

NETWORKS

BST228

TODAY

- Review of ignorability.
- Statistical network models: Erdős–Rényi model, (soft) random geometric graphs and latent space models, stochastic block models, ...
- Mechanistic network models (maybe)

Speaker notes

- Unlike previous topics we studied in depth, this is an overview of networks without going into the technical details.

IGNORABILITY FOR DATA MAR

Suppose data are MAR such that $p(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}}, \boldsymbol{\phi}) = p(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi})$ and the parameters of our model $\boldsymbol{\theta}$ and of the missingness model $\boldsymbol{\phi}$ are independent a priori, i.e., $p(\boldsymbol{\theta}, \boldsymbol{\phi}) = p(\boldsymbol{\theta})p(\boldsymbol{\phi})$. Using the expression for the joint posterior,

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}) &= \int d\boldsymbol{\phi} d\mathbf{y}_{\text{mis}} p(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}) \\ &\propto \int d\boldsymbol{\phi} d\mathbf{y}_{\text{mis}} p(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi}) p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) p(\boldsymbol{\phi}). \end{aligned}$$

The integral with respect to \mathbf{y}_{mis} and $\boldsymbol{\phi}$ is separable such that

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}, \mathbf{I}) &\propto \left[\int d\boldsymbol{\phi} p(\mathbf{I} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\phi}) p(\boldsymbol{\phi}) \right] \left[\int d\mathbf{y}_{\text{mis}} p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \right] \\ &\propto \int d\mathbf{y}_{\text{mis}} p(\mathbf{y}_{\text{obs}}, \mathbf{y}_{\text{mis}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto p(\mathbf{y}_{\text{obs}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \end{aligned}$$

The second line follows because the first integral does not depend on $\boldsymbol{\theta}$ and can be absorbed by the proportionality. The third line follows by the law of total probability.

The missing data model does not appear in the posterior, and we say the missing data mechanism is *ignorable*. This means we do not need to model $p(\mathbf{I} \mid \dots)$, but it does *not* mean that we can ignore that data are missing.

SO WHY NETWORKS?

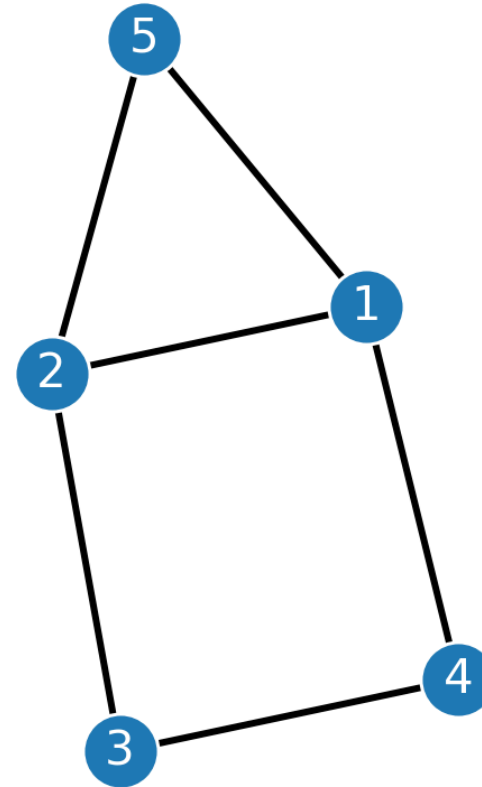
- Social networks and isolation.
- Contact networks and infection.
- Referral networks for specialist care.
- ...

Speaker notes

- The surgeon general declared social isolation a public health emergency. Social isolation is fundamentally a network question.
- Disease transmission in the real world does not follow common assumptions of homogeneous mixing, and the structure of the network affects transmission dynamics.
- Who gets referred to which doctor can have a health impact. Do you get care as fast as possible, or does the health care system prioritize making connections to in-network providers?

WHAT IS A NETWORK?

- A graph has n nodes with labels $i \in \{1, \dots, n\}$.
- If two nodes i and j are connected, $A_{ij} = 1$ in the adjacency matrix \mathbf{A} .
If not connected, $A_{ij} = 0$.



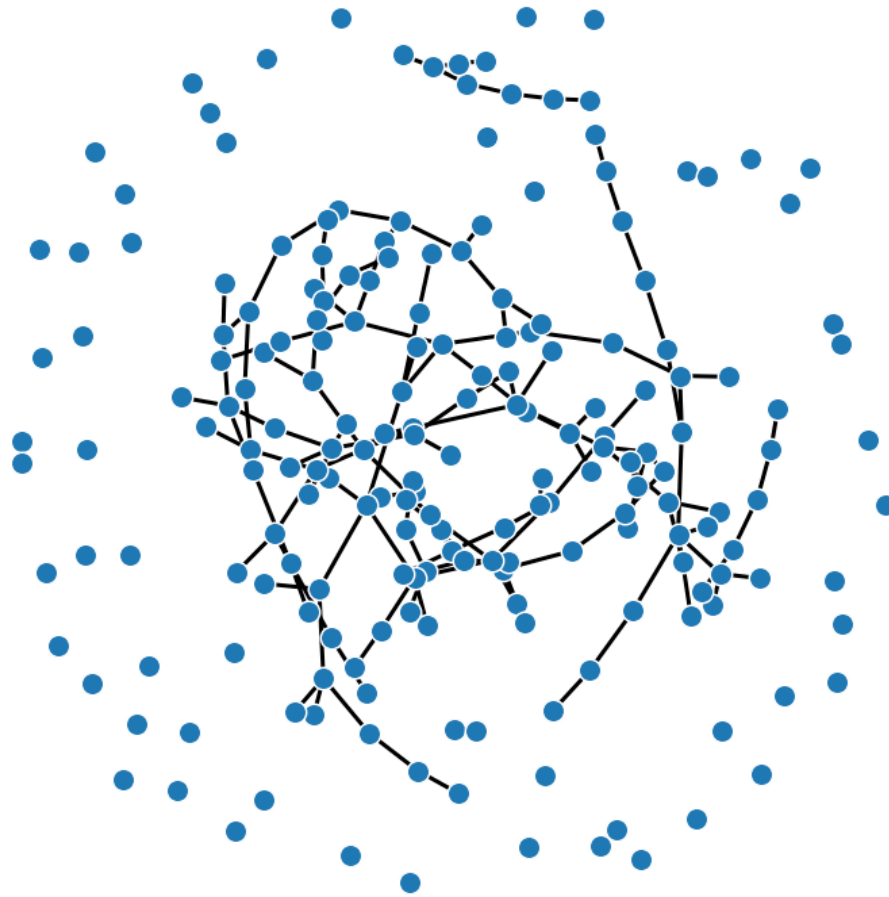
THE SIMPLEST OF MODELS: ERDŐS–RÉNYI

In an Erdős–Rényi model, all edges are independent and present with probability λ such that

$$p(\mathbf{A} \mid \lambda) = \prod_{i,j=1}^n \lambda^{A_{ij}} (1 - \lambda)^{1 - A_{ij}} .$$

Speaker notes

- ER models were originally developed in the context of statistical mechanics and much of the literature considers the sensitivity of network properties to the connection probability λ . E.g., how large should λ be for most of the network to be connected. How does clustering change with increasing λ ?
- From an inference perspective, we can use a beta prior for the connection probability λ because the data comprise i.i.d. binary outcomes.



Speaker notes

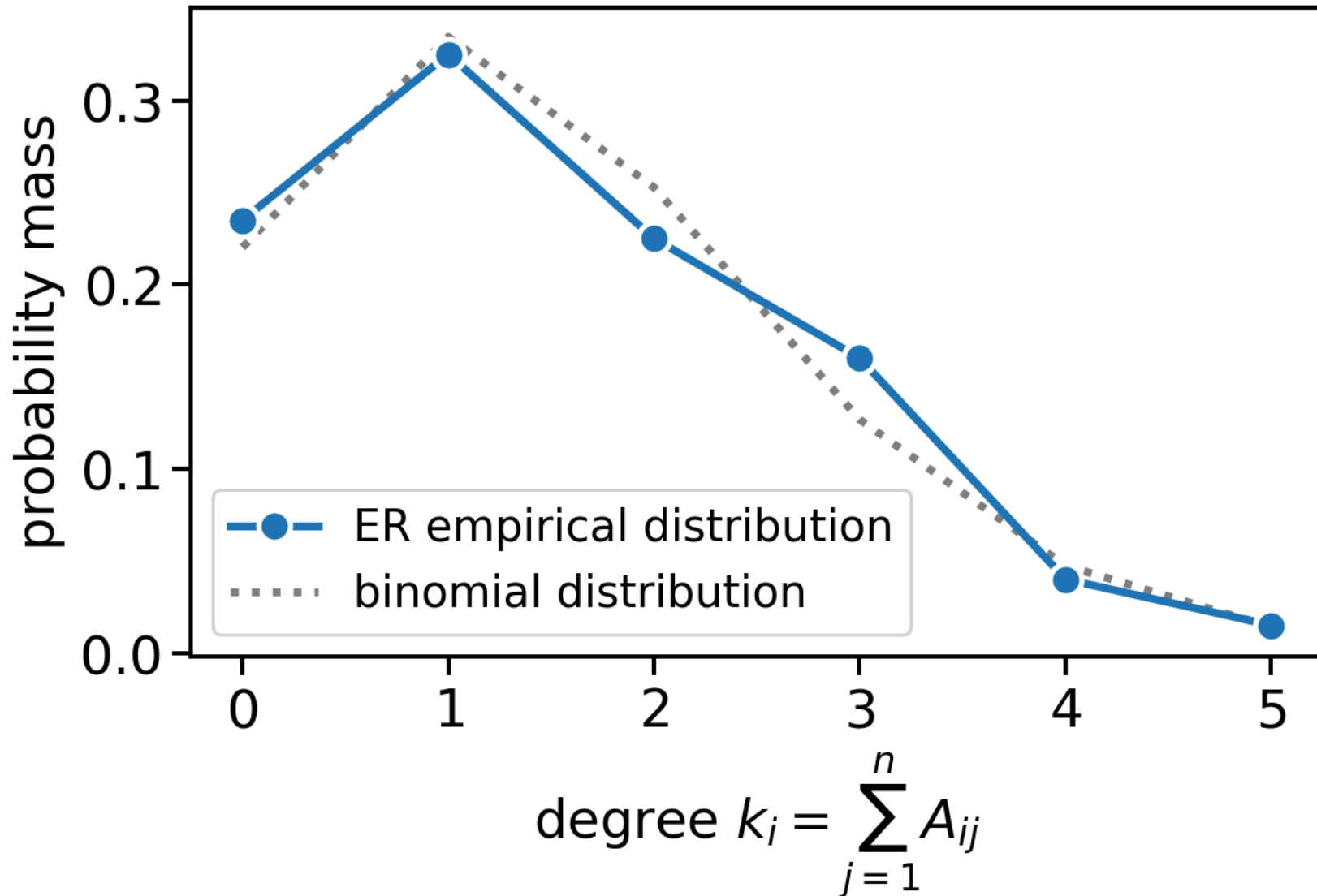
- An illustration of an ER graph with 200 nodes and connection probability just high-enough for a giant connected component to form.
- At this still relatively low connection probability, a non-negligible number of nodes remain isolated.

DEGREE DISTRIBUTION OF THE ERDŐS-RÉNYI MODEL

- For an Erdos Renyi model, the degree is binomially distributed because the number of connections is a sum of independent Bernoulli trials, i.e., $k_i \sim \text{Binomial}(n - 1, \lambda)$. The term $n - 1$ arises because we assume nodes cannot connect to themselves.

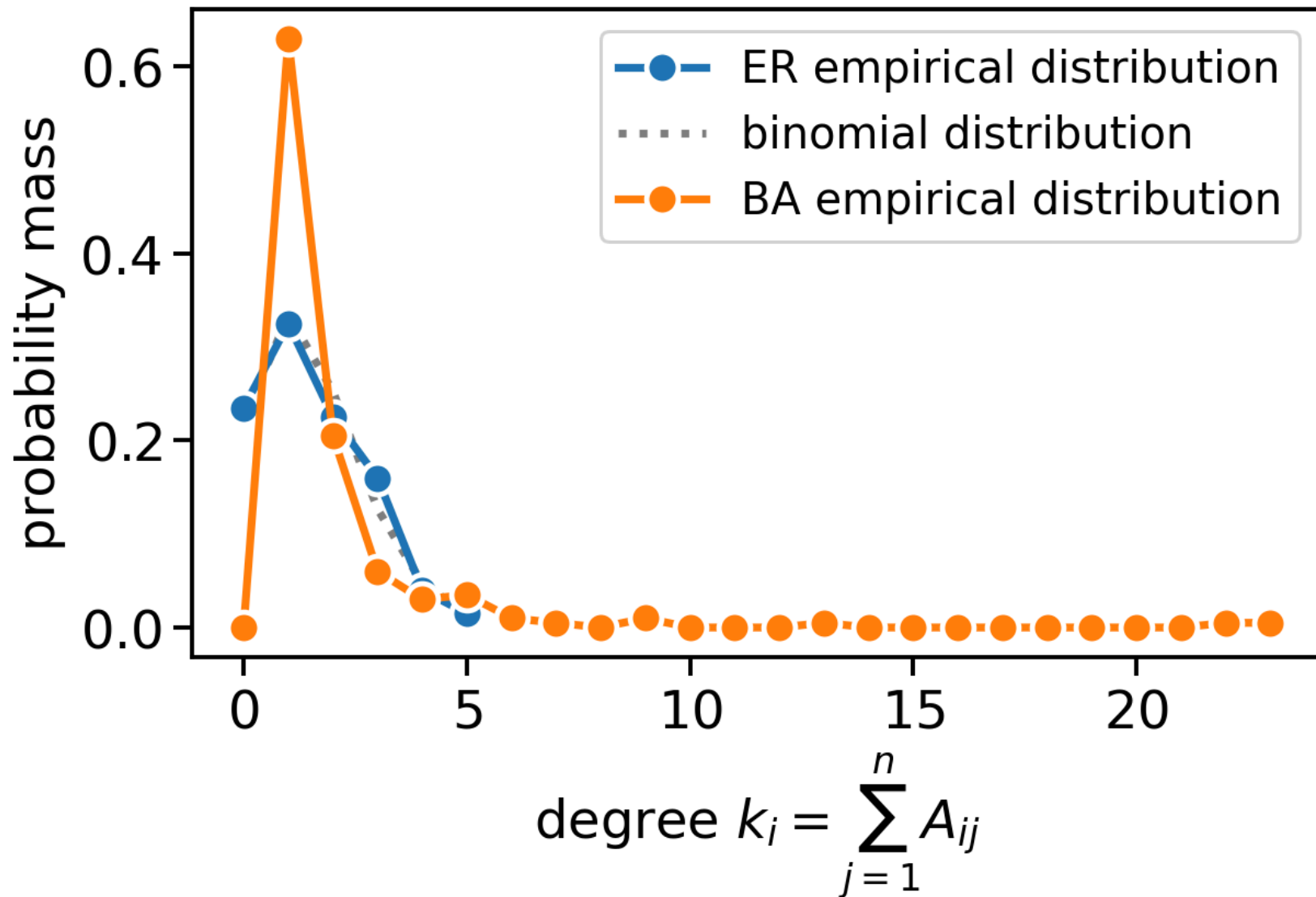
The degree of node i is its number of connections

$$k_i = \sum_{j=1}^n A_{ij}.$$



Speaker notes

- The empirical degree distribution of the network in the previous figure is shown in blue.
- It agrees with our expectation of a binomial degree distribution with appropriate parameters.
- This degree distribution is not realistic because observed degree distributions are often very heavy-tailed. E.g., celebrities on social media sometimes have millions of followers whereas most accounts only have a few.
- Heavy-tailed degree distributions are particularly important for sexually transmitted infections where “hubs” can lead to large-scale spreading events.
- While amenable to theoretical analysis, the ER model is not particularly useful in public health applications.



Speaker notes

- Like the previous figure, a binomial degree distribution is shown in blue.
- The distribution in orange is the degree of a [Barabási–Albert model](#) which is a classic model to reproduce realistic degree distributions.

CONDITIONALLY INDEPENDENT EDGE MODELS

In a *conditionally* independent edge model,

$$p(\mathbf{A} \mid \mathbf{X}) = \prod_{i,j=1}^n p(A_{ij} \mid \mathbf{x}_i, \mathbf{x}_j),$$

where \mathbf{X} is a matrix of node features.

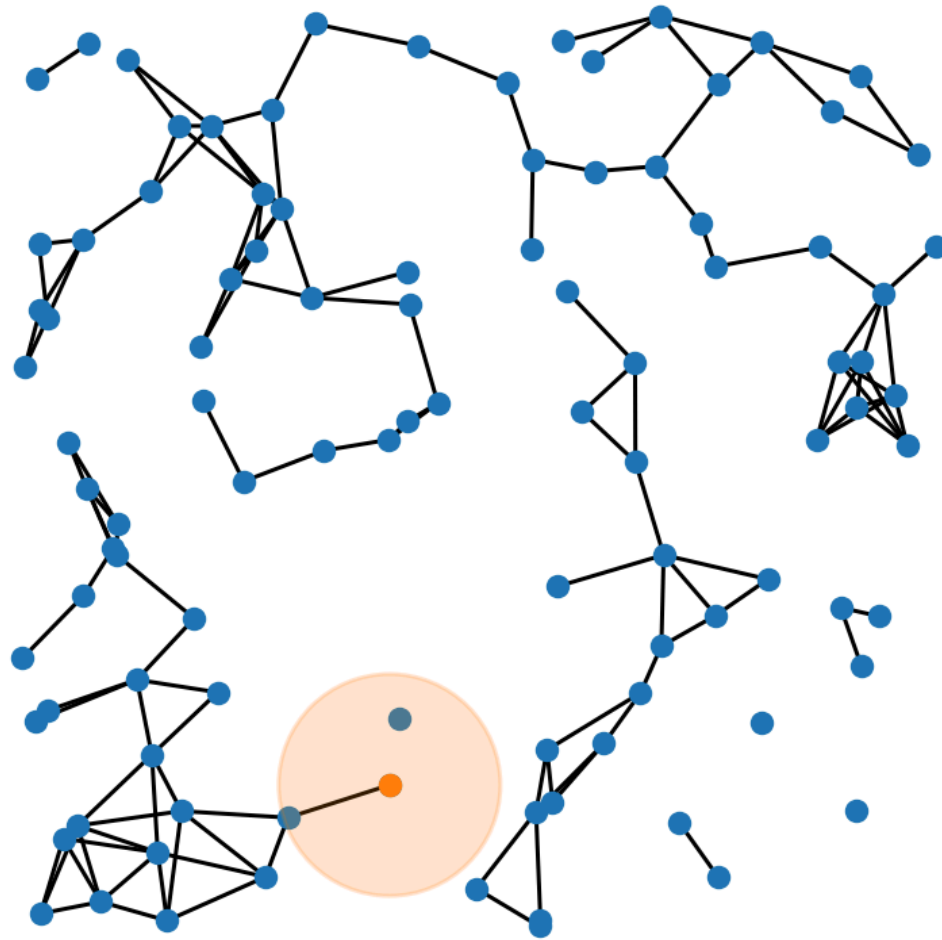
Speaker notes

- This generalizes the i.i.d. edges in the ER model.
- We can incorporate more interesting structure because the presence or absence of edges may depend on node attributes or features.
- E.g., in heterosexual contact networks, the probability to connect depends on sex and generates a bipartite network.
- In contacts relevant for airborne transmission, the likelihood of interaction may depend on age, occupation, ...

- Examples of random geometric graphs include the [internet of things](#) where devices may only be able to interact if they are within a certain distance.

RANDOM GEOMETRIC GRAPHS

Each node i has coordinates \mathbf{x}_i in a q -dimensional space, and nodes are connected with probability λ if $|\mathbf{x}_i - \mathbf{x}_j| < \theta$ for some threshold θ .



Speaker notes

- The figure shows a realization of a soft random geometric graph with $\lambda = 0.8$. Nodes are not connected if they are too far apart. If they are close enough (within the distance indicated by the orange circle), they may connect.

LOGISTIC REGRESSION FOR NETWORKS

$$A_{ij} \mid \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\lambda_{ij})$$
$$\text{logit}(\lambda_{ij}) = a + \mathbf{b}^\top \mathbf{f}(\mathbf{x}_i, \mathbf{x}_j),$$

where the intercept a controls overall edge density, \mathbf{b} are regression coefficients, and \mathbf{f} maps node attributes \mathbf{x} to features.

Speaker notes

- An example of a logistic regression model for networks is connections forming based on demographics. Demographics are represented as node attributes \mathbf{x} , and features for the regression may include, for example, the age difference between individuals.

SOCIAL SEPARATION

Consider the social separation

$$\psi(\mathbf{x}_i, \mathbf{x}_j) = \text{logit}(\lambda_{jj}) - \text{logit}(\lambda_{ij}),$$

which measures how isolated i is from j .

Speaker notes

- This measure of social separation is based on [Hoffmann and Jones \(2020\)](#).
- If the probability for connection between nodes at different locations in the space is low (second term), the separation is large (remember that logit (small probability) = large negative number).
- If there is no dependence on attributes, $\psi = 0$.
- If there is heterophily, i.e., a tendency for nodes to connect with others who are different, there is negative separation.

SOCIAL ISOLATION

Consider the social isolation

$$\phi(\mathbf{x}_i) = \int d\mathbf{x}_j \psi(\mathbf{x}_i, \mathbf{x}_j) p(\mathbf{x}_j)$$

which measures how separated i is from others on average.

Speaker notes

- Social separation is not necessarily indicative of social isolation.
- Members of a majority group may have large separation to some members of society, but their average distance to others is small because they belong to the majority.
- Members of a minority group may have the same large separation to some members of society, but their average distance to others is large because they do not belong to the majority.
- Isolation can be very different depending on who you are even if separations are the same.

SOCIAL STRAIN

Consider the social strain

$$\Phi = \int d\mathbf{x}_i p(\mathbf{x}_i) \phi(\mathbf{x}_i)$$

which measures how isolated people are on average.

Speaker notes

- This is a utilitarian definition of social strain because we focus on average social isolation.
- One might argue that other measures of social strain are more appropriate.
- Maybe we instead want to minimize the isolation of the most isolated person.
- Maybe we want to instead make sure the 75th percentile of isolation is below some threshold.
- We would have to define another measure in this case.

INFERENCE FROM SURVEY DATA

- We ask respondents about their demographics and also about the demographics of their “friends.”
- We treat these pairs as “cases” and randomly pair respondents as “controls.”
- If we properly account for the missing data (all the present or absent edges we didn’t observe), we can infer the parameters of the logistic regression model.

Speaker notes

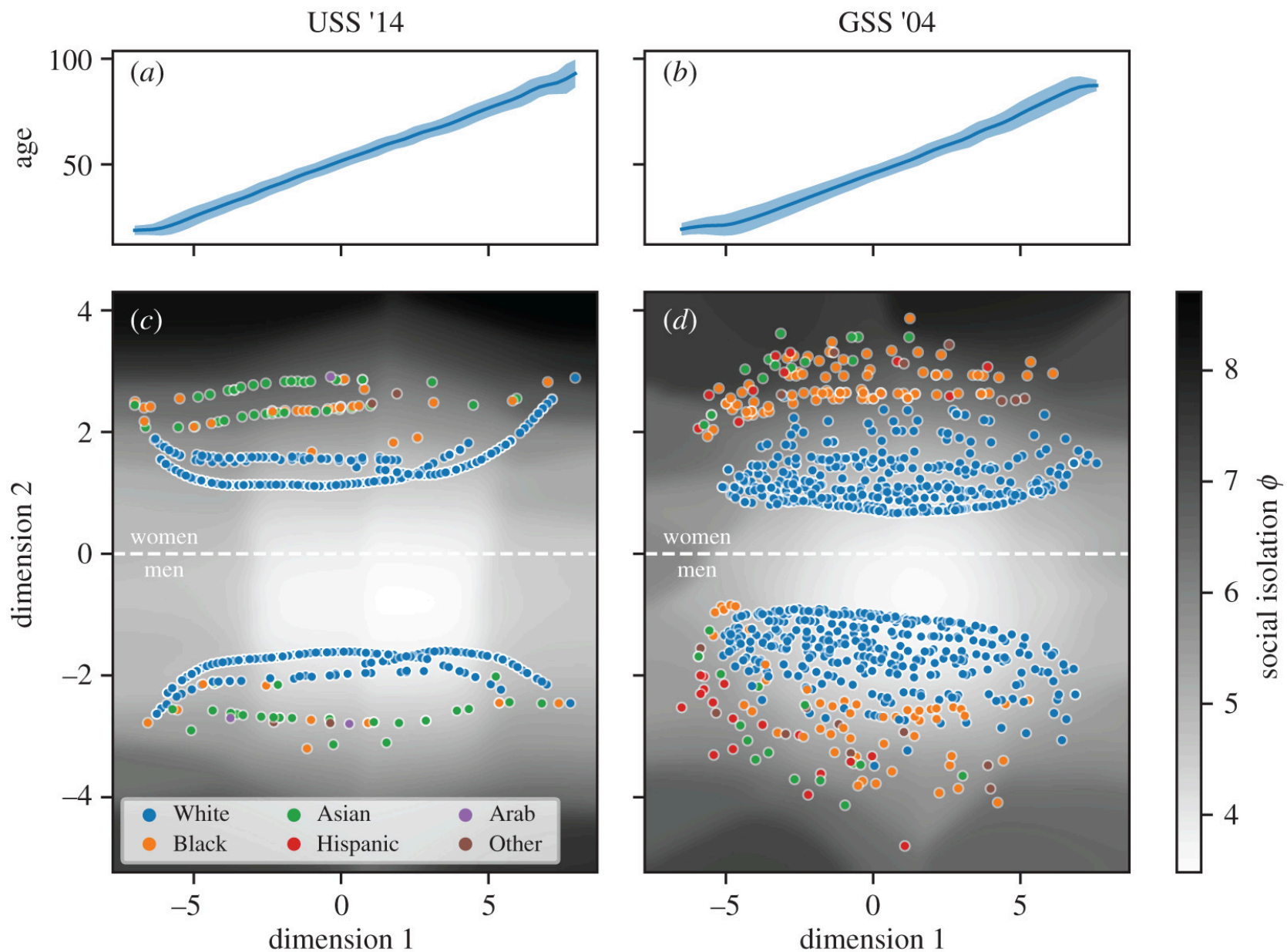
- Connecting back to Bayesian inference, a nice consequence of our model-based approach is that we naturally get uncertainty quantification. But we first need to infer the model parameters.
- The definition of friend depends on the exact question wording and how respondents interpret the question. Questions could include “someone you discuss important matters with” or simply “close friend.”
- Kin are usually excluded from analyses.

MODEL-BASED UNCERTAINTY QUANTIFICATION

- Because ψ , ϕ and Φ are defined in terms of the model, we automatically get uncertainty quantification.
- Unlike regression coefficients, these quantities have real-world implications and interpretation.
- Even for non-parametric logistic regression, the notion of social separation, isolation, and strain remain valid and useful.

Speaker notes

- For surveys in the UK (left) and US (right), we fitted a logistic network regression model using demographic differences as features.
- We then embedded individuals in a low-dimensional space based on the separation between them using standard dimensionality reduction techniques (MDS).
- In this space, there is clear separation between sexes and also a strong effect of age on friendship formation.

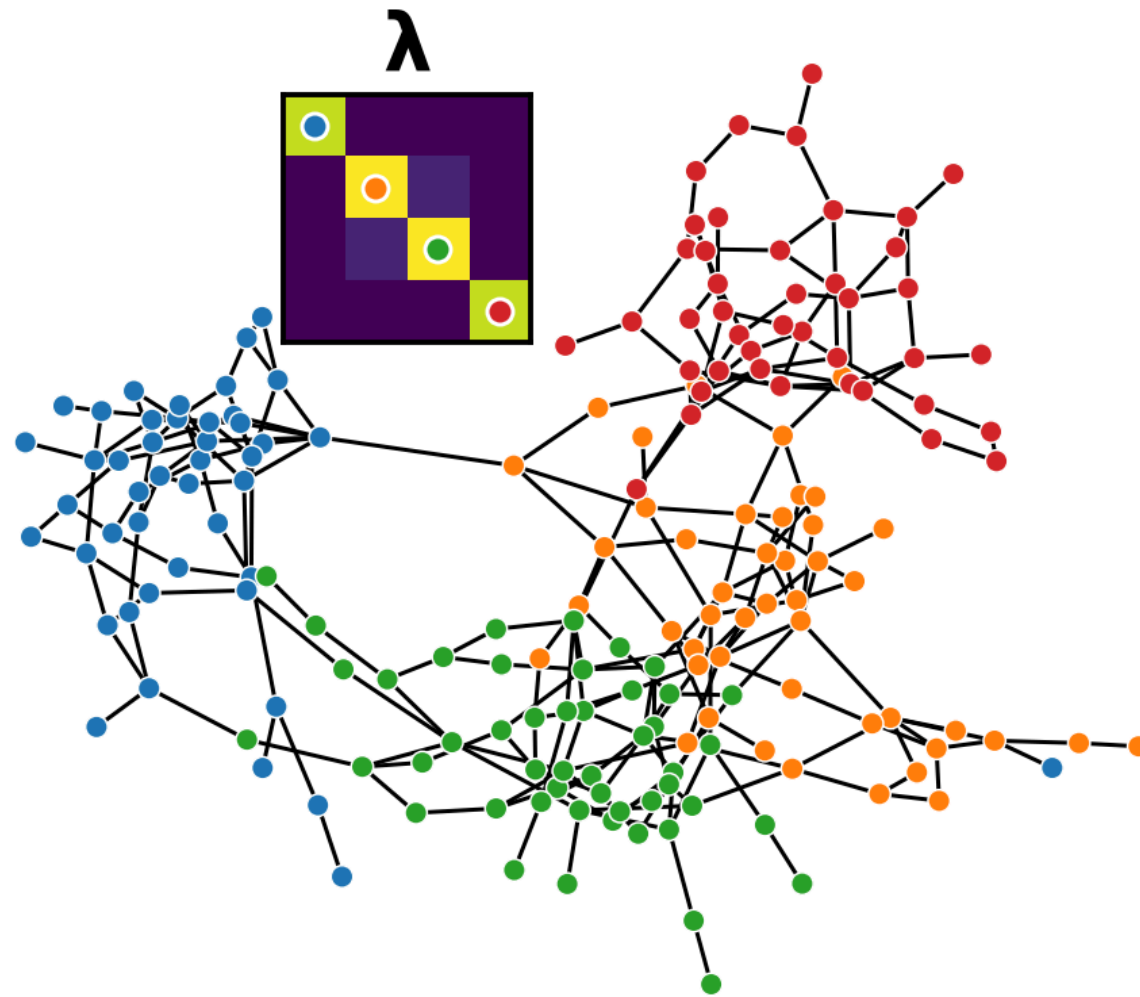


STOCHASTIC BLOCK MODELS (SBMS)

For a stochastic block model, features $x \in \{1, \dots, q\}$ indicate block membership and

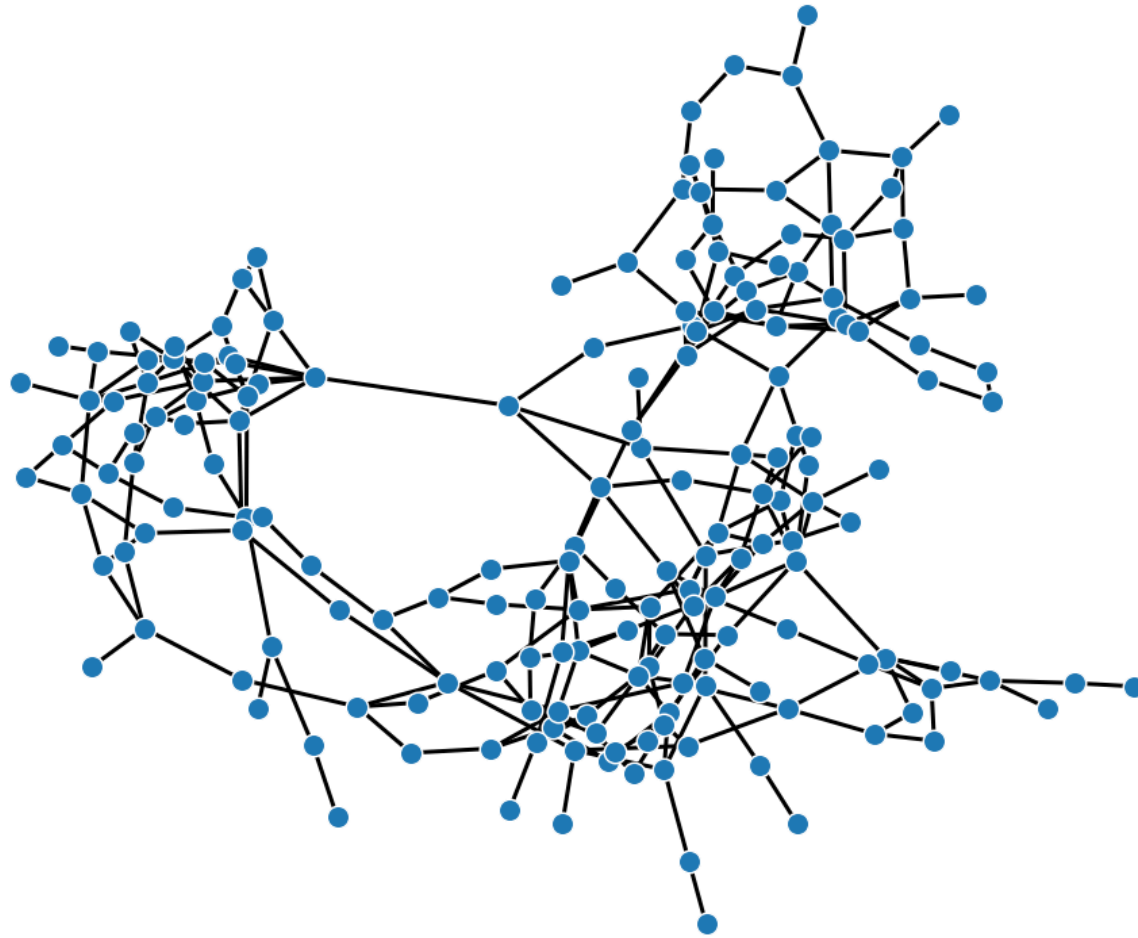
$$A_{ij} \sim \text{Bernoulli}(\lambda_{x_i x_j}),$$

where $\boldsymbol{\lambda}$ is a $q \times q$ matrix of connection probabilities.



Speaker notes

- The figure shows a realization of a stochastic block model with strong community structure (i.e., connections within communities are more likely than between communities).
- The corresponding matrix of connection probabilities λ is shown as an inset heatmap.



Speaker notes

- Empirical network data often do not include ground truth community labels.
- We may want to find clusters—a task known as community detection. E.g., proteins that interact, genes that are co-expressed, clusters of financial transactions, or asset classes in financial markets can form communities.

SBMS FOR COMMUNITY DETECTION

We treat both the block memberships \mathbf{x} and connection probabilities $\boldsymbol{\lambda}$ as parameters of the model:

$$A_{ij} \sim \text{Bernoulli}(\lambda_{x_i x_j})$$

$$\lambda_{rs} \sim \text{Beta}(1, 1)$$

$$x_i \sim \text{Uniform}(1, q)$$

Speaker notes

- Sometimes people introduce another parameter $\boldsymbol{\rho}$ representing the size of clusters. Then $x_i \sim \text{Categorical}\{\rho_1, \dots, \rho_q\}$.
- This model favors communities with similar sizes a priori.

GIBBS SAMPLER FOR COMMUNITY DETECTION

We can sample from the conditional distributions

$$\mathbf{x} \mid \mathbf{A}, \boldsymbol{\lambda} \sim \text{Categorical}(\dots)$$

$$\boldsymbol{\lambda} \mid \mathbf{x}, \mathbf{A} \sim \text{Beta}(\dots)$$

to obtain a Gibbs sampler for community detection.

Speaker notes

- Stan cannot solve this problem because Stan relies on partial derivatives of the density of the target distribution with respect to parameters of the model. That is not possible for discrete parameters like community membership, and Gibbs samplers are a more suitable tool.

APPLICATIONS OF NETWORK MODELS

- Answering science questions using domain-specific models, e.g., studying gene co-expression.
- Predict links using $p(\mathbf{A}^{\text{pred}} \mid \mathbf{A})$. Or identify surprising edges.
- Predict attributes using $p(\mathbf{x}^{\text{pred}} \mid \mathbf{A}, \mathbf{x})$.
- Sensitivity analysis for infectious disease transmission dynamics by replicating networks using $p(\mathbf{A}^{\text{repl}} \mid \mathbf{A})$

Speaker notes

- Network science often implicitly assumes that the adjacency matrix \mathbf{A} is measured without error, i.e., it represents the *true* network. Even if such an unambiguous truth exists, we usually have measurement error. It can be helpful to identify edges that are missing (link prediction) or ones that should not be present (identifying spurious edges).
- Alternatively, we can use network information to constrain attributes of individuals if we know how people connect based on attributes.
- When estimating key quantities like epidemic size, we often repeat stochastic simulations to assess variability but keep the network fixed. Replicating networks and running simulations can account for the additional uncertainty due to network variability.

MECHANISTIC NETWORK MODELS

- All of the previous models are *statistical* network models with tractable likelihood.
- But they cannot (easily) reproduce a number of features of real-world data (like degree distributions or temporal effects).
- They also pose a challenge for assessing policy interventions.
- Another approach is to use *mechanistic* network models whose evolution is described by a set of rules or mechanisms.

Speaker notes

- Even if we spend time in the same classroom together and one of us catches an infection, the rest of us are safe from airborne transmission because of the direction of time. Many network models have no notion of time or causality.
- There are temporal statistical network models, but mechanistic network models are often more intuitive.
- Mechanisms vary from extremely simple (pick a person at random and form a connection with probability ρ , remove existing connections with probability σ) to very complex (whole population simulations).

SEXUAL CONTACT NETWORK MODELS

```
1 n <- 1000 # Expected number of nodes.  
2 mu <- 0.001 # Probability of leaving population.  
3 rho <- 0.1 # Prob. to form relationship.  
4 sigma <- 0.05 # Prob. for relationship to dissolve.  
5 omega0 <- 0.2 # Prob. for casual contact for singles.  
6 omega1 <- 0.1 # Prob. for casual contact for paired.  
7 T <- 10000 # Number of simulation steps.
```

Speaker notes

- This code block defines parameters for a mechanistic network model for men who have sex with men from [10.1016/j.epidem.2019.02.001](https://doi.org/10.1016/j.epidem.2019.02.001).

SEXUAL CONTACT NETWORK MODELS

```
1 graph <- empty()
2 for (t in 1:T) {
3   graph <- remove_nodes_with_prob(graph, mu)
4   graph <- remove_steady_with_prob(graph, sigma)
5   graph <- remove_casual(graph)
6
7   n_new_nodes <- rpoisson(1, mu * n)
8   graph <- add_new_nodes(graph, n_new_nodes)
9   singles <- get_nodes_with_zero_degree(graph)
10  graph <- add_steady_with_prob(graph, singles, rho)
11  singles <- get_nodes_with_zero_degree(graph)
12  graph <- add_casual_with_prob(
13    graph, singles, omega0)
14  paired <- get_nodes_with_nonzero_degree(graph)
15  graph <- add_casual_with_prob(paired, omega1)
16
17  # Disease dynamics ...
18 }
```

Speaker notes

- This pseudo-code does not run but conveys the gist of a mechanistic network model.
- The first block removes nodes (population turnover), steady relationships (relationships breaking up), and all casual contacts (we treat them as instantaneous for each time step).
- We then add new nodes to keep a constant expected population size, create new steady relationships, and add casual contacts.
- Separately, we may run disease transmission dynamics on the network (omitted here because we are focusing on networks).

BENEFITS

- We can easily run sensitivity analyses.
- We can simulate intervention policies.
- The model is easily interpretable.

Speaker notes

- How do the results change if, for example, there is more population turnover or the likelihood of casual interactions decreases?
- Intervention on the contact network may be ethically complex, e.g., should we, as policy makers, try to reduce casual contacts or reduce concurrency of multiple steady partners?
- Intervention on the transmission dynamics are ethically more straightforward and can be studied using these dynamic network models. E.g., how does pre-exposure prophylaxis affect transmission in light of specific network structure?
- Mechanistic network models are also interpretable and can be developed in collaboration with domain experts.

INFERENCE CHALLENGES

- Even though we use simple building blocks, the likelihood of the model is not tractable.
- Markov chain Monte Carlo methods are not applicable because we cannot evaluate the posterior density.
- Inference for these models is an open challenge and the topic of research in simulation-based inference.

RECAP AND FURTHER TOPICS

- What are networks and why should we care?
- Statistical network models: Erdős–Rényi, (soft) random geometric graphs, conditionally independent edge models, logistic regression, stochastic block models.
- Mechanistic models, their advantages, and challenges.
- Further topics: Directed vs undirected networks, temporal networks, multilayer networks, hypergraphs, ...

Speaker notes

- Directed networks may not be reciprocated (think of follower structure on X/Twitter) whereas undirected networks are always symmetric (think of friendships on Facebook).
- Multilayer networks try to capture the same individuals across different modalities of interaction, e.g., at work, in a social context, at a sports club, on social media, ... Do different modalities complement each other or are they substitutes?
- Networks are inherently pairwise interactions. Hypergraphs consider higher-order interactions, e.g., are interactions amongst three people qualitatively different from three pairwise interactions between the same people?