Lecture 6 (supplemental): Stochastic Block Models

Aaron Clauset
@aaronclauset
Assistant Professor of Computer Science
University of Colorado Boulder
External Faculty, Santa Fe Institute

what is structure?

what is structure?

- makes data different from noise
  - makes a network different from a random graph

what is structure?

- makes data different from noise
  - makes a network different from a random graph
- helps us compress the data
  - describe the network succinctly
  - capture most relevant patterns

what is structure?

- makes data different from noise
  - makes a network different from a random graph
- helps us compress the data
  - describe the network succinctly
  - capture most relevant patterns
- helps us generalize,
  from data we've seen to data we haven't seen:
  i. from one part of network to another
  ii. from one network to others of same type
  iii. from small scale to large scale (coarse-grained structure)
  iv. from past to future (dynamics)

statistical inference

- imagine graph G is drawn from an ensemble or **generative model**: a probability distribution $\Pr(G \mid \theta)$ with parameters θ
- θ can be continuous or discrete; represents structure of graph

statistical inference

- imagine graph G is drawn from an ensemble or **generative model**: a probability distribution $\Pr(G \mid \theta)$ with parameters θ
- θ can be continuous or discrete; represents structure of graph
- inference (MLE): given G, find θ that maximizes $\Pr(G \mid \theta)$
- inference (Bayes): compute or sample from posterior distribution $\Pr(\theta \mid G)$

statistical inference

- imagine graph G is drawn from an ensemble or **generative model**: a probability distribution $\Pr(G \,|\, \theta)$ with parameters θ
- θ can be continuous or discrete; represents structure of graph
- inference (MLE): given G, find θ that maximizes $\Pr(G \,|\, \theta)$
- inference (Bayes): compute or sample from posterior distribution $\Pr(\theta \,|\, G)$

- if θ is partly known, constrain inference and determine the rest
- if G is partly known, infer θ and use $\Pr(G \,|\, \theta)$ to generate the rest
- if model is good fit (application dependent), we can generate synthetic graphs structurally similar to G
- if part of G has low probability under model, flag as possible anomaly

statistical inference

statistical inference = principled approach to learning from data

combines tools from statistics, machine learning, information theory, and statistical physics

quantifies uncertainty

separates the model from the learning

- imagine graph G
  model $\Pr(G\,|\,\theta)$ θ
- θ can be continuous or discrete, represents structure of graph
- inference (MLE): given G, find θ that maximizes $\Pr(G\,|\,\theta)$
- inference (Bayes): compute or sample from posterior distribution $\Pr(\theta\,|\,G)$

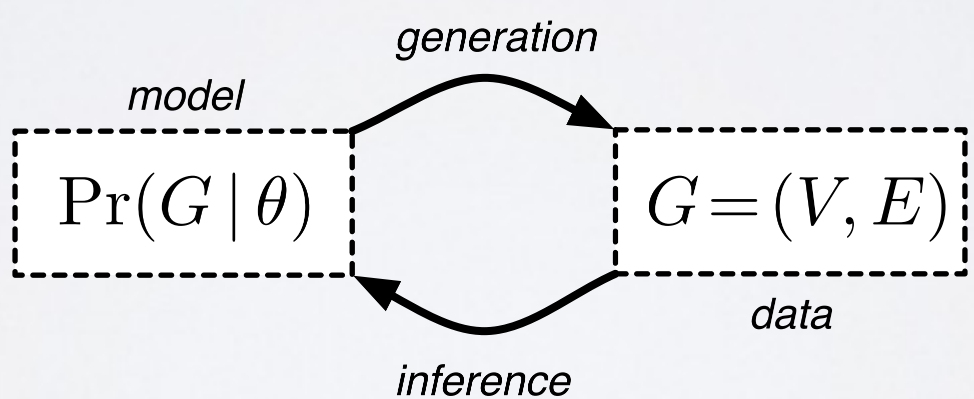
- if θ is partly known, constrain inference and determine the rest
- if G is partly known, infer θ and use $\Pr(G\,|\,\theta)$ to generate the rest
- if model is good fit (application dependent), we can generate synthetic graphs structurally similar to G
- if part of G has low probability under model, flag as possible anomaly

statistical inference: key ideas

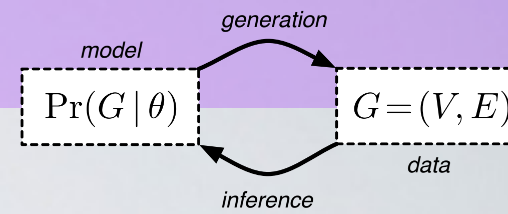
- interpretability
  model parameters have meaning for scientific questions
- auxiliary information
  node & edge attributes, temporal dynamics (beyond static binary graphs)
- scalability
  fast algorithms for fitting models to big data (methods from physics, machine learning)
- model selection
  which model is better? is this model bad? how many communities?
- partial or noisy data
  extrapolation, interpolation, hidden data, missing data
- anomaly detection
  low probability events under generative model

generative models for complex networks

- define a parametric probability distribution over networks $\Pr(G \mid \theta)$
- ***generation*** : given θ, draw G from this distribution
- ***inference*** : given G, choose θ that makes G likely



generative models for complex networks



general form

$$\Pr(G \mid \theta) = \prod_{ij} \Pr(A_{ij} \mid \theta)$$

edge generation function

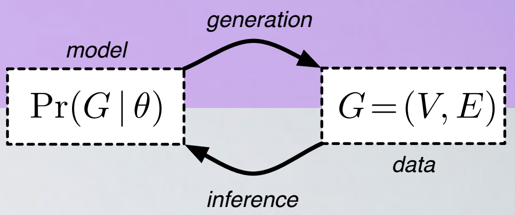
assumptions about "structure" go into $\Pr(A_{ij} \mid \theta)$

consistency $\lim_{n \to \infty} \Pr\left(\hat{\theta} \neq \theta\right) = 0$

requires that edges be conditionally independent [Shalizi, Rinaldo 2011]

two general classes of these models

generative models for complex networks



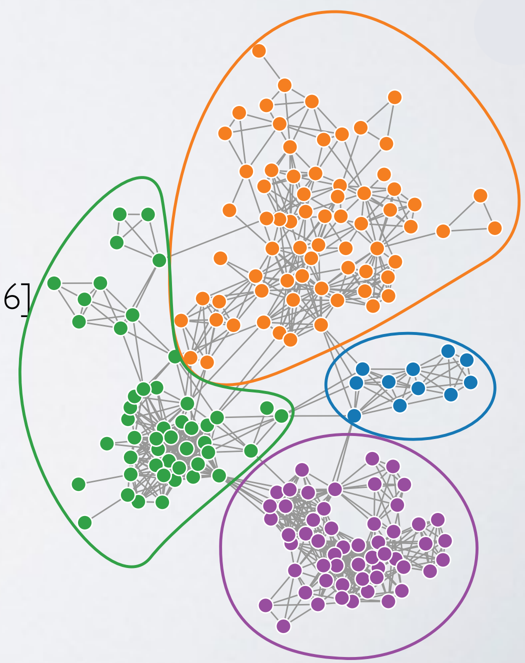
stochastic block models

k types of vertices, $\Pr(A_{ij} \mid M, z)$ depends only on node types z_i, z_j

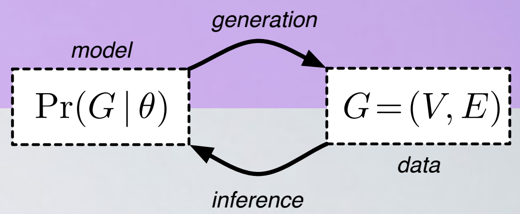
originally invented by sociologists [Holland, Laskey, Leinhardt 1983]

many, many flavors, including

- binomial SBM [Holland et al. 1983, Wang & Wong 1987]
- simple assortative SBM [Hofman & Wiggins 2008]
- mixed-membership SBM [Airoldi et al. 2008]
- hierarchical SBM [Clauset et al. 2006,2008, Peixoto 2014]
- fractal SBM [Leskovec et al. 2005]
- infinite relational model [Kemp et al. 2006]
- degree-corrected SBM [Karrer & Newman 2011]
- SBM + topic models [Ball et al. 2011]
- SBM + vertex covariates [Mariadassou et al. 2010, Newman & Clauset 2016]
- SBM + edge weights [Aicher et al. 2013,2014, Peixoto 2015]
- bipartite SBM [Larremore et al. 2014]
- multilayer SBM [Peixoto 2015, Valles-Catata et al. 2016]
- and many others



generative models for complex networks



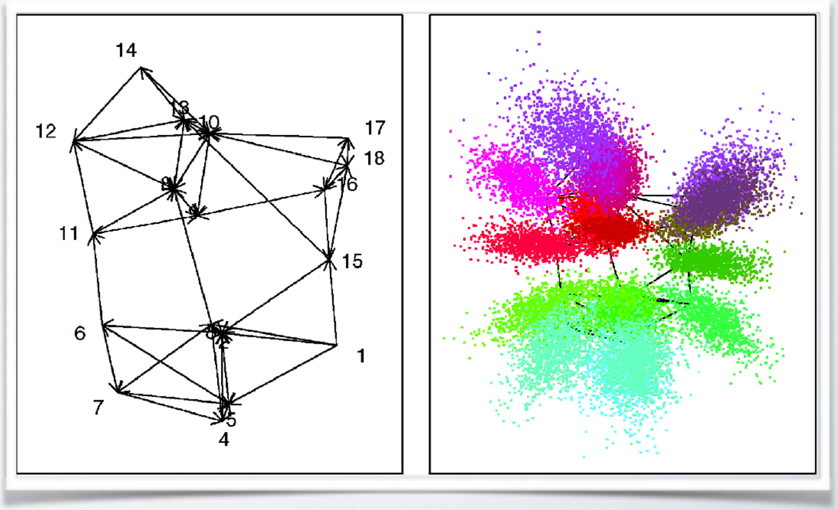
latent space models

nodes live in a latent space, $\Pr(A_{ij} \mid f(x_i, x_j))$ depends only on vertex-vertex proximity

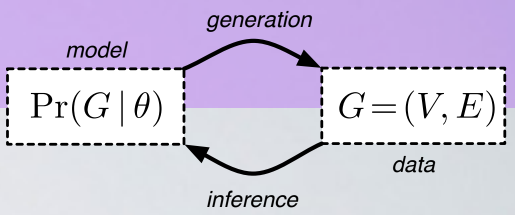
originally invented by statisticians [Hoff, Raftery, Handcock 2002]

many, many flavors, including

- logistic function on vertex features [Hoff et al. 2002]
- social status / ranking [Ball, Newman 2013]
- nonparametric metadata relations [Kim et al. 2012]
- multiplicative attribute graphs [Kim & Leskovec 2010]
- nonparametric latent feature model [Miller et al. 2009]
- infinite multiple memberships [Morup et al. 2011]
- ecological niche model [Williams et al. 2010]
- hyperbolic latent spaces [Boguna et al. 2010]
- and many others

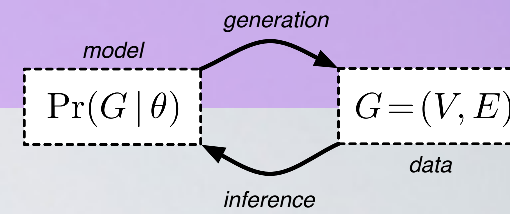


opportunities and challenges



- richly annotated data
  - edge weights, node attributes, time, etc.
  - = new classes of generative models
- generalize from $n = 1$ to ensemble
  - useful for modeling checking, simulating other processes, etc.
- many familiar techniques
  - frequentist and Bayesian frameworks
  - makes probabilistic statements about observations, models
  - predicting missing links $\approx$ leave-*k*-out cross validation
  - approximate inference techniques (EM, VB, BP, etc.)
  - sampling techniques (MCMC, Gibbs, etc.)
- learn from partial or noisy data
  - extrapolation, interpolation, hidden data, missing data

opportunities and challenges

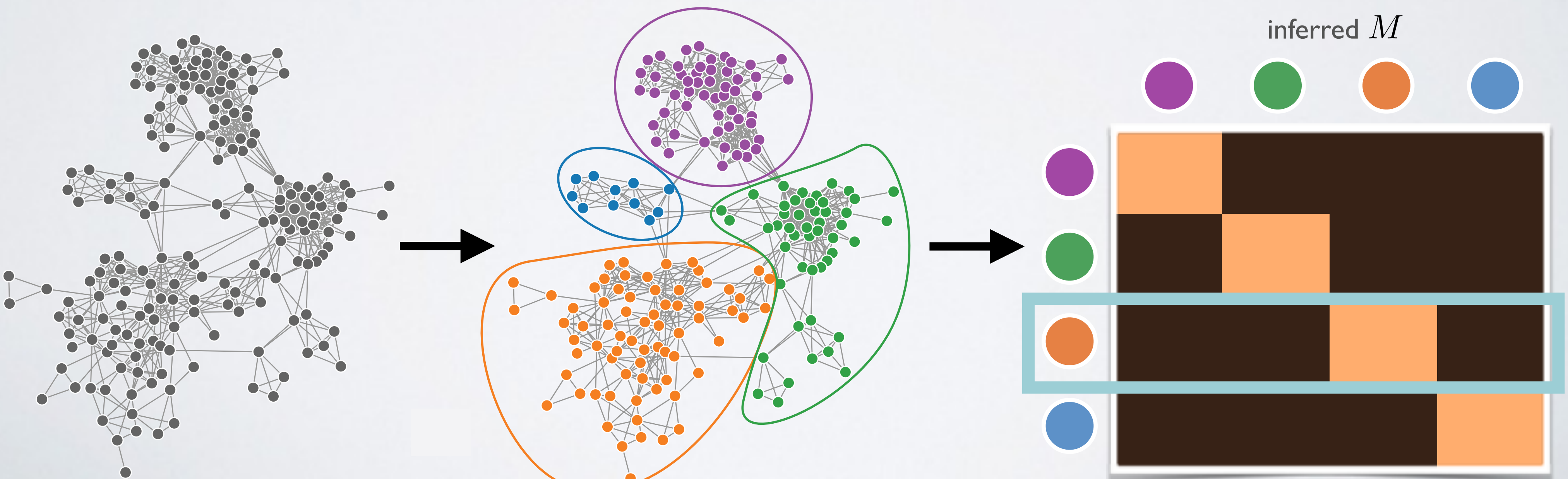


- only two classes of models
  - stochastic block models (categorical latent variables)
  - latent space models (ordinal / continuous latent variables)
- bootstrap / resampling for network data
  - critical missing piece
  - depends on what is independent in the data
- model comparison
  - naive AIC, BIC, marginalization, LRT can be wrong for networks
  - what is goal of modeling: realistic representation or accurate prediction?
- model assessment / checking?
  - how do we know a model has done well? what do we check?
- what is *v*-fold cross-validation for networks?
  - Omit n^2/v edges? Omit n/v nodes? What?

the stochastic block model

- each vertex i has type $z_i \in \{1, \ldots, k\}$ (k vertex types or groups)
- stochastic block matrix M of group-level connection probabilities
- probability that i, j are connected = M_{z_i, z_j}

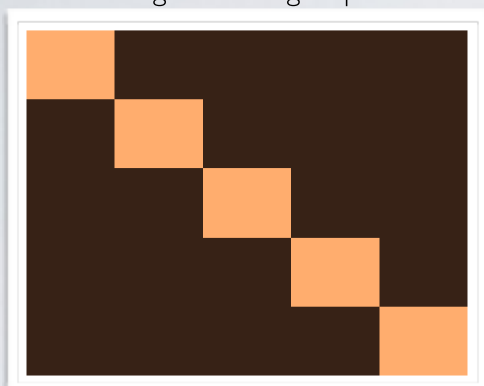
community = vertices with same pattern of inter-community connections

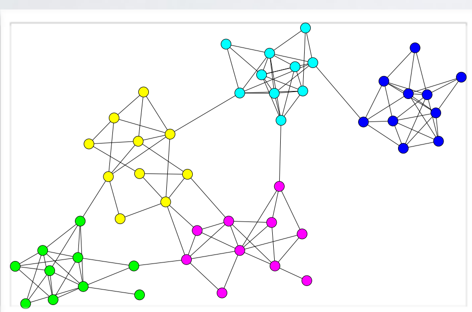


the stochastic block model

assortative

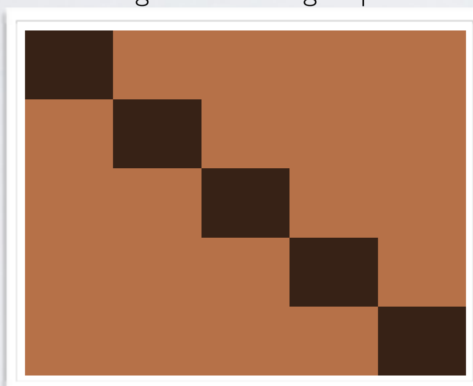
edges within groups

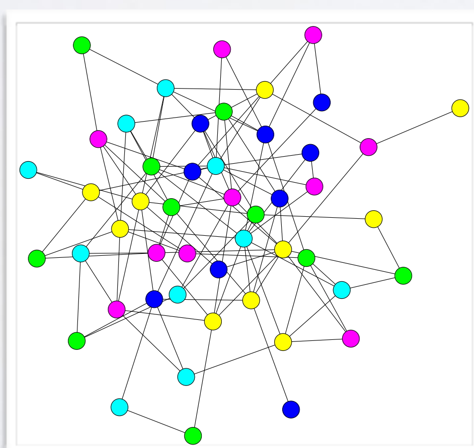




disassortative

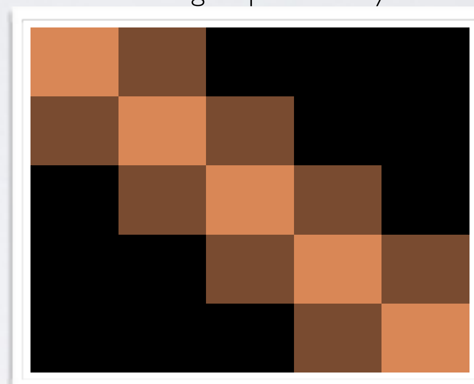
edges between groups

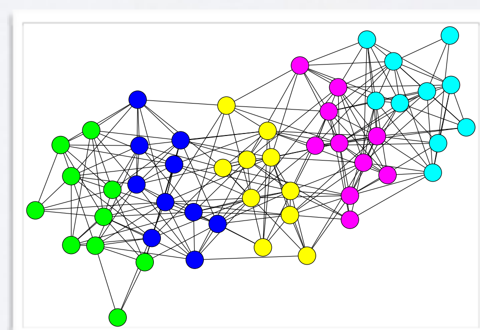




ordered

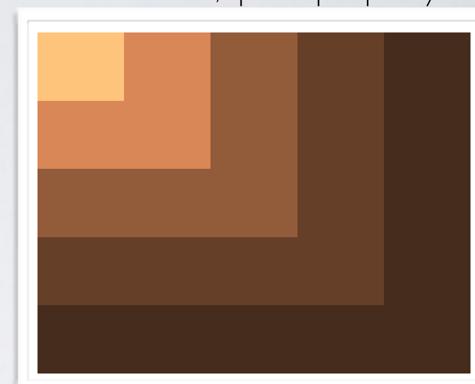
linear group hierarchy

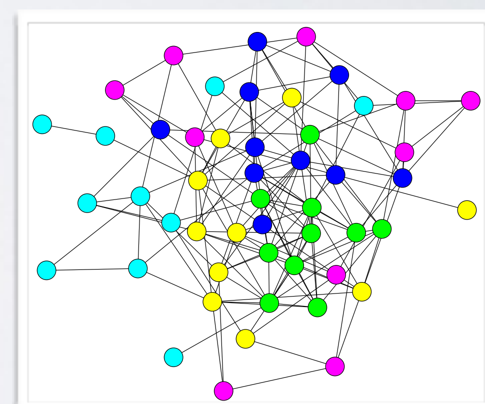




core-periphery

dense core, sparse periphery





the stochastic block model

likelihood function

the probability of G given labeling z and block matrix M

$$\Pr(G \,|\, z, M) = \underbrace{\prod_{(i,j) \in E} M_{z_i, z_j}}_{\text{edge}} \underbrace{\prod_{(i,j) \notin E} (1 - M_{z_i, z_j})}_{\text{non-edge probability}}$$

edge / non-edge probability

the stochastic block model

likelihood function

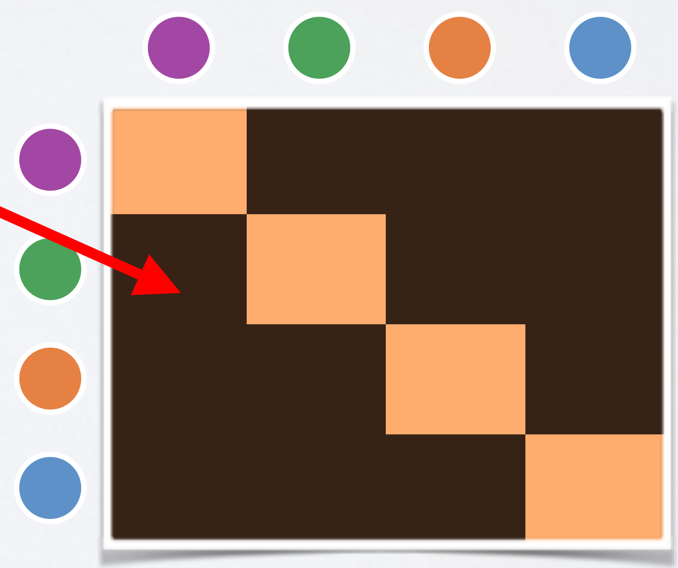
the probability of G given labeling z and block matrix M

$$\Pr(G \,|\, z, M) = \prod_{(i,j) \in E} M_{z_i,z_j} \prod_{(i,j) \notin E} (1 - M_{z_i,z_j})$$

$$= \prod_{rs} M_{r,s}^{e_{r,s}} \, (1 - M_{r,s})^{n_s n_r - e_{r,s}}$$

(Bernoulli edges)

Bernoulli random graph with parameter $M_{r,s}$



the stochastic block model

the most general SBM

$$\Pr(A \mid z, \theta) = \prod_{i,j} f\big(A_{ij} \mid \theta_{\mathcal{R}(z_i, z_j)}\big)$$

A_{ij} : value of adjacency

$\mathcal{R}$: partition of adjacencies

f : probability function

$\theta_{a,*}$: pattern for a-type adjacencies

Binomial = simple graphs
Poisson = multi-graphs
Normal = weighted graphs
etc.

θ_{11}	θ_{12}	θ_{13}	θ_{14}
θ_{21}	θ_{22}	θ_{23}	θ_{24}
θ_{31}	θ_{32}	θ_{33}	θ_{34}
θ_{41}	θ_{42}	θ_{43}	θ_{44}

the stochastic block model

degree-corrected SBM (f = Poisson)

Karrer, Newman, *Phys. Rev. E* (2011)

the stochastic block model

degree-corrected SBM (f = Poisson)

key assumption $\Pr(i \to j) = \theta_i \theta_j \omega_{z_i, z_j}$

stochastic block matrix $\omega_{r,s}$

(degree) propensity of node θ_i

likelihood:

$$\Pr(A \mid z, \theta, \omega) = \prod_{i<j} \frac{\left(\theta_i \theta_j \omega_{z_r, z_j}\right)^{A_{ij}}}{A_{ij}!} \exp\left(-\theta_i \theta_j \omega_{z_r, z_j}\right)$$

where $\underbrace{\hat{\theta}_i = \frac{k_i}{\sum_j k_j \delta_{z_i, z_j}}}_{\text{fraction of } i\text{'s group's stubs on } i}$ $\quad$ $\underbrace{\hat{\omega}_{rs} = m_{rs}}_{\text{total number of edges between } r \text{ and } s} = \sum_{ij} A_{ij} \delta_{z_i, r} \delta_{z_j, s}$

Karrer, Newman, *Phys. Rev. E* (2011)

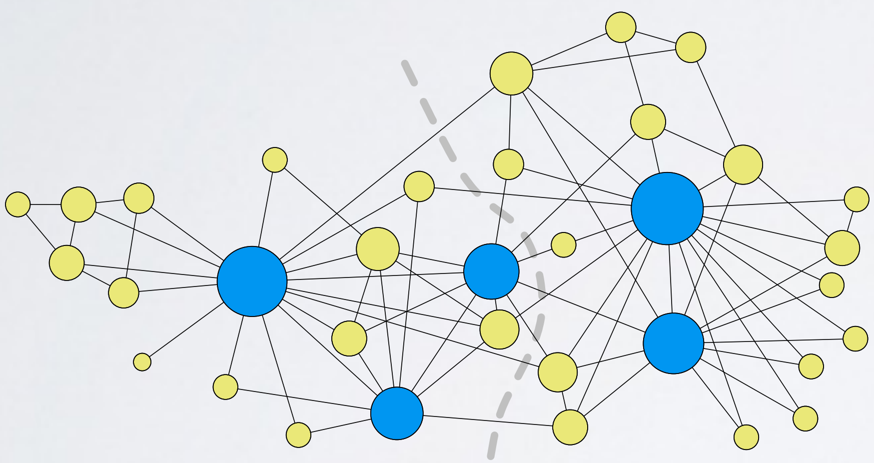
the stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club

Karrer, Newman, *Phys. Rev. E* (2011)

the stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



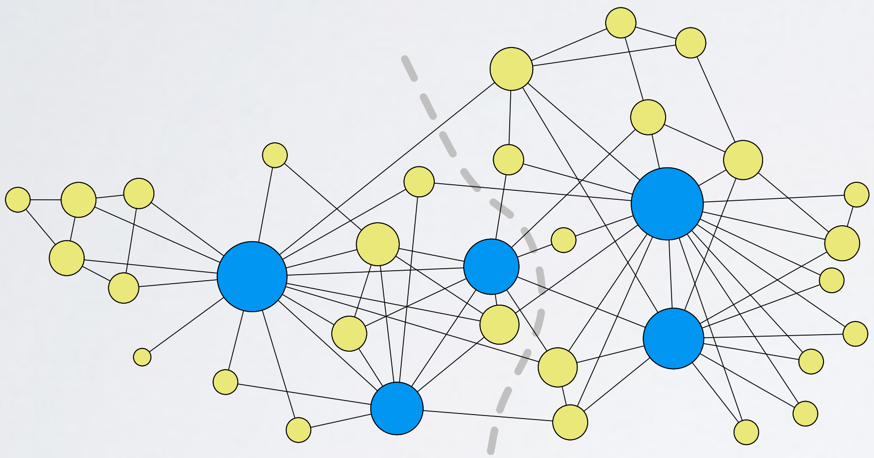
SBM
leader/follower division

DC-SBM
social group division

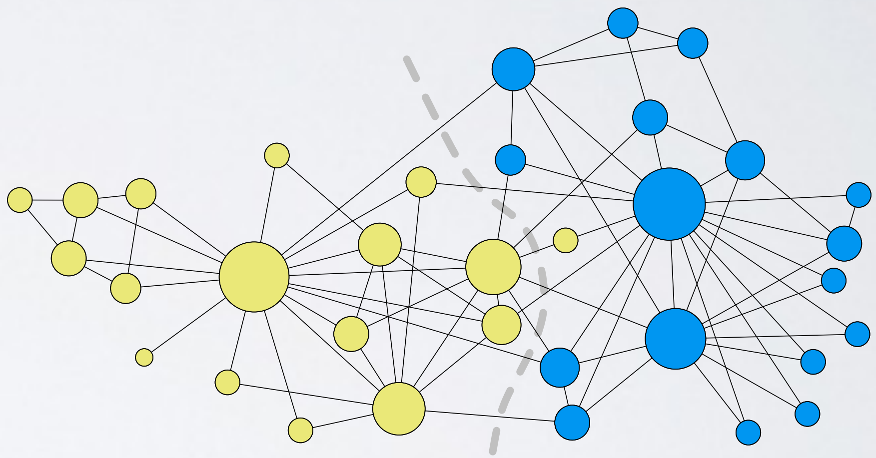
Karrer, Newman, *Phys. Rev. E* (2011)

the stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



SBM
leader/follower division

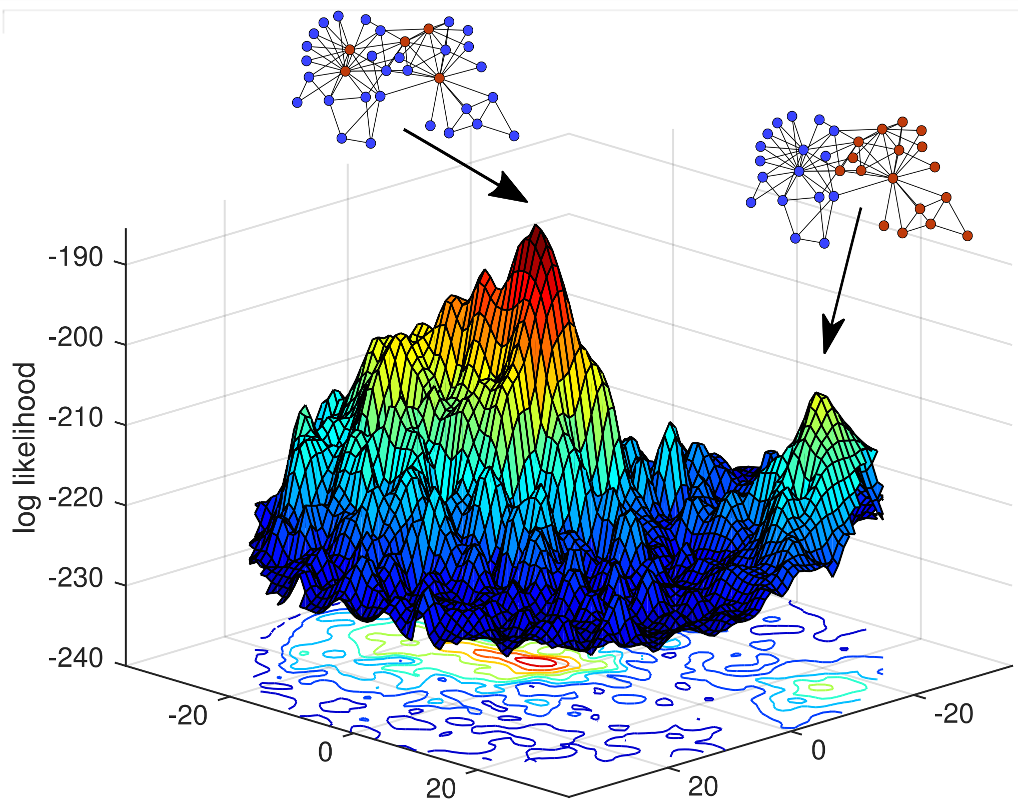


DC-SBM
social group division

Karrer, Newman, *Phys. Rev. E* (2011)

the stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



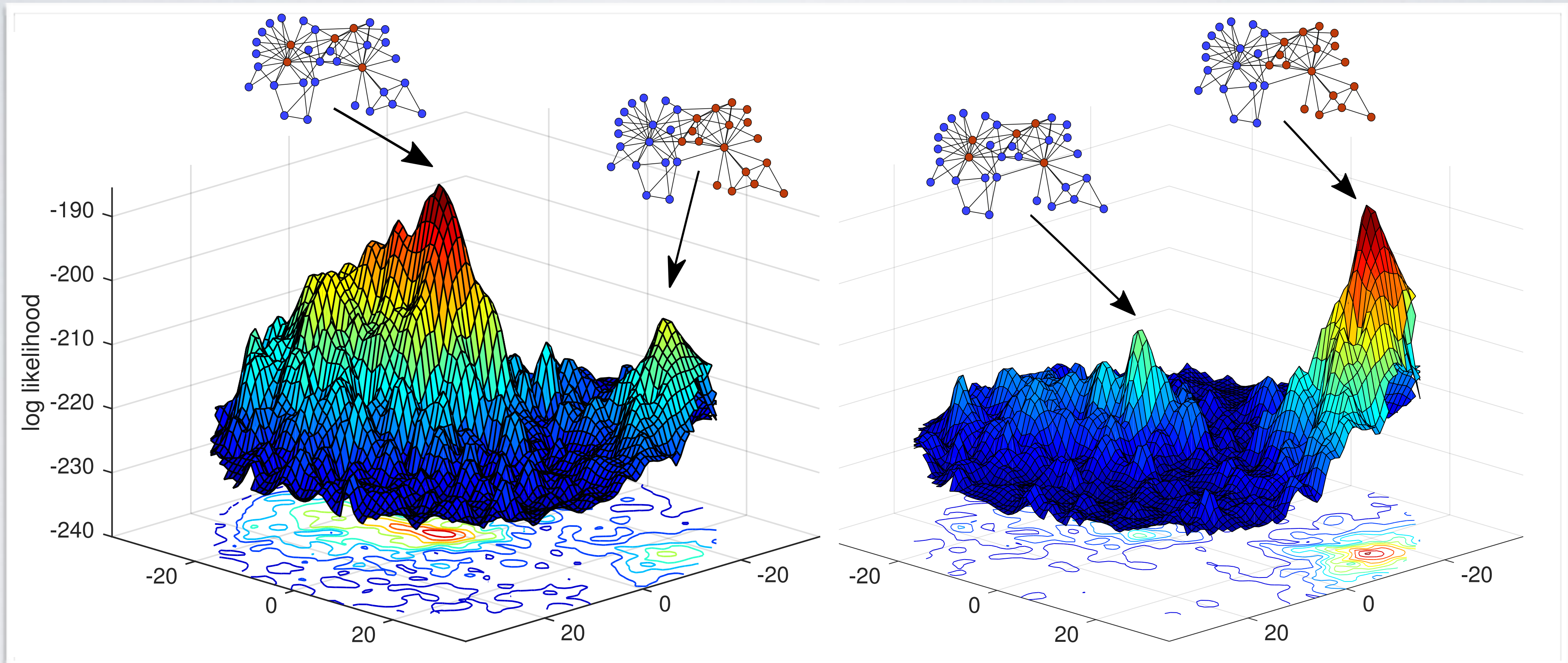
SBM
leader/follower division

DC-SBM
social group division

Peel, Larremore, Clauset, forthcoming (2016)

the stochastic block model

comparing SBM vs. DC-SBM : Zachary karate club



SBM
leader/follower division

DC-SBM
social group division

Peel, Larremore, Clauset, forthcoming (2016)

extending the SBM

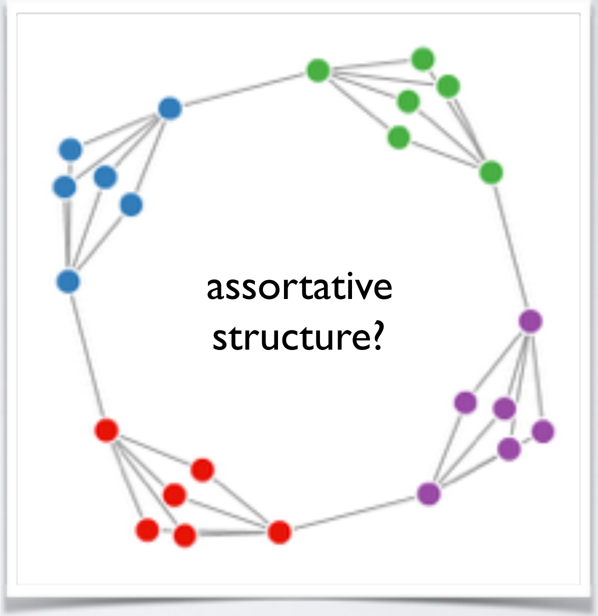
many variants! we'll cover three:

- bipartite community structure
- weighted community structure
- hierarchical community structure

bipartite networks

many networks are bipartite

- scientists and papers (co-authorship networks)
- actors and movies (co-appearance networks)
- words and documents (topic modeling)
- plants and pollinators
- genes and genomes
- etc.



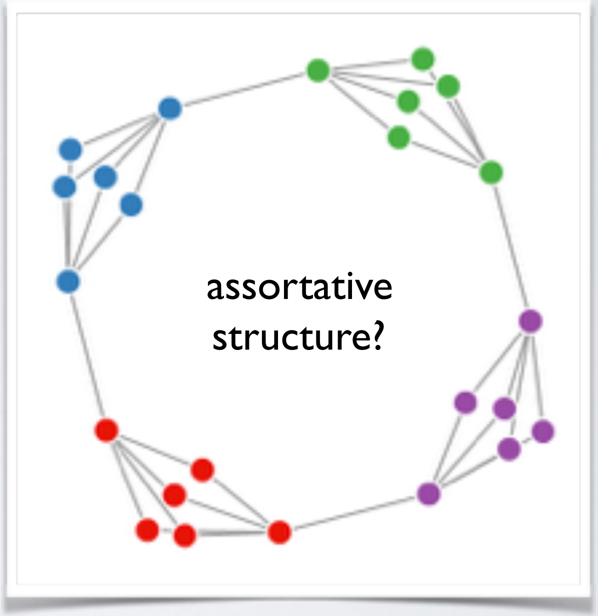
Larremore et al., *Phys. Rev. E* **90**, *012805* (2014)

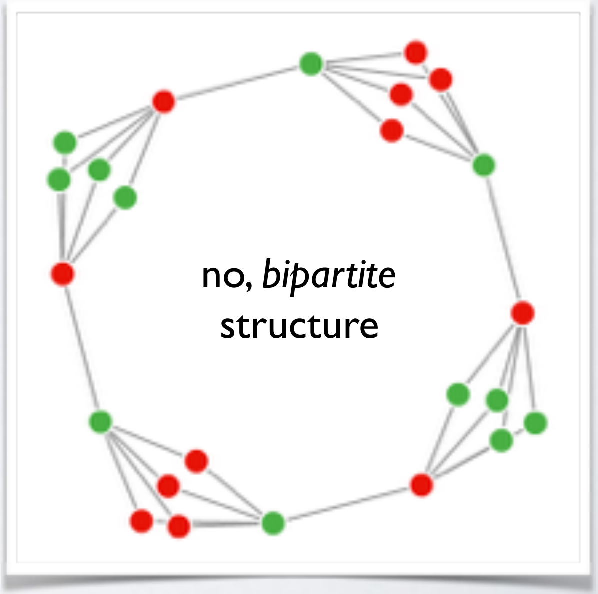
bipartite networks

many networks are bipartite

- scientists and papers (co-authorship networks)
- actors and movies (co-appearance networks)
- words and documents (topic modeling)
- plants and pollinators
- genes and genomes
- etc.

most analyses focus on one-mode projections which discard information





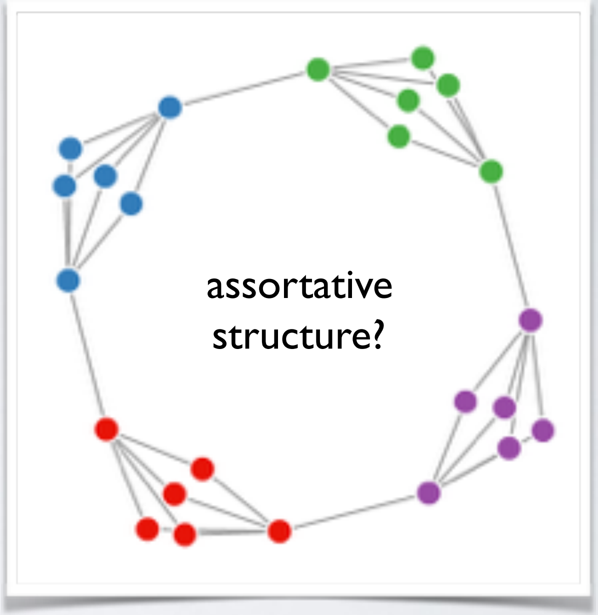
Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

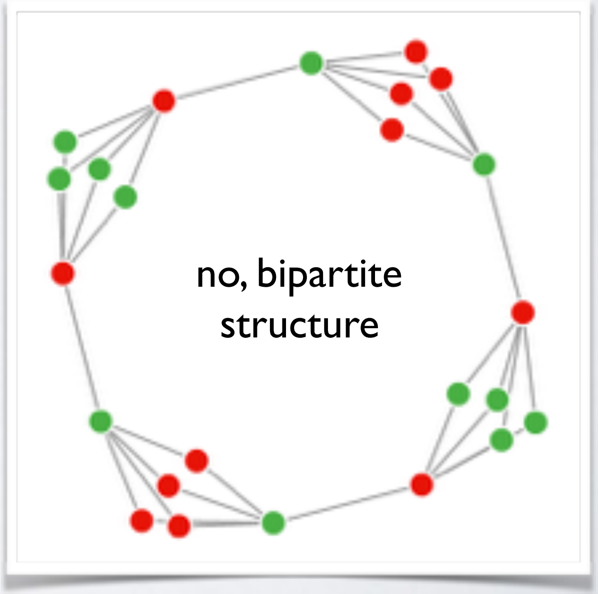
bipartite networks

bipartite stochastic block model (biSBM)

- exactly the SBM, but model knows network is bipartite
- if $\mathrm{type}(z_i) = \mathrm{type}(z_j)$

  then require $M_{z_i,z_j} = 0$
- inference proceeds as before



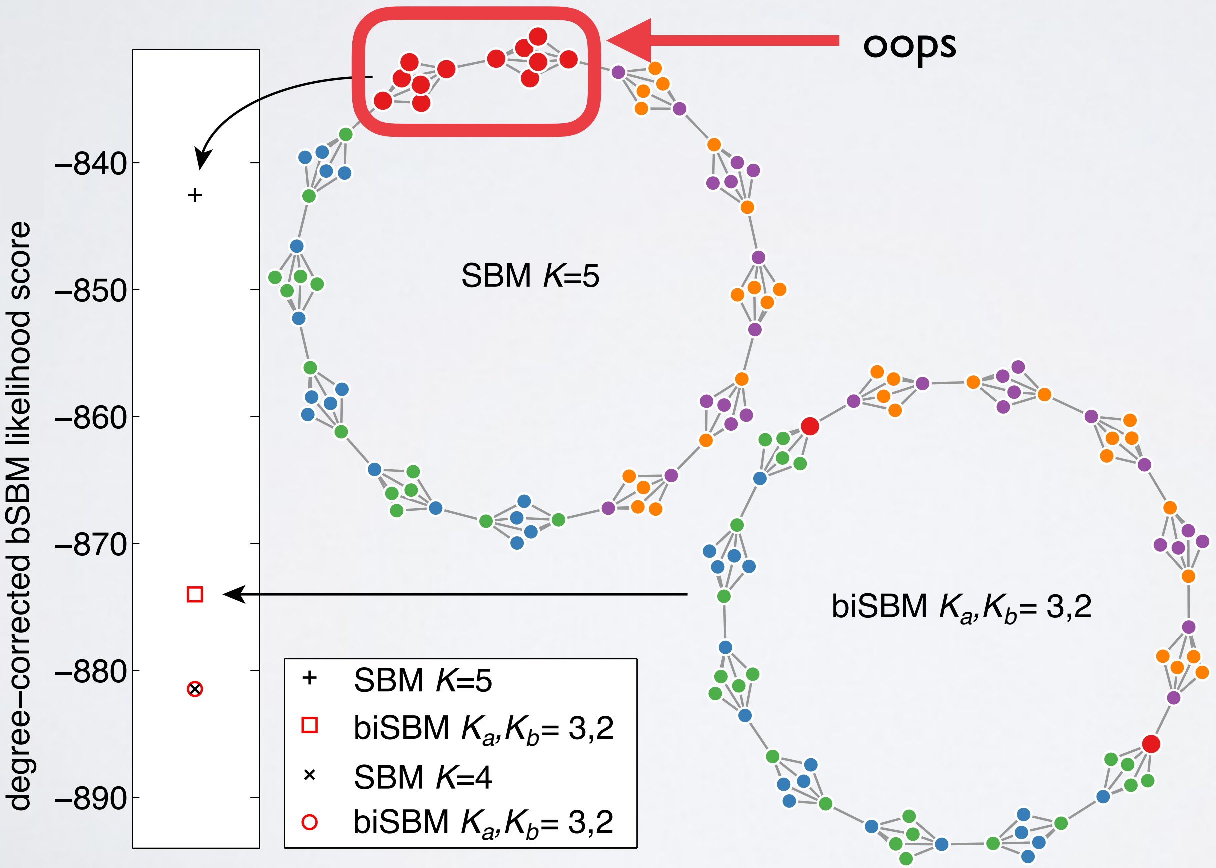


Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

SBM can learn bipartite structure on its own

but often over fits or returns mixed-type groups



Larremore et al., *Phys. Rev. E* **90**, *012805* (2014)

bipartite networks

bipartite stochastic block model (biSBM)

- how do we know it works well?

Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

bipartite stochastic block model (biSBM)

- how do we know it works well?
- synthetic networks with *planted partitions*

$$\omega^{\text{planted}} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \alpha & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \beta & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & \gamma & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & \delta \\ \alpha & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & \beta & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \gamma & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \delta & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

easy case:
4x4 easy-to-see communities

$$\omega^{\text{planted}} = \begin{pmatrix} \cdot & \cdot & \cdot & \epsilon & 0 \\ \cdot & \cdot & \cdot & 0 & \epsilon \\ \cdot & \cdot & \cdot & \gamma & \gamma \\ \epsilon & 0 & \gamma & \cdot & \cdot \\ 0 & \epsilon & \gamma & \cdot & \cdot \end{pmatrix}$$

hard case:
3x2 hard-to-see communities

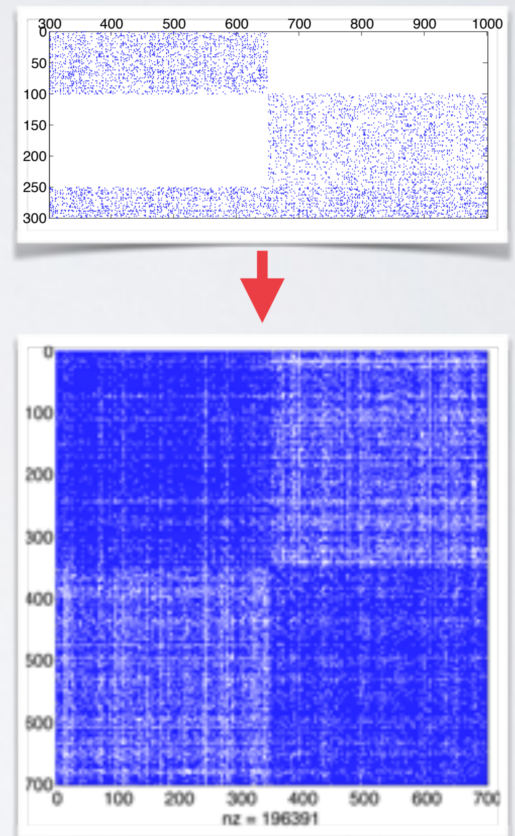
Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

bipartite stochastic block model (biSBM)

- how do we know it works well?
- synthetic networks with *planted partitions*

$$\omega^{\text{planted}} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \alpha & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \beta & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & \gamma & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & 0 & \delta \\ \alpha & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & \beta & 0 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \gamma & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \delta & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$



easy case:
4x4 easy-to-see communities

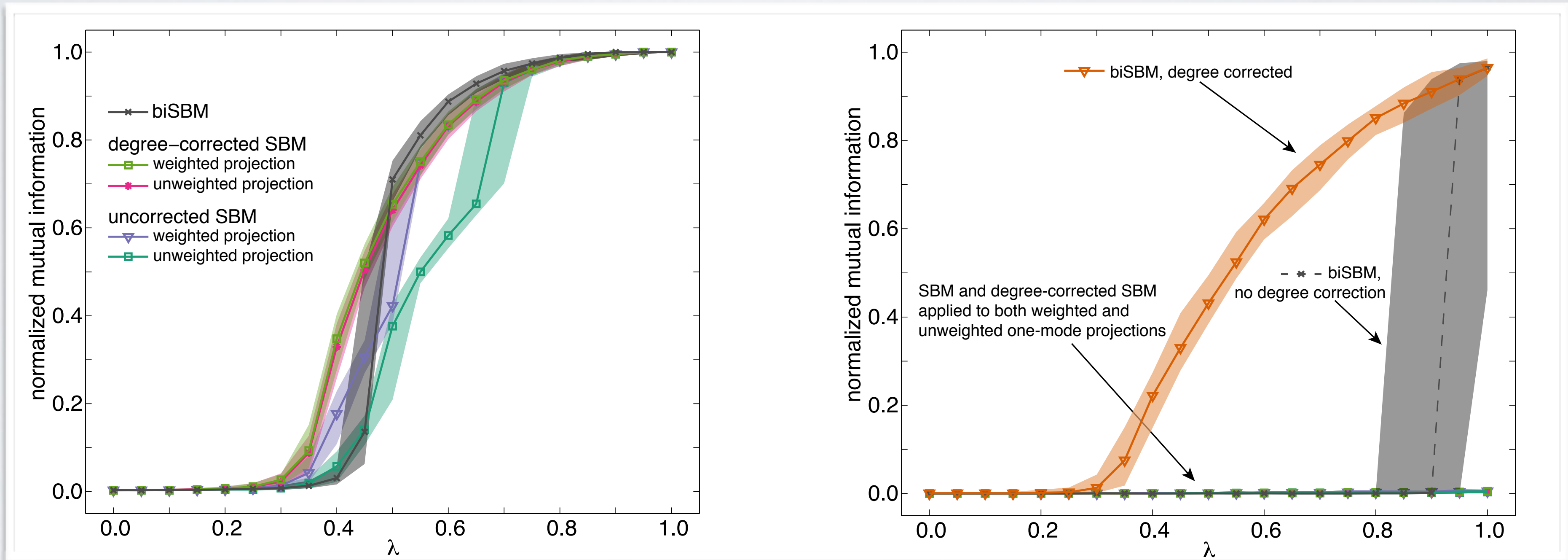
hard case:
3x2 hard-to-see communities

Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

bipartite stochastic block model (biSBM)

- how do we know it works well?
- synthetic networks with *planted partitions*



easy case:
4x4 easy-to-see communities

hard case:
3x2 hard-to-see communities

Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

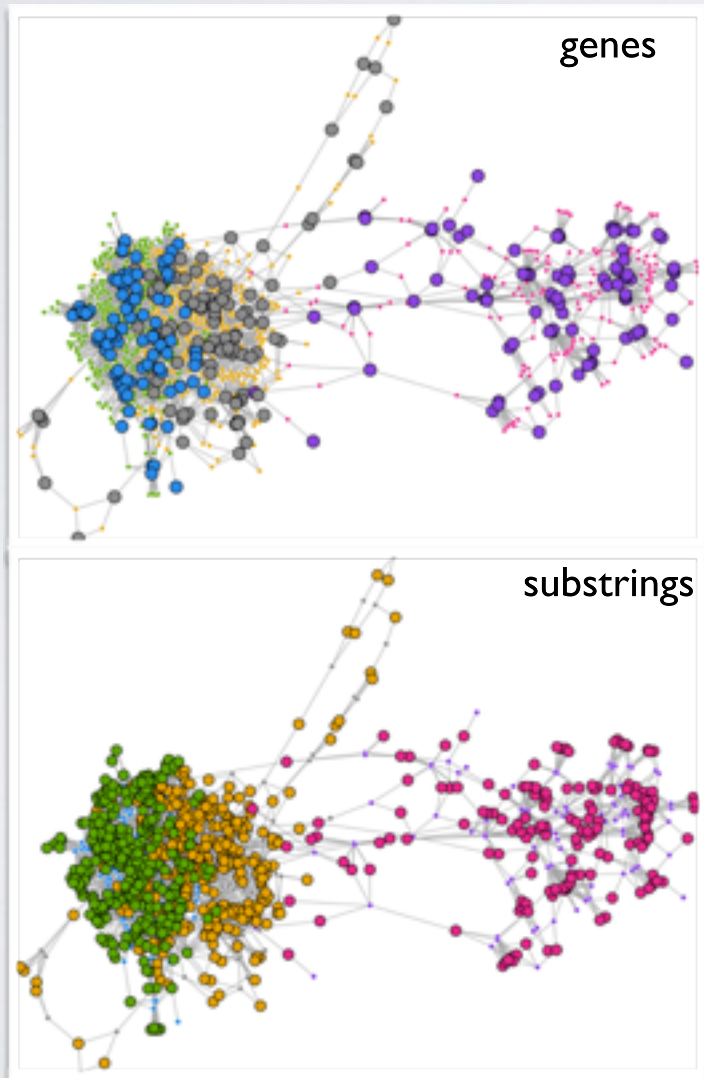
bipartite stochastic block model (biSBM)

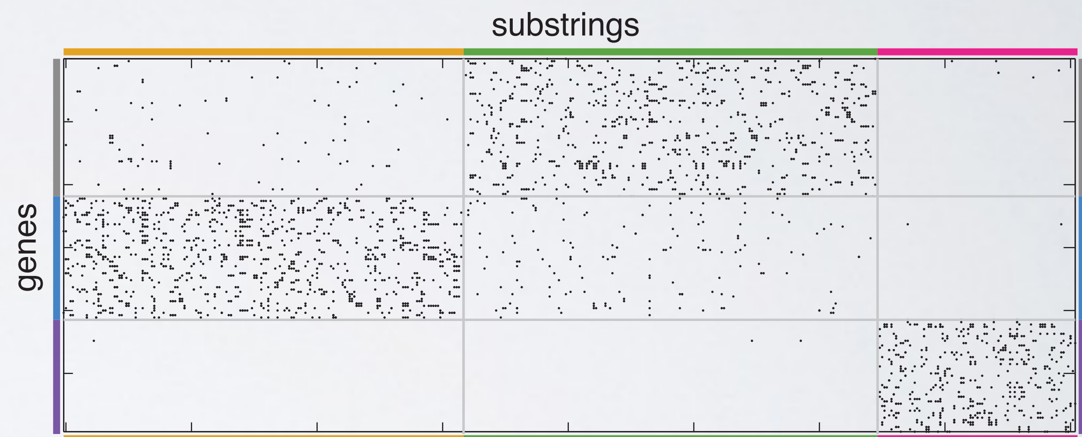
- always find pure-type communities
- more accurate than modeling one-mode projections (even weighted projections)
- finds communities in *both* modes

Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

example 1: malaria gene network

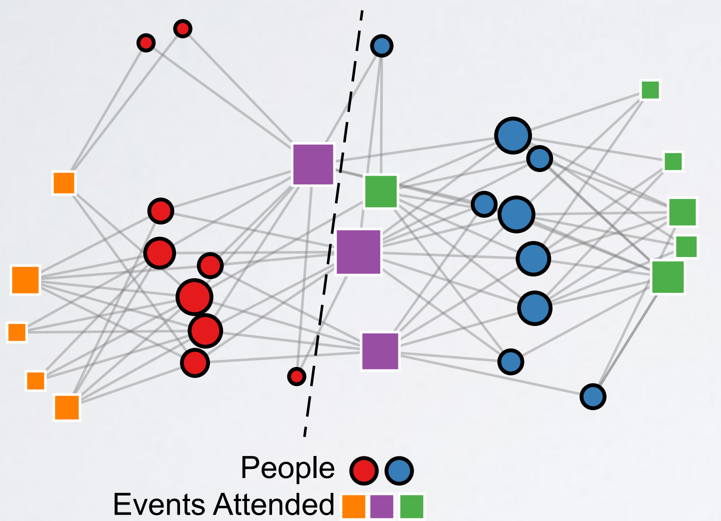


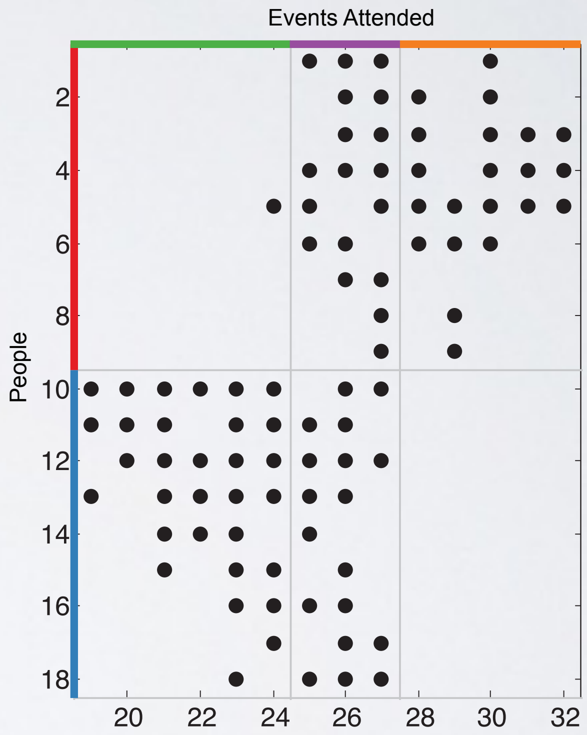


Larremore et al., *Phys. Rev. E* **90**, *012805* (2014)

bipartite networks

example 2: Southern Women network

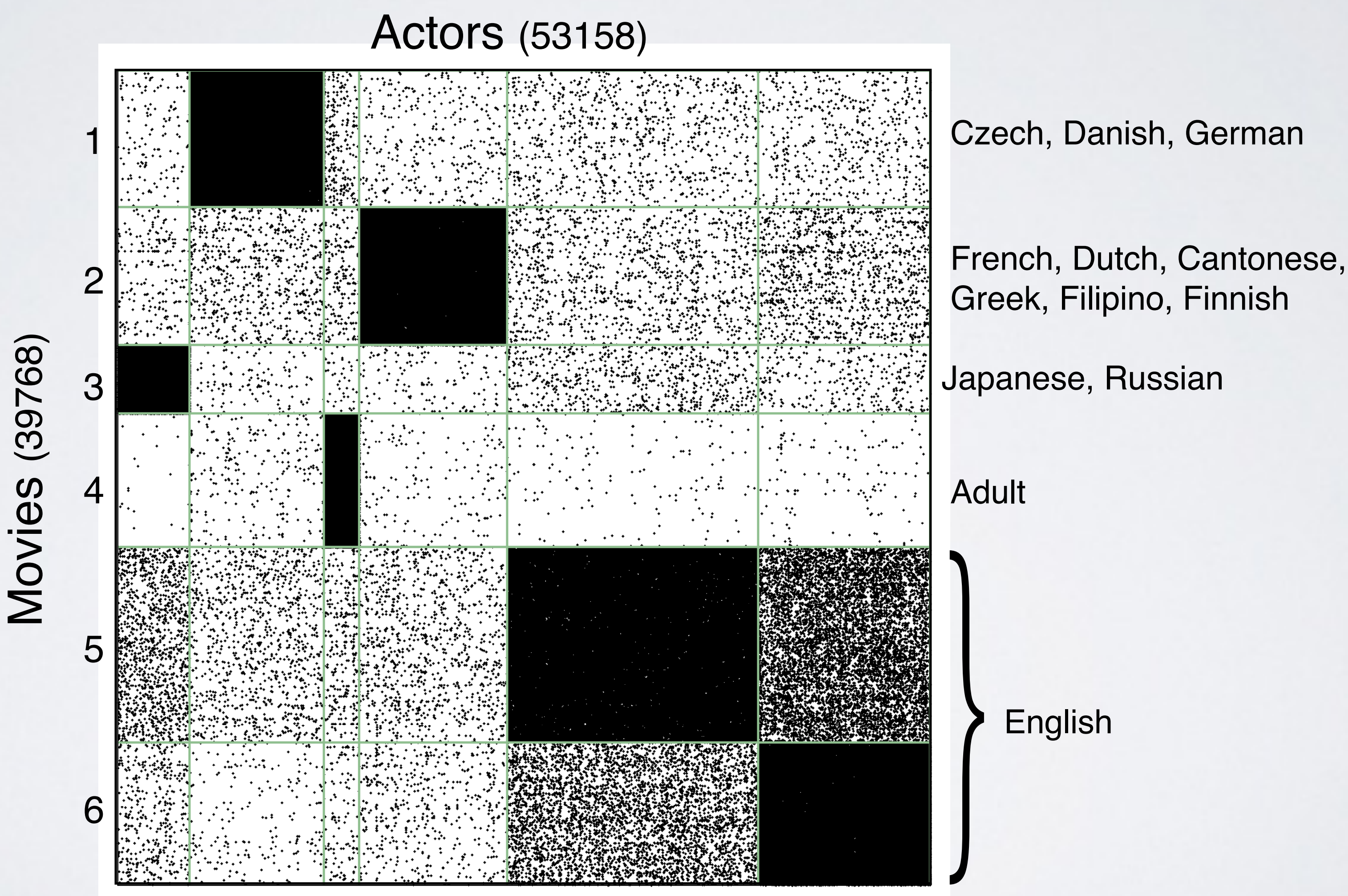




Larremore et al., *Phys. Rev. E* ***90***, *012805* (2014)

bipartite networks

example 3: IMDb



Larremore et al., *Phys. Rev. E* ***90****, 012805* (2014)

bipartite networks

other approaches

PRL **110,** 148701 (2013) PHYSICAL REVIEW LETTERS week ending 5 APRIL 2013

Parsimonious Module Inference in Large Networks

Tiago P. Peixoto*

Institut für Theoretische Physik, Universität Bremen, Hochschulring 18, D-28359 Bremen, Germany

(Received 19 December 2012; published 5 April 2013; publisher error corrected 5 April 2013)

minimum description length (MDL) principle

learns that a network is bipartite

OPEN ACCESS Freely available online PLOS | ONE

Predicting Human Preferences Using the Block Structure of Complex Social Networks

Roger Guimerà[1,2]*, Alejandro Llorente[3], Esteban Moro[4,5,3], Marta Sales-Pardo[2]

marginalize over bipartite SBM parameterizations

Peixoto, *Phys. Rev. Lett.* ***110***, *148701* (2013)
Guimera et al., *PLOS ONE* **7**(9), e44620 (2012)

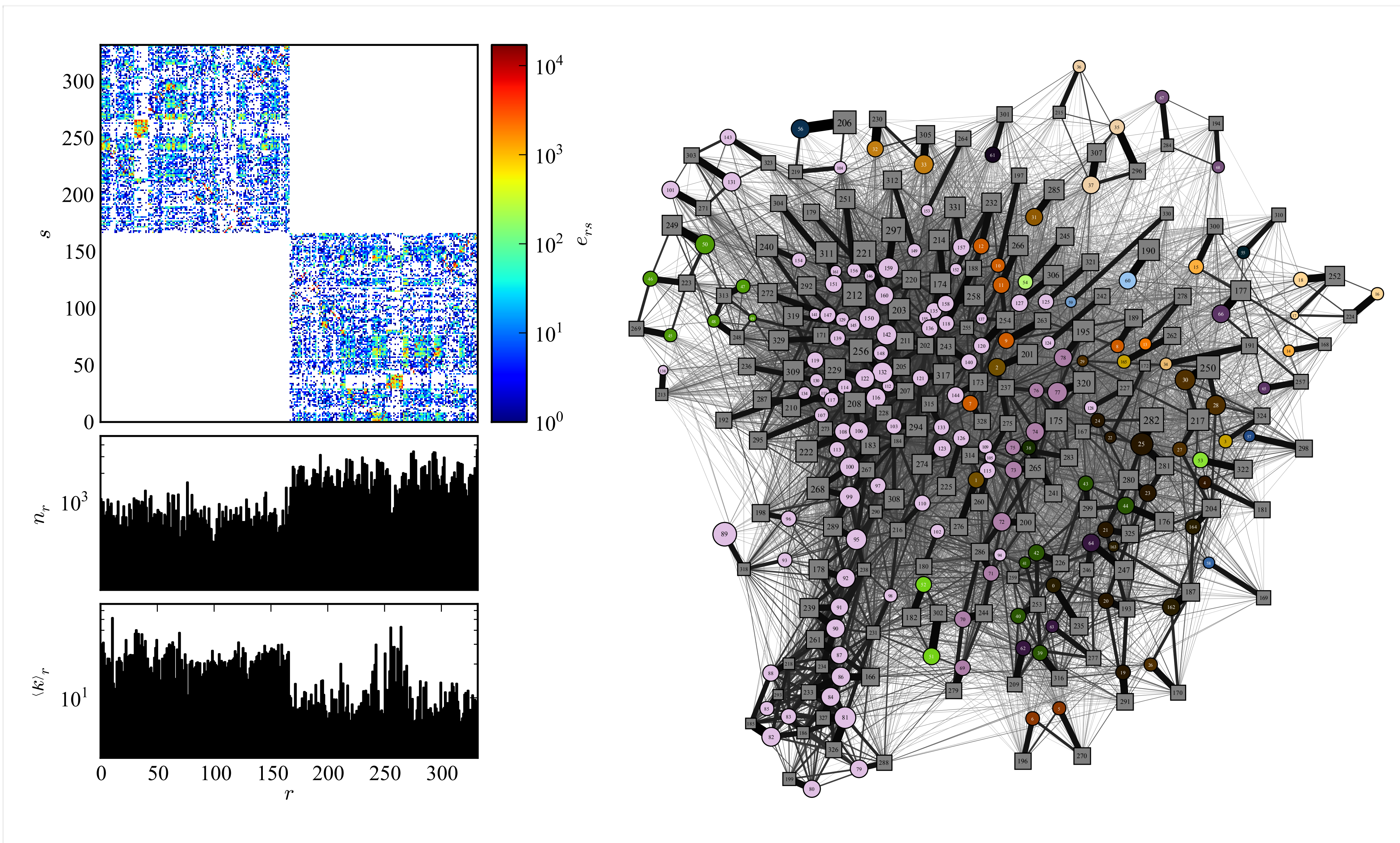
bipartite networks

PRL 110, 148701 (2013) PHYSICAL REVIEW LETTERS week ending 5 APRIL 2013

Parsimonious Module Inference in Large Networks

Tiago P. Peixoto*
Institut für Theoretische Physik, Universität Bremen, Hochschulring 18, D-28359 Bremen, Germany

(Received 19 December 2012; published 5 April 2013; publisher error corrected 5 April 2013)



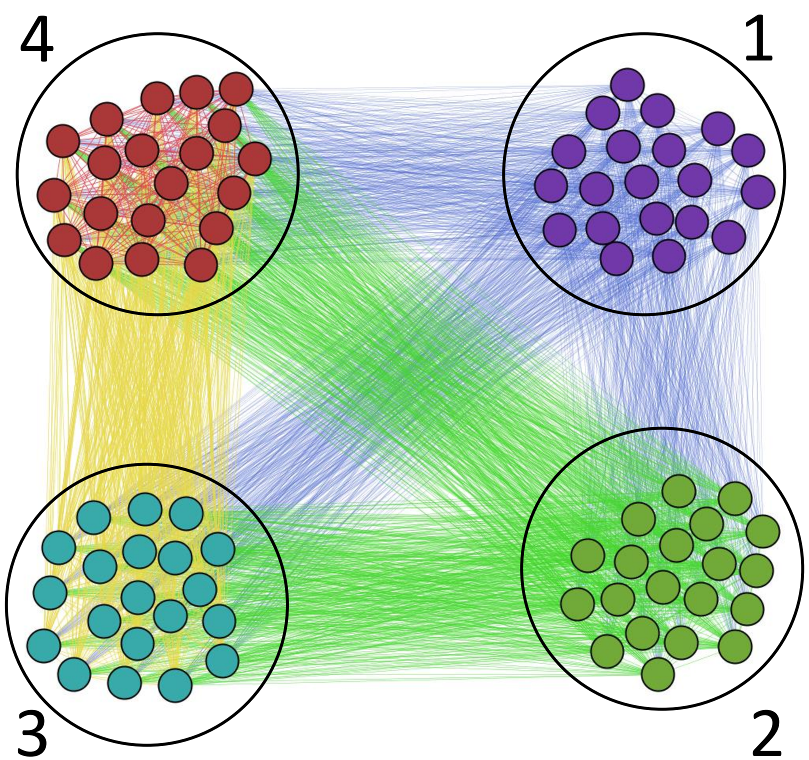
Peixoto, *Phys. Rev. Lett.* ***110***, *148701* (2013)

weighted networks

most interactions are weighted

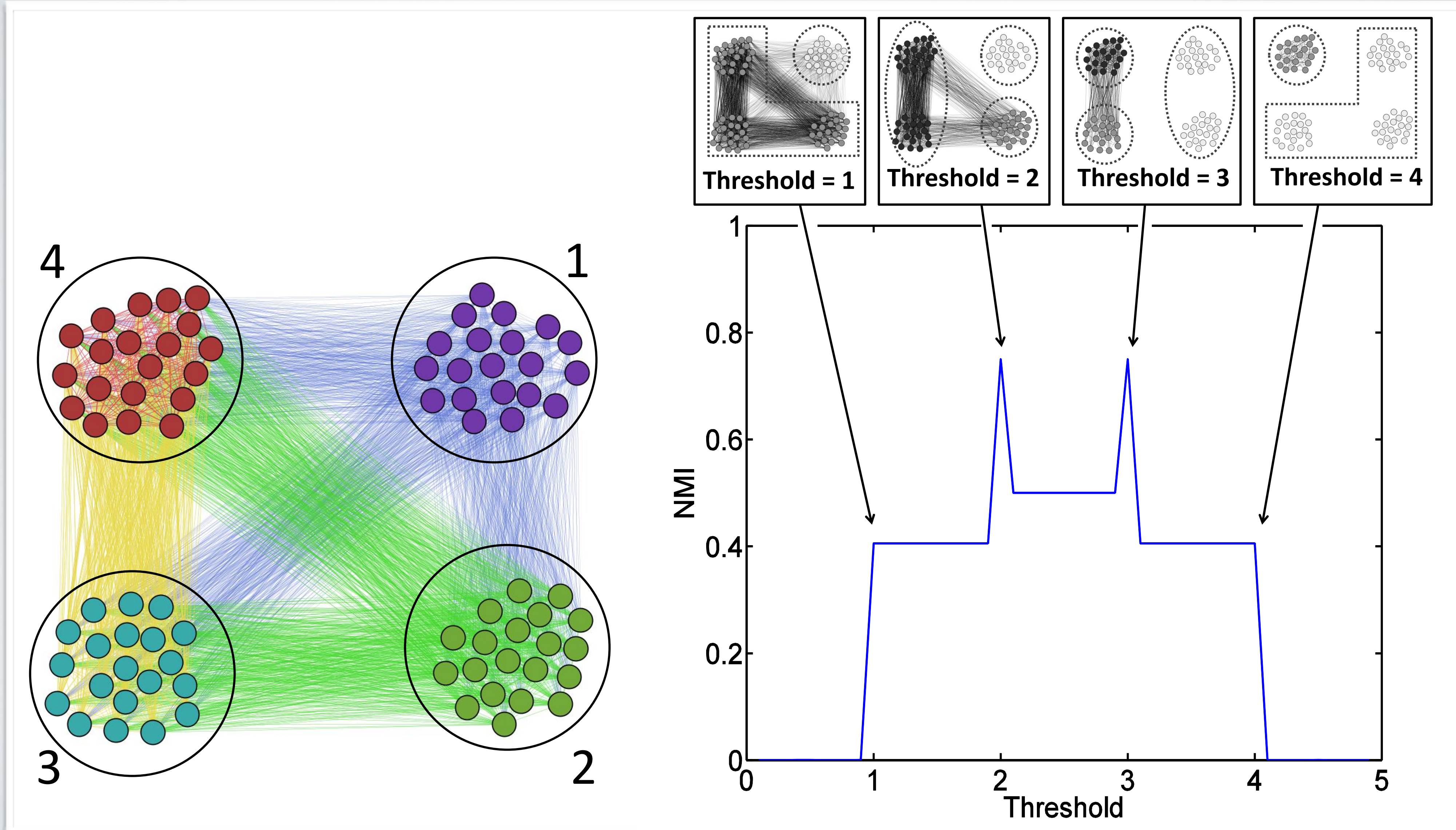
- frequency of interaction
- strength of interaction
- outcome of interaction
- etc.
- **but!** thresholding discards information and can obscure underlying structure

weighted networks



Aicher et al., *J. Complex Networks* ***3***, *221-228* (2015).

weighted networks



Aicher et al., *J. Complex Networks* ***3***, *221-228* (2015).

weighted networks

Valued Ties Tell Fewer Lies: Why Not To Dichotomize Network Edges With Thresholds*

Andrew C. Thomas† Joseph K. Blitzstein‡

- how will the results depend on the threshold?
- what impact does noise have, under threshold?

Thomas & Blitzstein, *arxiv:1101.0788* (2011).

recall…

the most general SBM

$$\Pr(A \mid z, \theta) = \prod_{i,j} f\left(A_{ij} \mid \theta_{\mathcal{R}(z_i, z_j)}\right)$$

A_{ij} : value of adjacency

$\mathcal{R}$: partition of adjacencies

f : probability function

$\theta_{a,*}$: pattern for a-type adjacencies

Binomial = simple graphs
Poisson = multi-graphs
Normal = weighted graphs
etc.

θ_{11}	θ_{12}	θ_{13}	θ_{14}
θ_{21}	θ_{22}	θ_{23}	θ_{24}
θ_{31}	θ_{32}	θ_{33}	θ_{34}
θ_{41}	θ_{42}	θ_{43}	θ_{44}

weighted networks

weighted stochastic block model (WSBM)

- model edge *existence* and edge *weight* separately
- *edge existence*: SBM
- *edge weights*: exponential family distribution

log-likelihood:

edge-existence
[binomial distribution]
M_{z_i,z_j}

$$\begin{aligned}\ln \Pr(G \,|\, M, z, \theta, f) &= \alpha \ln \Pr(G \,|\, M, z) \\ &\quad + (1-\alpha) \ln \Pr(G \,|\, \theta, z, f)\end{aligned}$$

edge-weights
[exponential-family distribution]
θ_{z_i,z_j}

Poisson, Normal, Gamma,
Exponential, Pareto, etc.

Aicher et al., *J. Complex Networks* ***3****, 221-228* (2015).

weighted networks

weighted stochastic block model (WSBM)

- model edge *existence* and edge *weight* separately
- *edge existence*: SBM
- *edge weights*: exponential family distribution

log-likelihood:

$$\ln \Pr(G \,|\, M, z, \theta, f) = \alpha \ln \Pr(G \,|\, M, z) + (1-\alpha) \ln \Pr(G \,|\, \theta, z, f)$$

mixing parameter

$\alpha = 1$ only model edge existence (ignore weights)

$\alpha = 0$ only model edge weights (ignore non-edges)

Aicher et al., *J. Complex Networks* ***3***, *221-228* (2015).

weighted networks

American football, NFL 2009 season

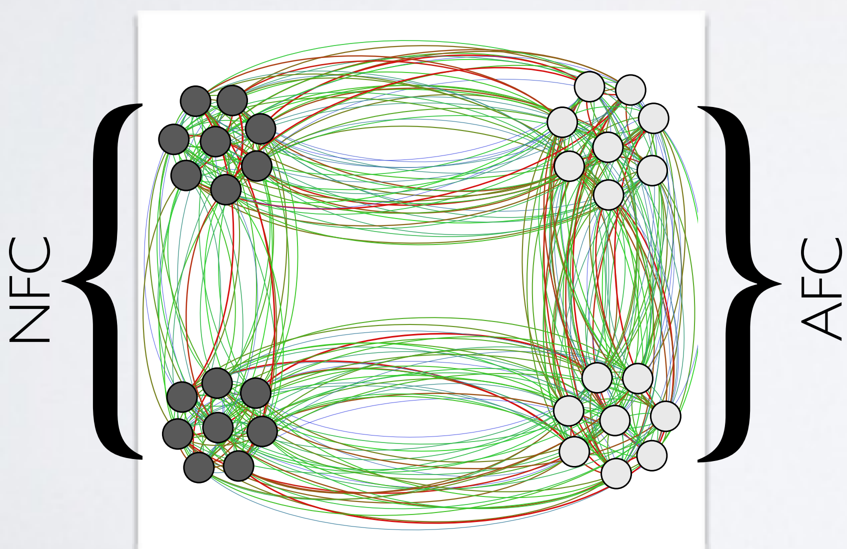
- 32 teams, 2 "divisions", 4 "subdivisions"
- *edge existence*: who plays whom
- *edge weight*: mean score difference

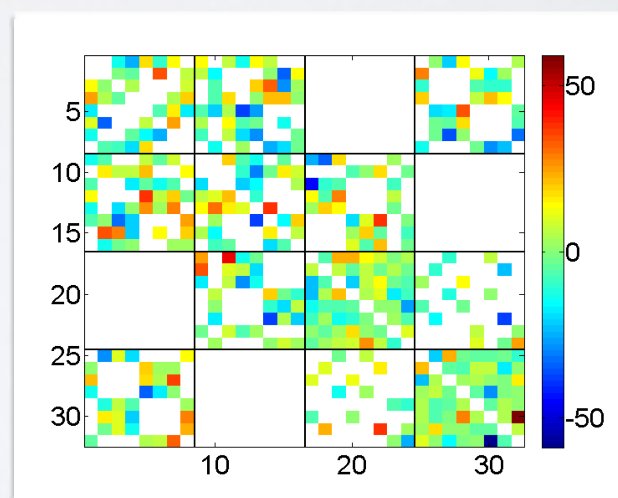
Aicher et al., *J. Complex Networks* ***3***, *221-228* (2015).

weighted networks

American football, NFL 2009 season

- 32 teams, 2 "divisions", 4 "subdivisions"
- SBM ($\alpha = 1$) recovers subdivisions perfectly





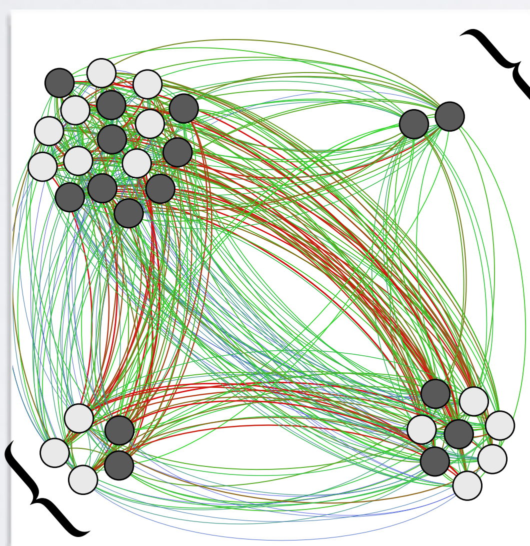
block matrix M, z

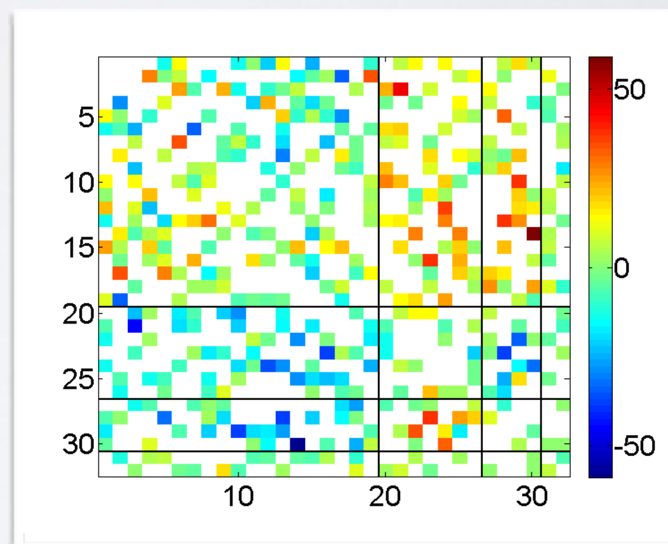
Aicher et al., *J. Complex Networks* ***3***, *221-228* (2015).

weighted networks

American football, NFL 2009 season

- 32 teams, 2 "divisions", 4 "subdivisions"
- WSBM ($\alpha = 0$) recovers team skill hierarchy





block matrix M, z

Aicher et al., *J. Complex Networks* **3**, 221-228 (2015).

weighted networks

adding weights to the SBM

- what does $A_{ij} = 0$ mean?

  no edge *or* weight=0 edge *or* non-observed edge?

- how will we model the distribution of edge weights?

- edge *existences* and edge *weights* may contain *different* large-scale structure (conference structure vs. skill hierarchy)

weighted networks

other approaches

weighted networks

other approaches

PHYSICAL REVIEW E **92**, 042807 (2015)

Inferring the mesoscale structure of layered, edge-valued, and time-varying networks

Tiago P. Peixoto*
Institut für Theoretische Physik, Universität Bremen, Hochschulring 18, D-28359 Bremen, Germany

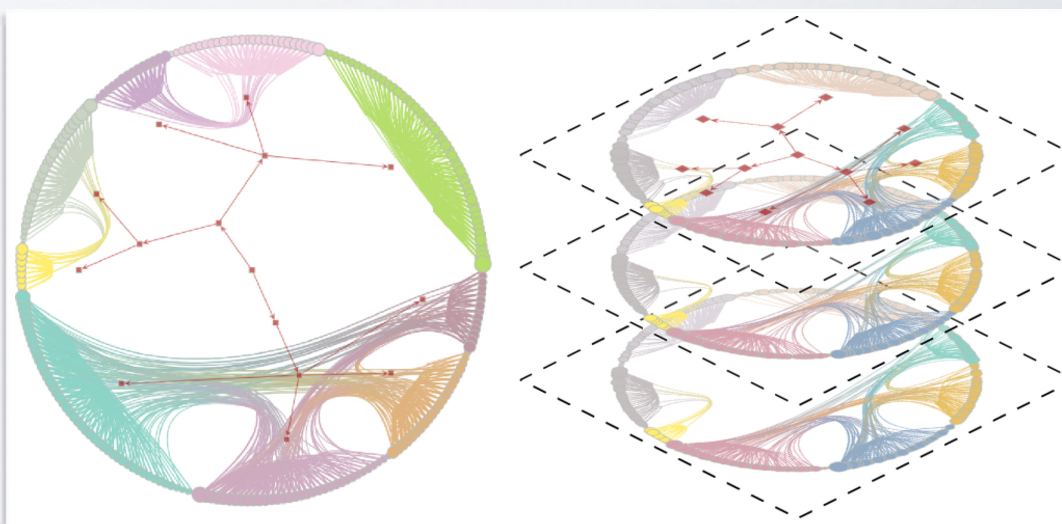
bin the edge weights into C ranges

C bins = C layers of a multi-layer SBM

common node labels

each layer has its own block matrix

infer C, M_c, z



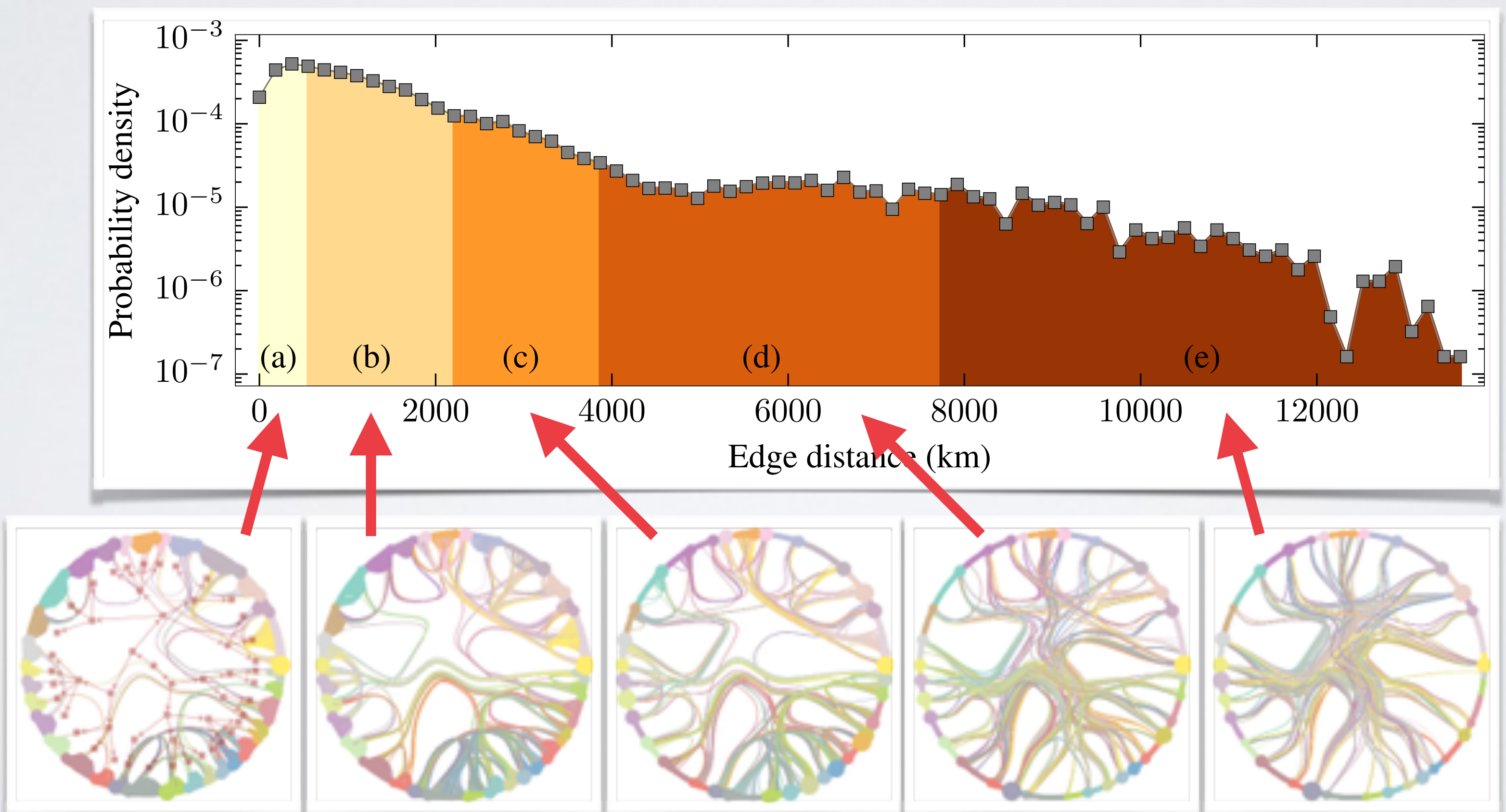
Peixoto, *Phys. Rev. E* **92**, *042807* (2015)

weighted networks

other approaches (discretized weights = SBM layers)

OpenFlights airport network

edge weights = distance (km)



Peixoto, *Phys. Rev. E* **92**, *042807* (2015)

weighted networks

other approaches (discretized weights = SBM layers)

OpenFlights airport network

Peixoto, *Phys. Rev. E* **92**, *042807* (2015)

fin