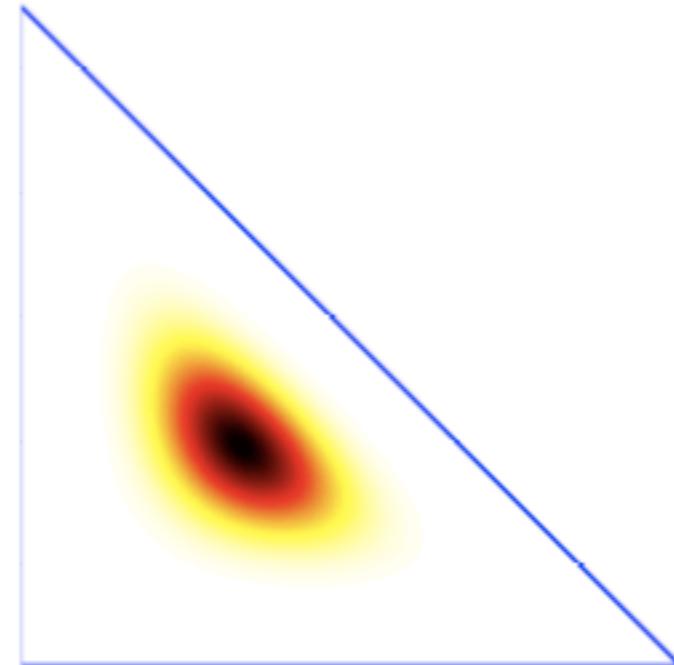
Dirichlet(1,1,1)



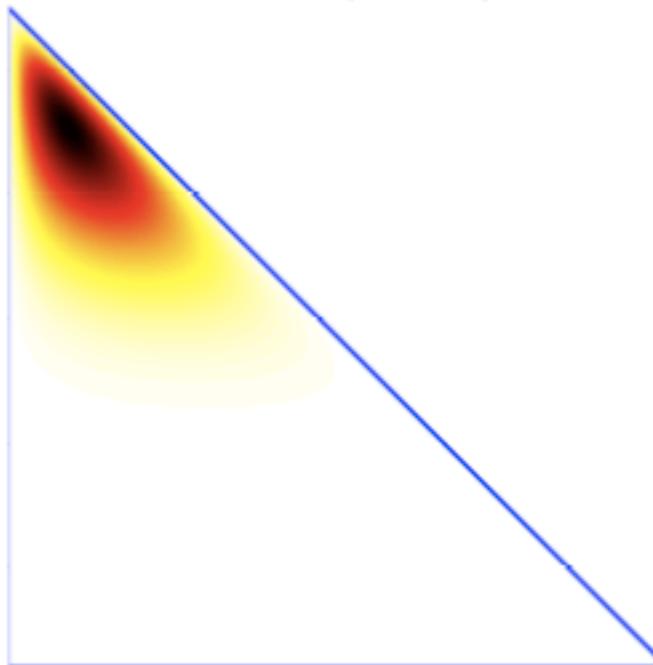
Dirichlet(2,2,2)



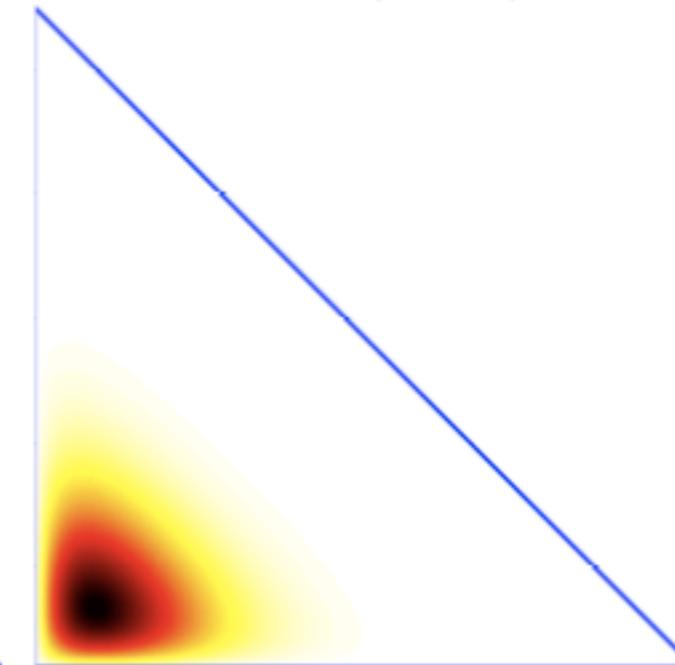
Dirichlet(10,10,10)



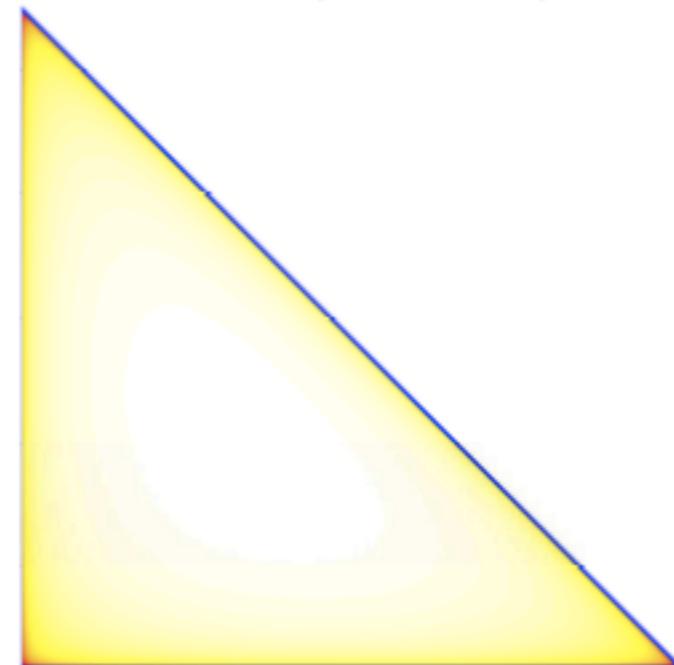
Dirichlet(2,10,2)



Dirichlet(2,2,10)



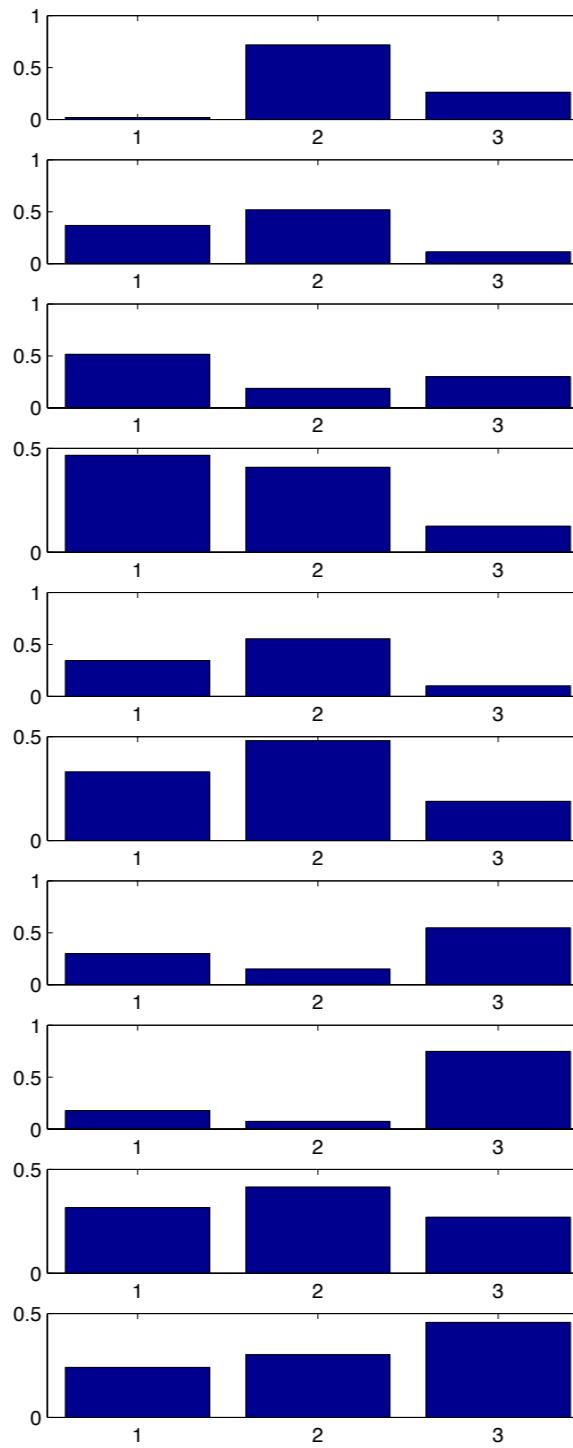
Dirichlet(0.9,0.9,0.9)



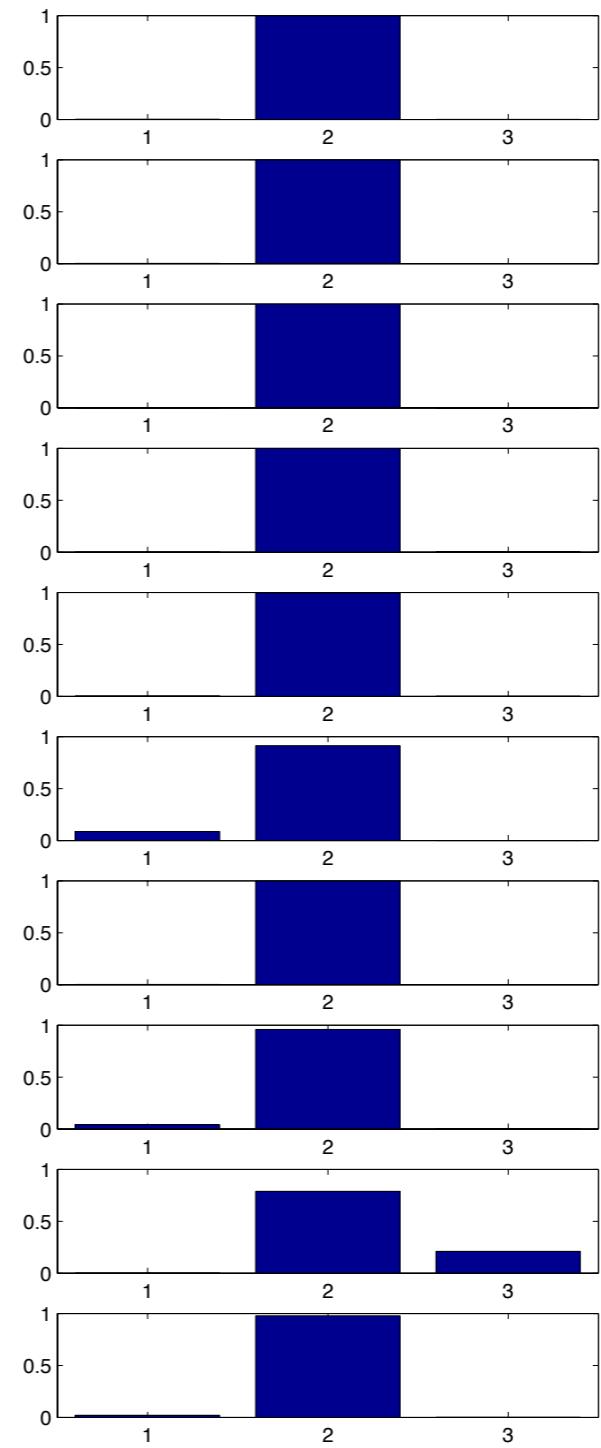
# Interpretation: samples from Dirichlet distribution



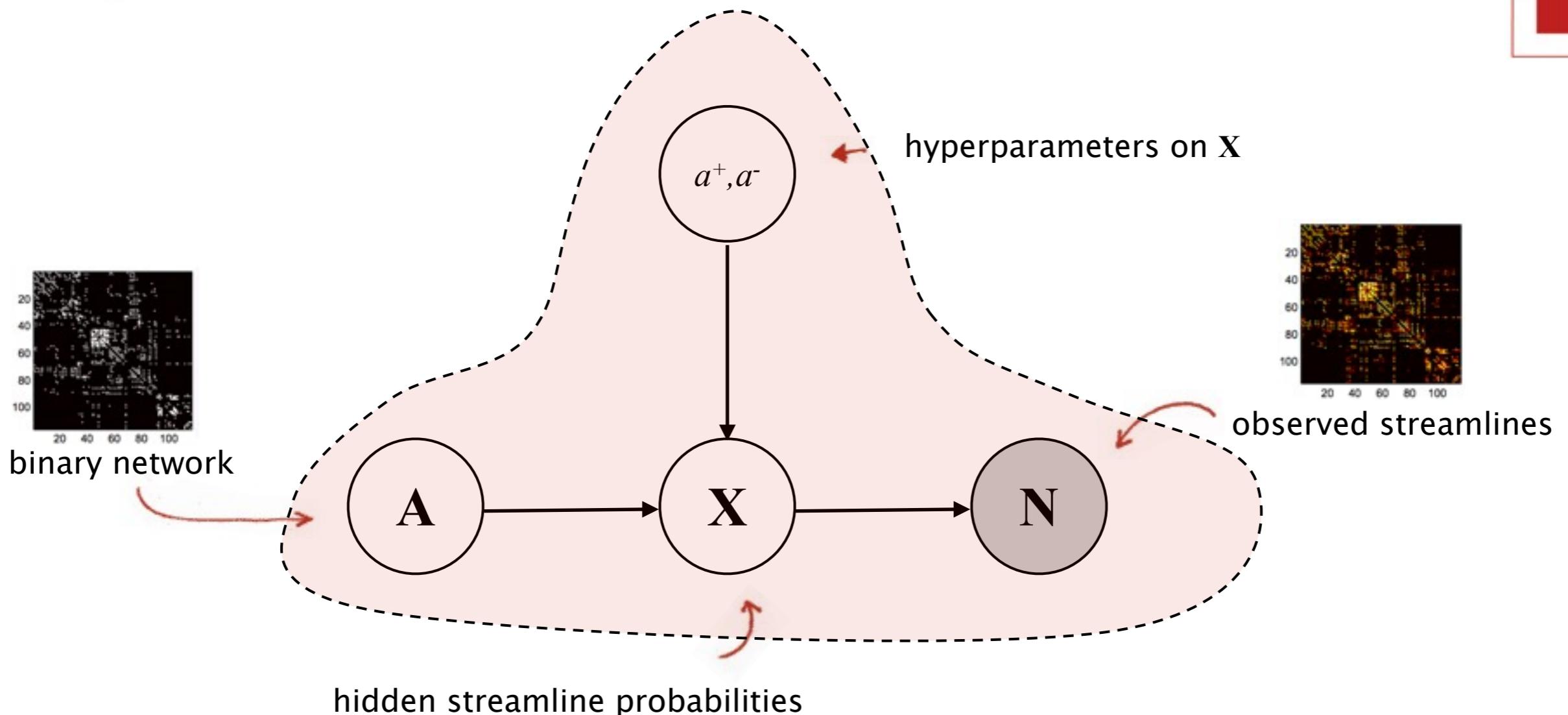
(2 2 2)



(0.1 2 0.1)

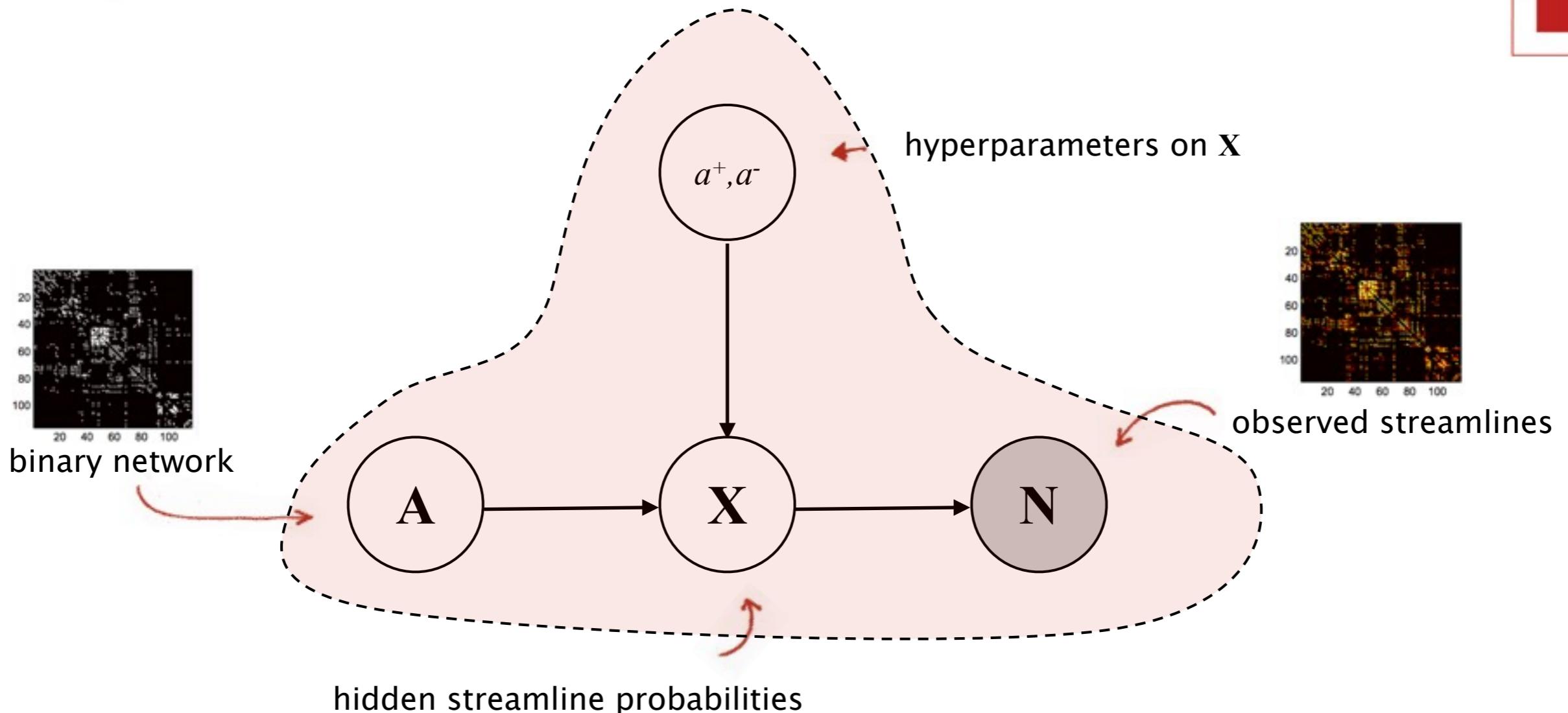


# Forward model for structural connectivity



$$\begin{aligned} P(\mathbf{N} \mid \mathbf{X}) &= \prod_i P(\mathbf{n}_i \mid \mathbf{x}_i) \\ &= \prod_i \text{Multinomial}(\mathbf{n}_i \mid \sum_k n_{ik}, \mathbf{x}_i) \\ &= \prod_i \frac{N_i}{n_{i1}! \cdots n_{iK}!} x_1^{n_{i1}} \cdots x_K^{n_{ik}} \end{aligned}$$

# Getting rid of the streamline probabilities

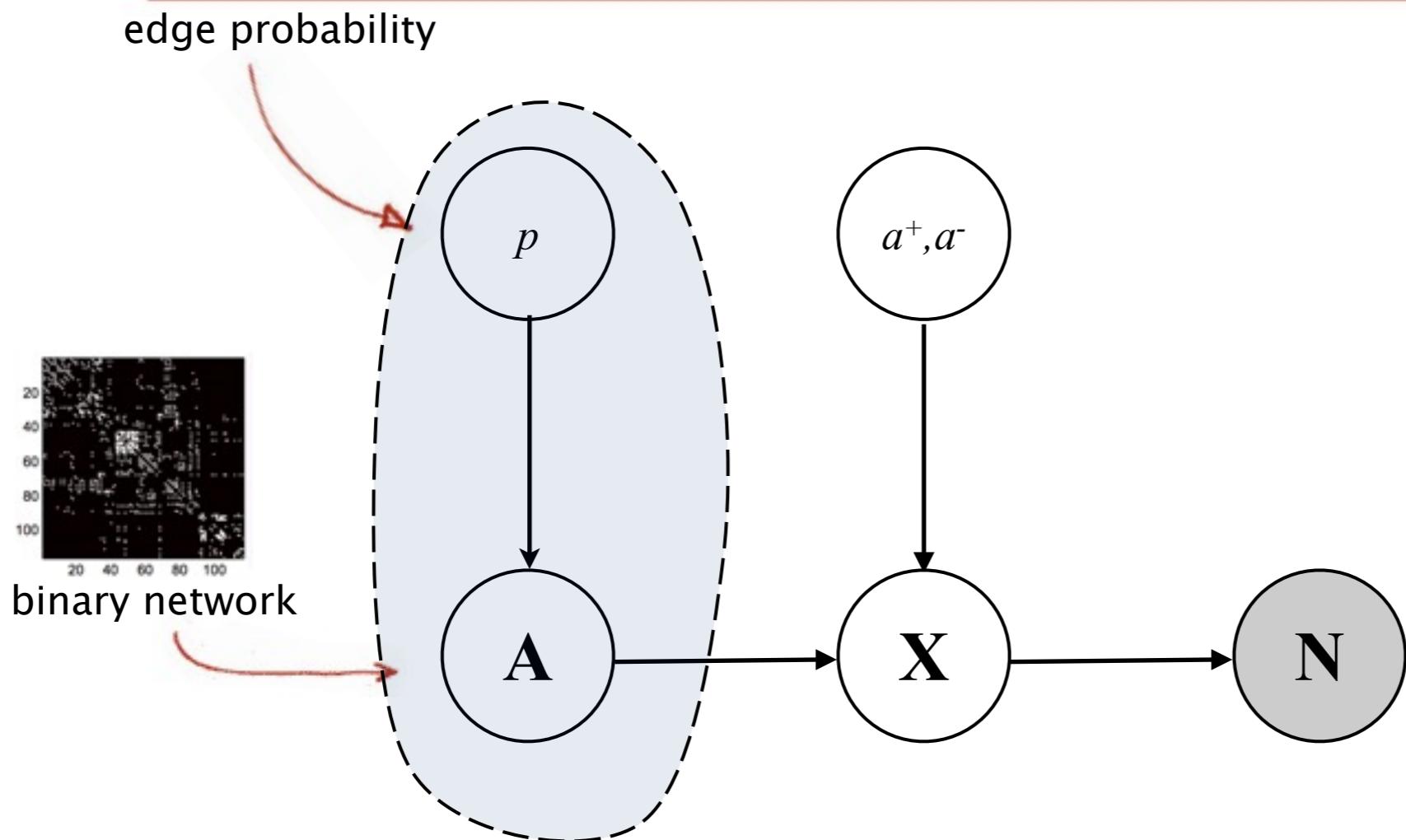


$$\mathbf{n}_i | \mathbf{x}_i \sim \text{Mult}(\mathbf{x}_i)$$

$$\mathbf{x}_i | \mathbf{a}_i, a^+, a^- \sim \text{Dir} (a^+ \mathbf{a}_i + a^- (\mathbf{1} - \mathbf{a}_i))$$

integrate out hidden streamline probabilities:

$$\begin{aligned} P(\mathbf{N} | \mathbf{A}, a^+, a^-) &= \int P(\mathbf{N} | \mathbf{X}) P(\mathbf{X} | \mathbf{A}, a^+, a^-) d\mathbf{X} \\ &= \prod_i \left[ \frac{N_i!}{\prod_j n_{ij}!} \frac{\Gamma(\sum_j a_{ij})}{\Gamma(\sum_j (a_{ij} + n_{ij}))} \right] \prod_j \frac{\Gamma(a_{ij} + n_{ij})}{\Gamma(a_{ij})} \end{aligned}$$



- We use the Erdős-Rényi (random graph) model as a prior on A:

$$P(a_{ij} | p) = p^{a_{ij}} (1 - p)^{1-a_{ij}}$$

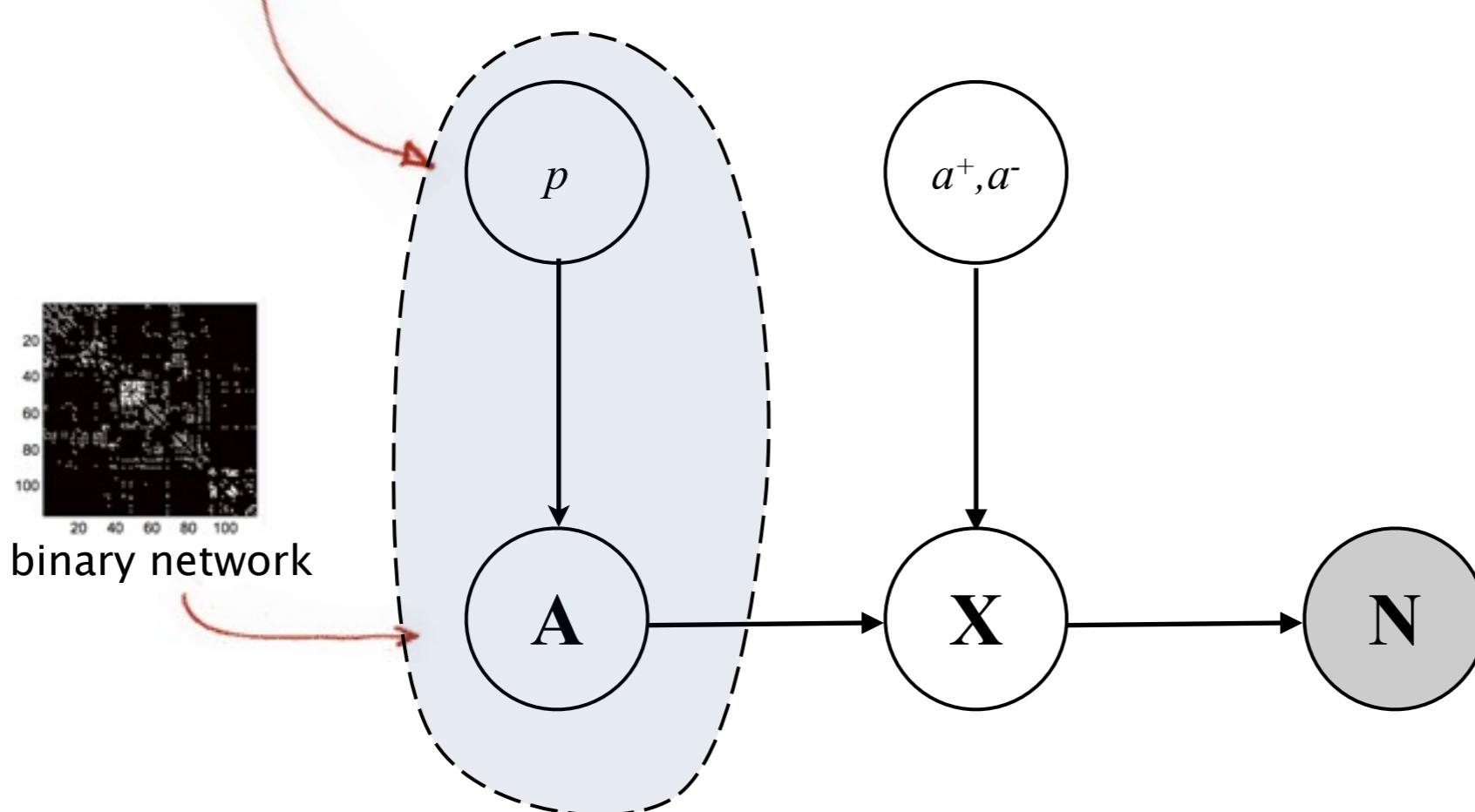
- If we say  $p=0.5$ , we obtain a flat prior.
- For  $p \neq 0.5$  we have a weighted coin-flip for each edge.
- More sophisticated priors are easily incorporated...



## Network prior



edge probability



$$\begin{aligned} P(A \mid p) &= \prod_{i < j} \text{Bernoulli}(a_{ij} \mid p) \\ &= \prod_{i < j} p^{a_{ij}} (1 - p)^{1 - a_{ij}} \end{aligned}$$



- We defined a generative model of how brain networks lead to observed data
- Now what?
- Compute the posterior distribution over structural networks given *observed* data and chosen hyper-parameters:

$$\text{structural networks} \quad \text{data} \quad \text{hyper-parameters}$$
$$P(\mathbf{A} | \mathbf{N}, \boldsymbol{\theta}) \propto P(\mathbf{N} | \mathbf{A}, a^+, a^-) P(\mathbf{A} | p)$$

with  $\boldsymbol{\theta} = (a^+, a^-, p)$

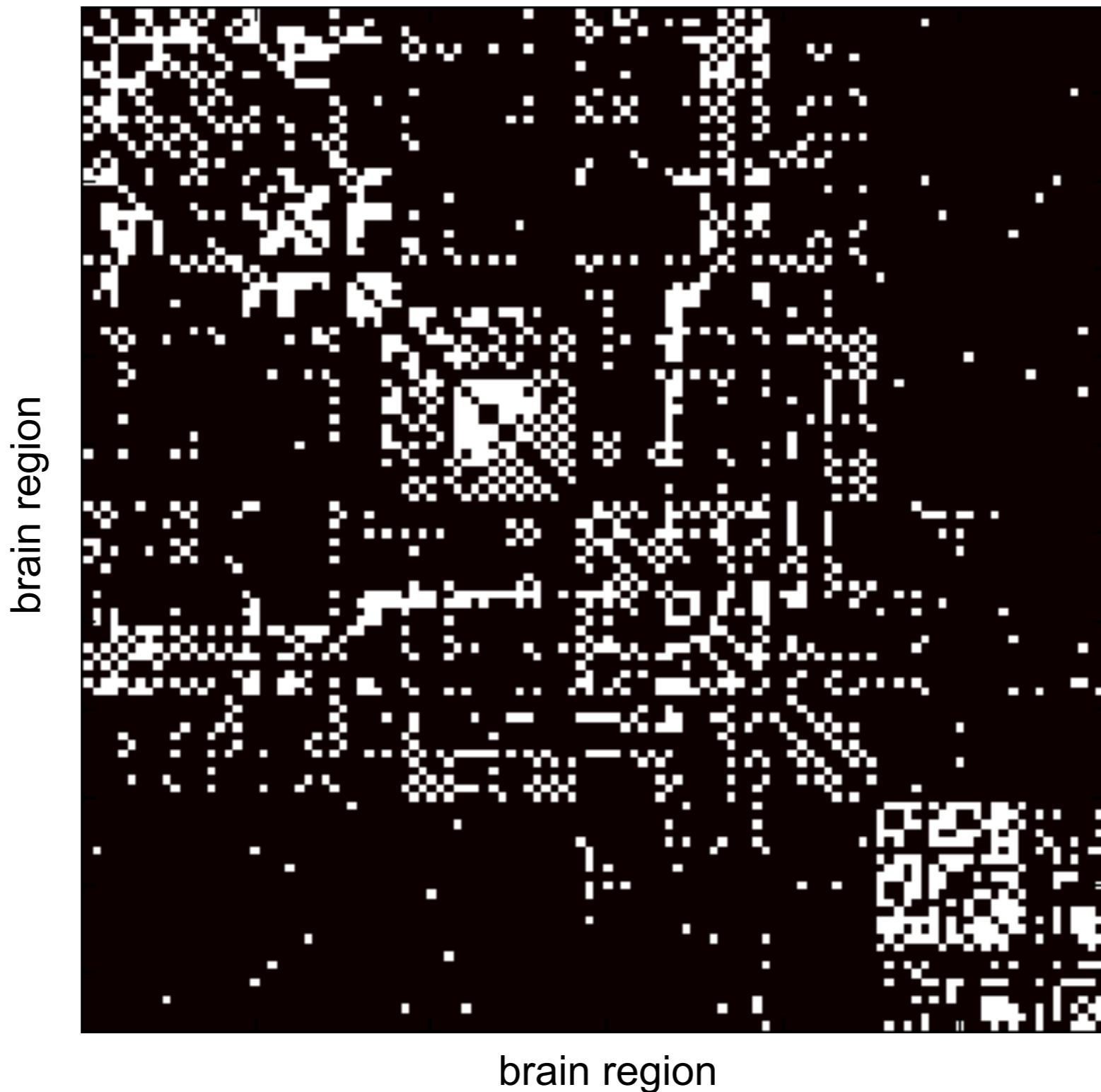
- We need to use *approximate* inference methods to estimate this distribution
- We use a Metropolis sampler - an example of a Markov chain Monte Carlo (MCMC) sampler



1. Set  $t=1$
2. Start with a random network  $\mathbf{A}^t$
3. Flip an edge  $a_{ij}$  to get a network  $\mathbf{A}'$
4. Accept the new network with probability

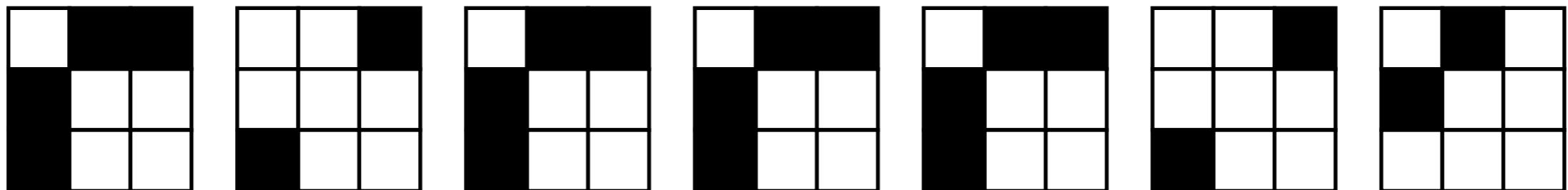
$$\gamma = \min \left\{ \frac{P(\mathbf{A}' | \mathcal{N}, \theta)}{P(\mathbf{A}^t | \mathcal{N}, \theta)}, 1 \right\}$$

5. If accepted, set  $\mathbf{A}^{t+1}=\mathbf{A}'$  else set  $\mathbf{A}^{t+1}=\mathbf{A}^t$
6. Set  $t=t+1$
7. Repeat from 3. hundreds of thousands of times
8. Discard the first few thousands of samples
9. Remaining samples approximate the posterior distribution





- We can compute any quantity of interest from the posterior
- E.g., the probability of a structural connection:



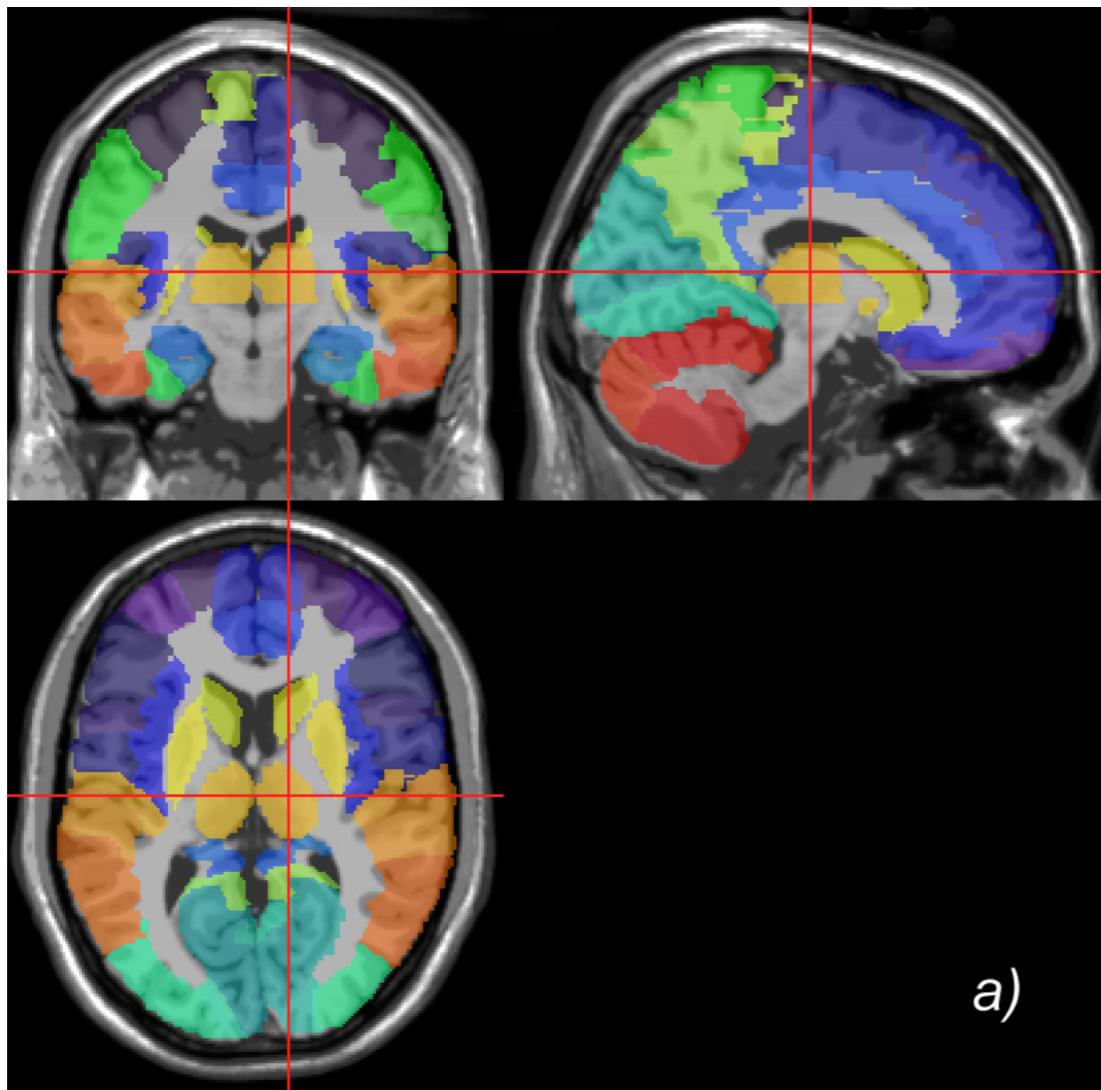
$$P(a_{12} \mid \mathcal{N}, \theta) \approx \frac{5}{7}$$

- Or selection of the most probable (MAP) graph in our set of samples
- Or computing confidence bounds on graph-theoretical quantities (e.g. small-worldness)
- This cannot be done with thresholding approaches
- Bayesian modeling is a very general technique to infer properties of unobservable variables of interest given observed data (cf. the scientific method)!



- We generated a set of samples that represent our structural networks
- Now what?
- VALIDATION!
- Compare different methods (e.g. thresholding vs Bayesian) by:
  - dissecting the brains for which diffusion imaging data has been collected and compare estimates with ground truth
  - look at replicability of estimates over subjects (but control for model complexity)
  - look how well independent data (e.g. functional data) are predicted from structural network estimates (but control for model complexity)

## Step 1: Node definition



- Automated Anatomical Labeling (AAL) atlas.
- Each node is the center of mass of one parcel.
- We count all streamlines connecting parcel A to parcel B.
- Intra-parcel streamlines are ignored.



- For each node, aside from streamlines we collect time series of BOLD activations
- Assume that the activation for a single node is modelled by a zero-mean Gaussian density with sparse inverse covariance (precision) matrix  $\mathbf{Q}$ :

$$P(\mathbf{y} \mid \mathbf{Q}) = (2\pi)^{-K/2} |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}^\top \mathbf{Q} \mathbf{y} \right\}$$

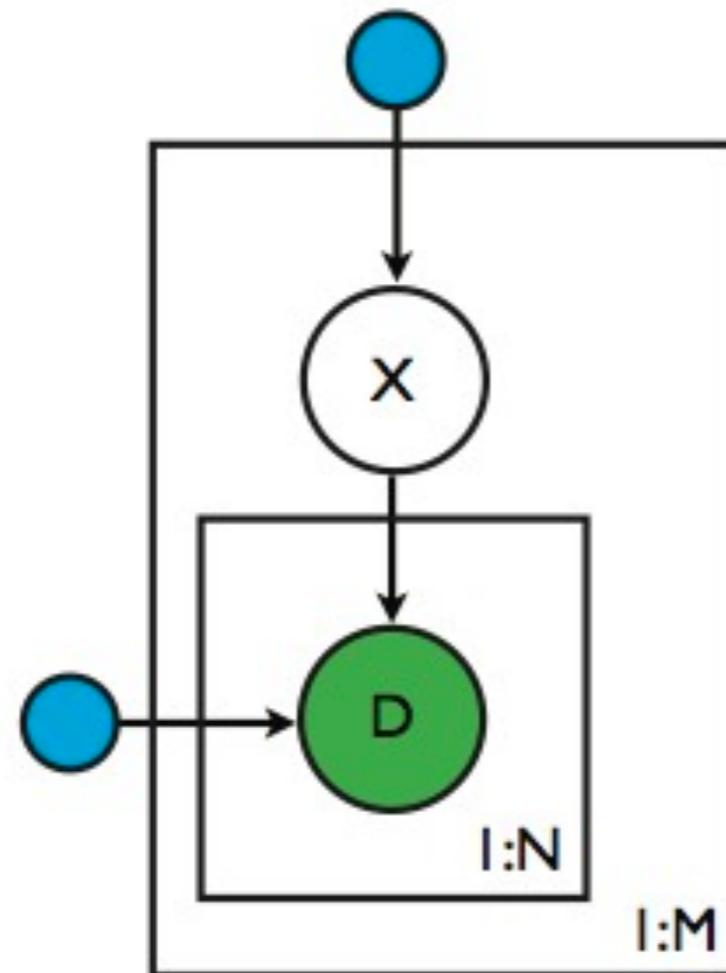
- The structural network  $G$  provides an estimate for the sparsity of  $\mathbf{Q}$ ;  $y_i$  and  $y_j$  are conditionally independent iff  $q_{ij} = 0$ , i.e.  $(i,j) \notin E_G$  (cf. Gaussian MRF)
- We can now compare different graphs by the loglikelihood of  $\mathbf{Q}$  constrained by  $G$ :

$$\ell(\mathbf{Q}) = (T/2)(\log \det \mathbf{Q} - \text{trace}(\mathbf{Q}\hat{\Sigma}))$$

- We compared thresholding with the Bayesian approach with thresholding



- We compared thresholding with the Bayesian approach with thresholding
- Threshold informed by the Bayesian approach (fixes model complexity)
- Both approaches have comparable average fit to the functional data
- But:
  - Thresholding was informed by Bayesian approach
  - Bayesian approach also gives confidence bounds on estimates
  - Bayesian approach can easily be extended to more complex models that can be shown to outperform thresholding
- Take-home message:
  - Proper thresholding can give quite reasonable estimates and is extremely fast
  - Bayesian approach is much richer and more precise but takes more time to compute



N = number of datasets

M = number of subjects

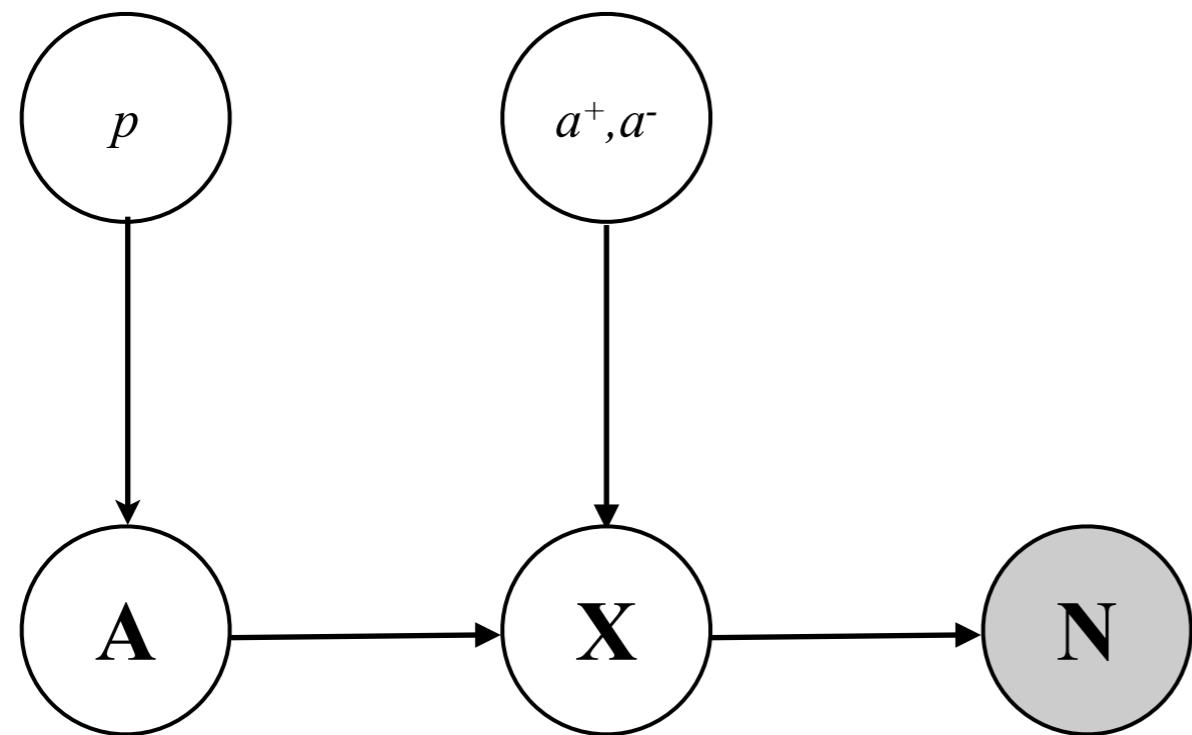
## Philosophy:

We should be able to write down a generative model which explains our data in terms of the states of latent (unobservable) random variables.

The goal of cognitive neuroscience should be to define these latent variables and to infer their states using Bayesian inference methods from all data hand.

## Examples of latent variables:

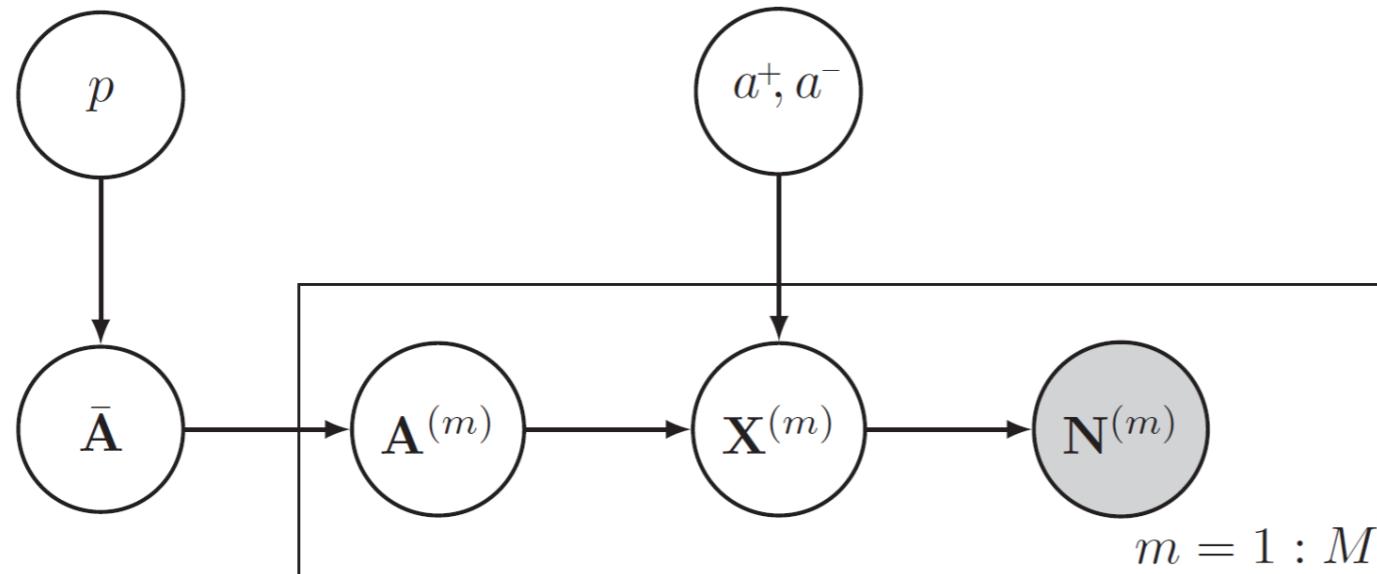
- structural connectivity
- functional interactions (cf. DCM)
- subject groups
- brain parcellations
- cognitive states



- Put priors on the hyper-parameters
- Model other sources of observed data
- Combine data collected for multiple subjects
- Use richer network priors
- Link estimates to behavioural data (e.g. cognitive performance, clinical states)



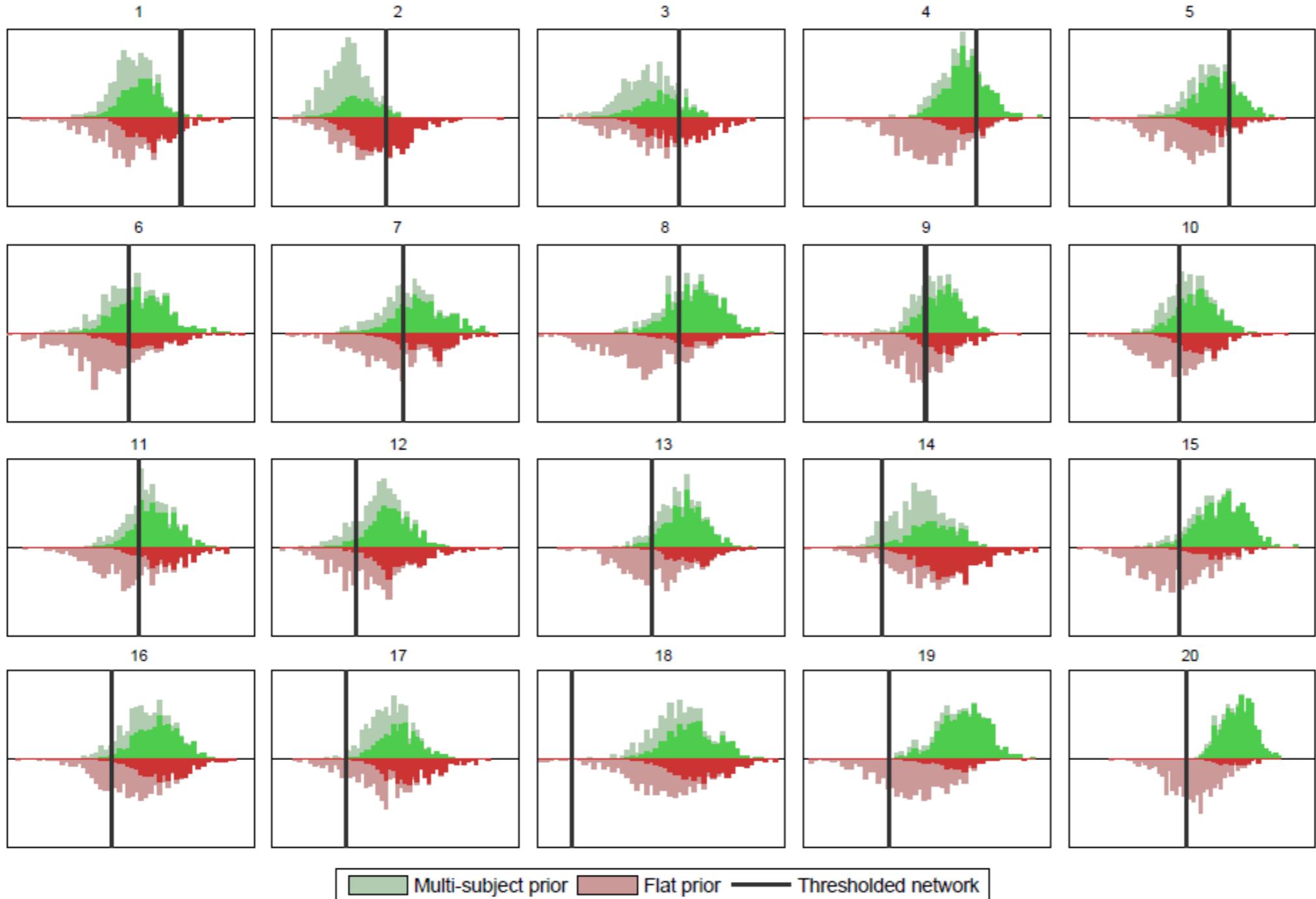
- If we obtain data for more than one subjects (as usual), it makes sense to infer all connectivity at once, as the connectivity is likely not independent:



- Sampling all connectivity for  $M$  subjects and a ‘parent’ network is expensive.
- Alternatively, we can derive a prior for a new subject  $M+1$ , based on connectivity for previously seen subjects.
- In this case  $p_{ij}$  is the probability of an edge given its occurrence in the maximum likelihood connectivity  $\hat{\mathbf{A}}$  for other subjects:

$$p_{ij} = \frac{\sum_{m=1}^M \hat{a}_{ij}^{(m)} + 1}{M + 2}$$

# Thresholding vs single-subject vs multi-subject

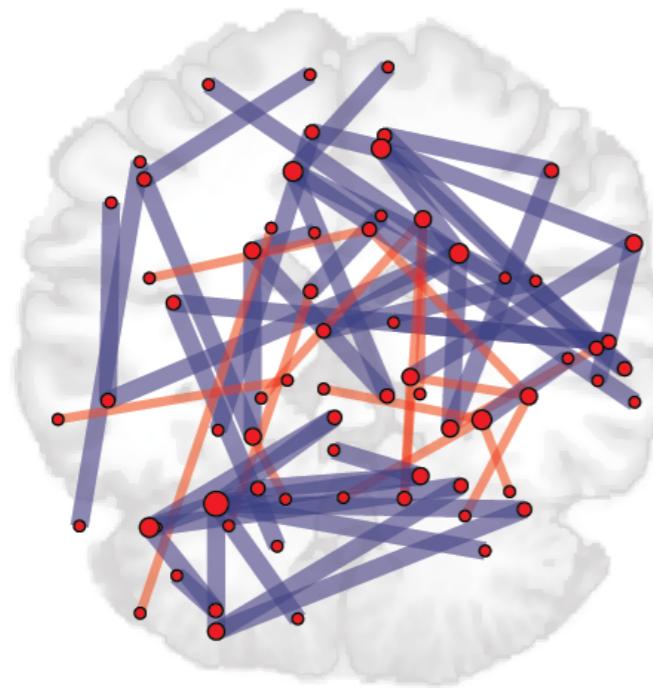


#	$f_{F-T}$	$f_{M-T}$	$f_{M-F}$
1	0.01	0.01	0.52
2	0.35	0.03	0.26
3	0.35	0.12	0.39
4	0.09	0.23	0.78
5	0.10	0.24	0.72
6	0.31	0.56	0.69
7	0.40	0.62	0.63
8	0.24	0.67	0.80
9	0.42	0.69	0.68
10	0.47	0.70	0.65
11	0.43	0.71	0.65
12	0.68	0.84	0.57
13	0.56	0.91	0.71
14	0.89	0.91	0.40
15	0.46	0.93	0.85
16	0.75	0.93	0.61
17	0.88	0.99	0.52
18	0.99	1.00	0.59
19	0.78	1.00	0.86
20	0.64	1.00	0.92

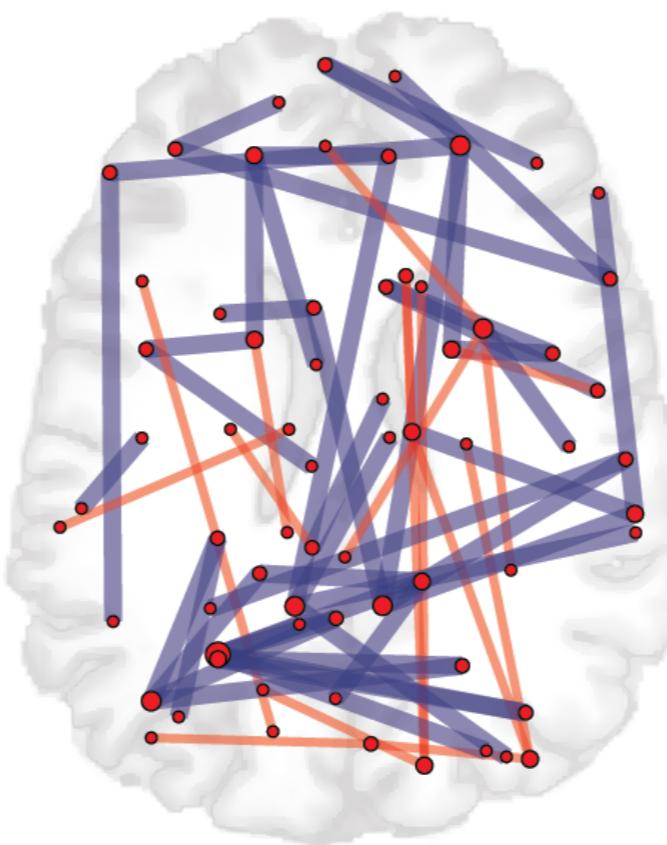
■ Multi-subject prior ■ Flat prior — Thresholded network



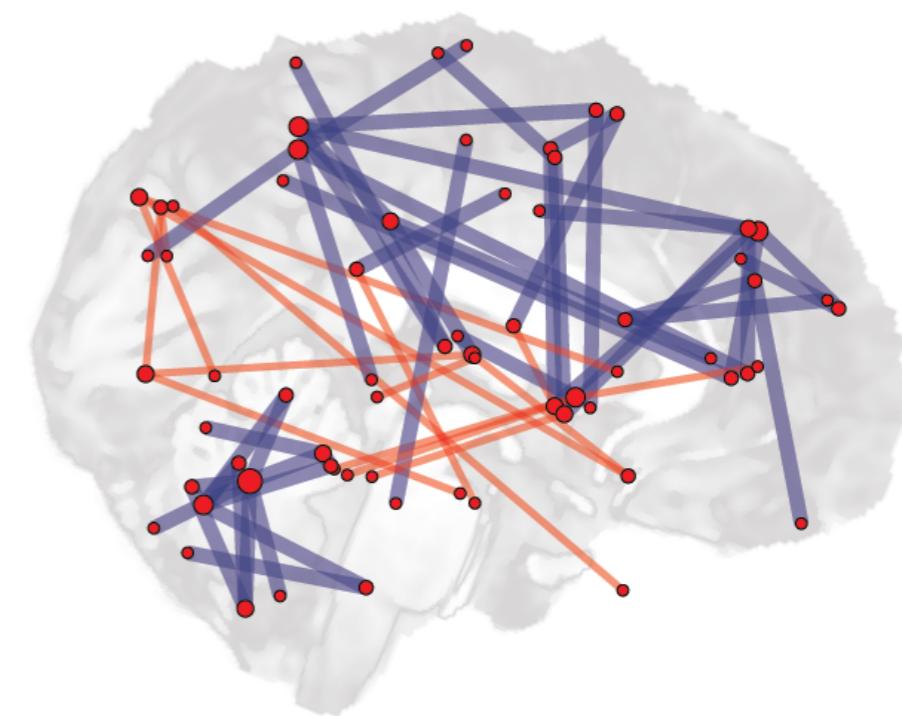
- **Blue:** edges present in many MAP networks, without being present in thresholded network.
- **Orange:** edges present in many thresholded networks, without being present in MAP network.
- In short: more anterior and cerebellar connections.



Coronal view (front-back)



Axial view (top-bottom)



Sagittal view (left-right)

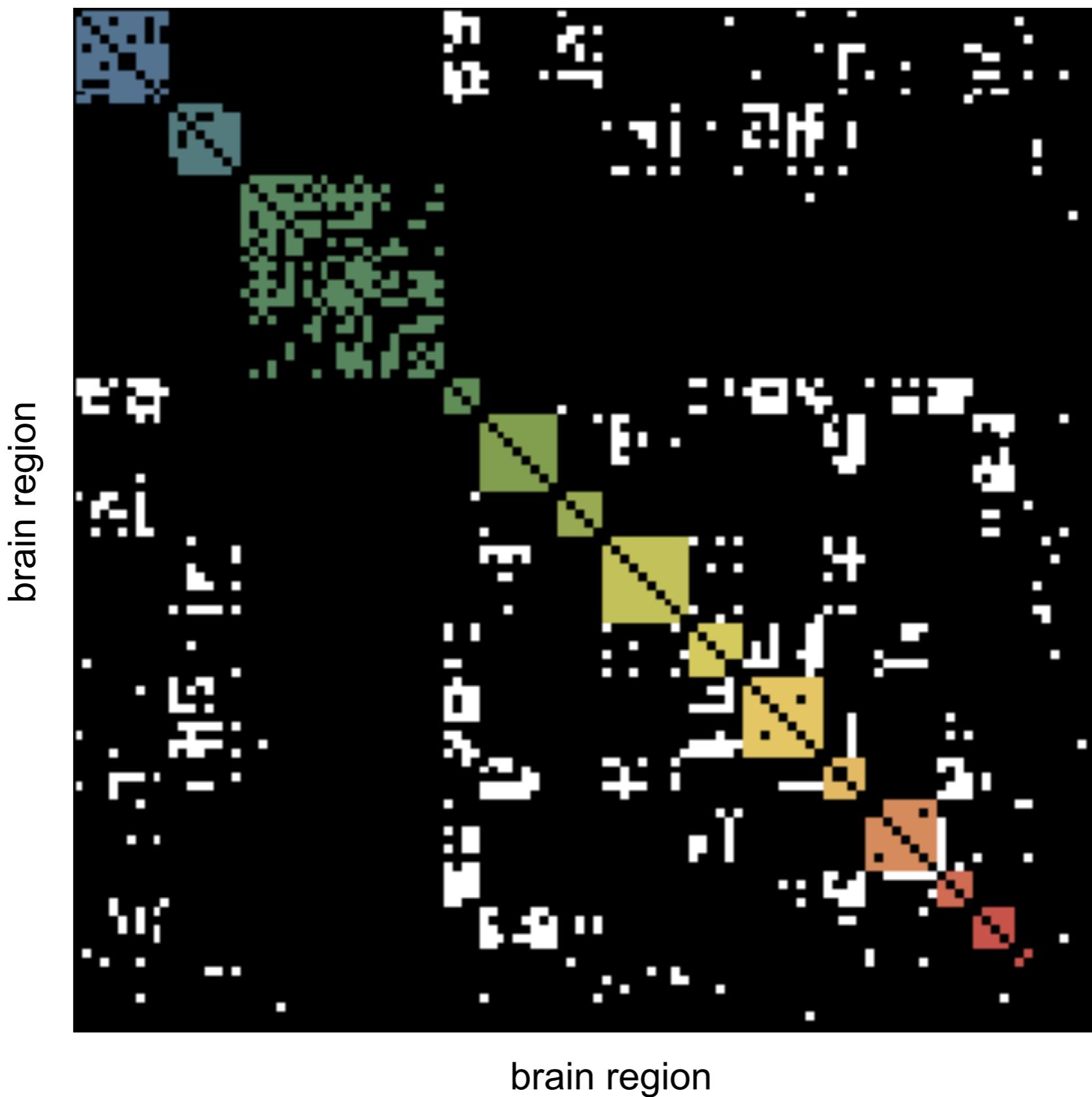


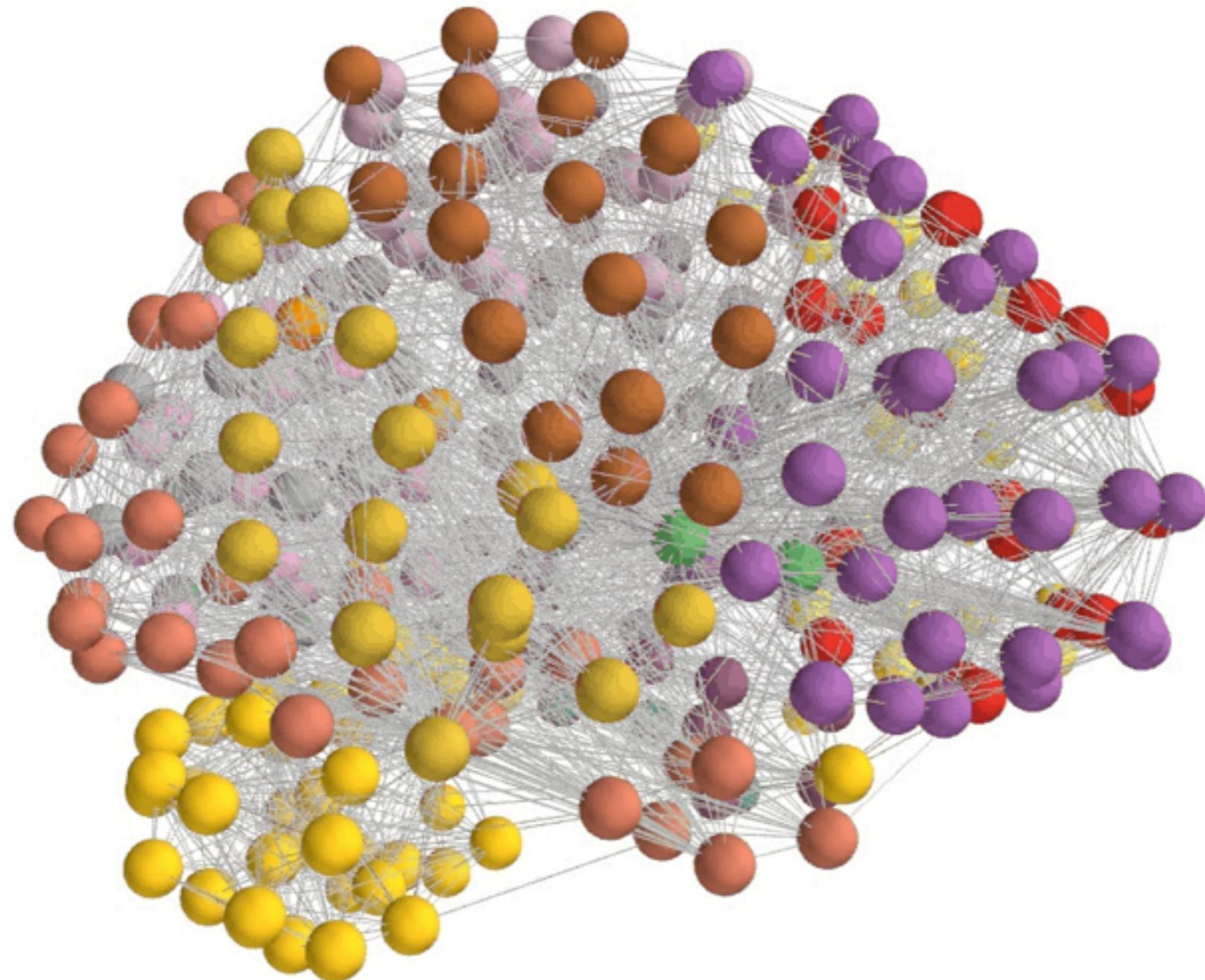
- Use the Bayesian framework to achieve connectivity-based parcellation
- Using the infinite relational model as a prior in our model
- Learns the structure of  $\mathbf{A}$  together with a clustering  $\mathbf{Z}$  ( $z_{ij}$  : node  $i$  is assigned cluster  $j$ ):

$$P(\mathbf{A}, \mathbf{Z} \mid \mathbf{N}, \xi) \propto P(\mathbf{N} \mid \mathbf{A}, a^+, a^-) P(\mathbf{A} \mid \mathbf{Z}, \beta^+, \beta^-) P(\mathbf{Z} \mid \alpha)$$

with  $\xi = \{\alpha, \beta^+, \beta^-, a^+, a^-\}$  hyperparameters.

- Running this on 1000 node random parcels
- Comparison with other parcellation methods

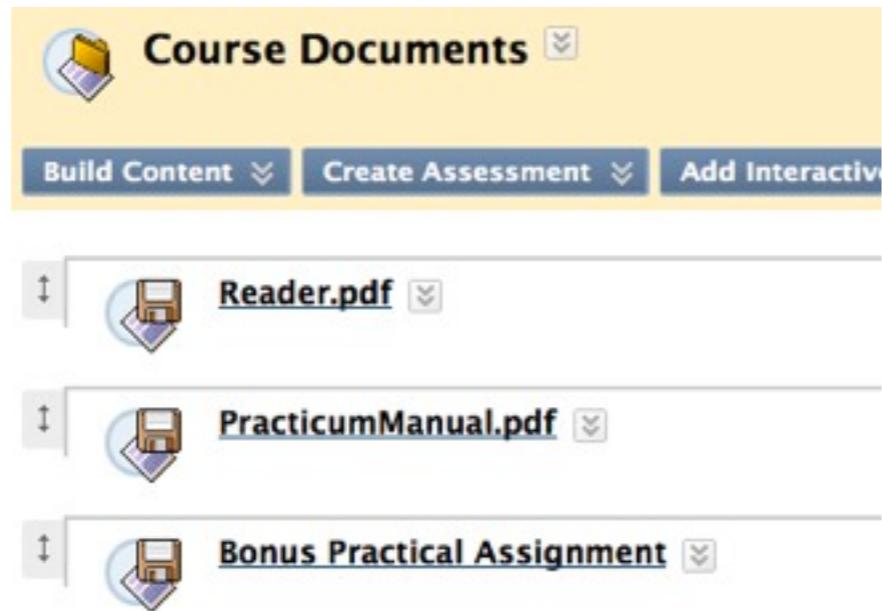






- Bayesian functional connectivity analysis
- Changes in connectivity over time
- Bayesian population stratification
- Bayesian connectivity-based prediction
- etc.

- Bonus practical assignment



- Next week: two guest lectures
- Week after: final lecture
- Last option to ask questions about course material
- Check practical and essay deadlines!