A Quick Survey of Statistics and Networks

Where We've Invested and Where We Might Next Invest

Eric D. Kolaczyk

McGill University

eric.kolaczyk@mcgill.ca @EKolaczyk

Introductory Overview Lecture, JSM 2022

Introduction

"Once Upon a Time . . ."

- Network analysis was a relatively small 'field' of study until $\sim$ 20 years ago
- Epidemic-like spread of interest in networks since mid-90s, across the sciences and humanities
- Arguably two key factors include
  - Increasingly systems-level perspective; and
  - Flood of high-throughput data (and the accompanying data science tools).

What Do We Mean by 'Network'?

Definition (OED): *A collection of inter-connected things.*

Formally, we typically use a graph $G = (V, E, W)$, where

- V is a set of $n_v = |V|$ vertices;
- E is a set of $n_e = |E|$ edges between vertex pairs; and
- $W = [W_{ij}]$ is an $n_v \times n_v$ matrix of (non-neg) weights.

A binary adjacency matrix $A = [A_{ij}]$ captures presence/absence of edges $\{i, j\} \in E$.

What Do We Mean by 'Network'?

Definition (OED): *A collection of inter-connected things.*

Formally, we typically use a graph $G = (V, E, W)$, where

- V is a set of $n_v = |V|$ vertices;
- E is a set of $n_e = |E|$ edges between vertex pairs; and
- $W = [W_{ij}]$ is an $n_v \times n_v$ matrix of (non-neg) weights.

A binary adjacency matrix $A = [A_{ij}]$ captures presence/absence of edges $\{i, j\} \in E$.

Caveat emptor: The term 'network' is often used in the literature to refer to the system, a graph, and even visualization(s) of the graph . . . sometimes in the same paper!

Our Focus

The statistical analysis of *network data*

i.e., analysis of measurements either of or from a system conceptualized as a network.

Core challenges include:

- relational aspect to the data;
- complex statistical dependencies (often the focus!);
- high-dimensional and often massive in quantity.

Statistics & Networks – Looking Back

Statistics started out as a comparatively minor player in network science 20 years ago.

Yet networks – as a form of complex data – are fundamentally data objects and hence the full taxonomy of statistical inquiry and analysis is relevant (e.g., sampling/design, inference, testing, prediction, modeling, visualization, etc.).

In the ensuing years, statisticians have since made substantial contributions to network science, particularly – as often the case – in a handful of core areas.

Goal for Today

Provide a (highly selective!) introduction and overview to several core topics at the interface of statistics and networks, with an eye towards where we've invested and where we might invest.

Chosen as a function of (i) depth / completeness of solution(s); and (ii) breadth of impact. Established, in the case of where we've invested, and anticipated, in the case of where we might invest.

Goal for Today

Provide a (highly selective!) introduction and overview to several core topics at the interface of statistics and networks, with an eye towards where we've invested and where we might invest.

Chosen as a function of (i) depth / completeness of solution(s); and (ii) breadth of impact. Established, in the case of where we've invested, and anticipated, in the case of where we might invest.

- **Where we've invested:** Network topology inference; community detection.
- **Where we might invest:** Multiple networks; noisy networks.

Apologies for the many topics / contributions we'll inevitably skip!

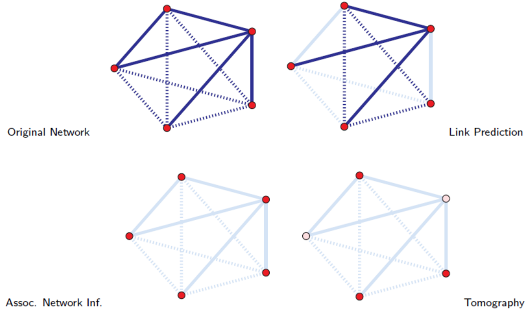
Modus Operandi

For each topic area, I will

1. Describe a canonical problem(s) through pictures;
2. Shine a "Spotlight on . . ." a key solution(s)  ;
3. Give a "Shout Out to . . ." other solutions  .

Where We’ve Invested

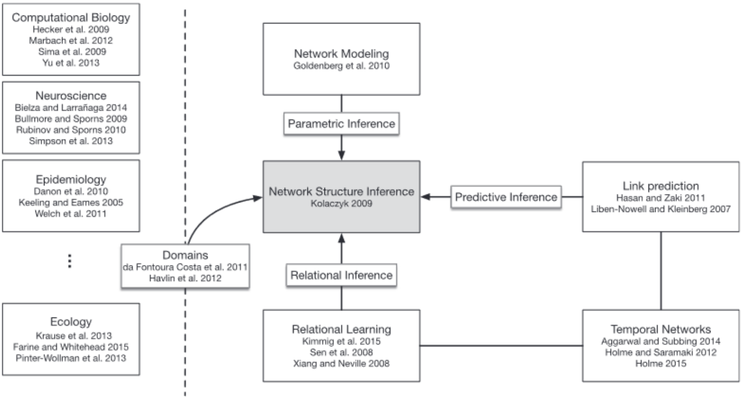
Problem in Pictures: Network Topology Inference



Question: Given available information, how might we infer unknown presence/absence of edges between vertex pairs?

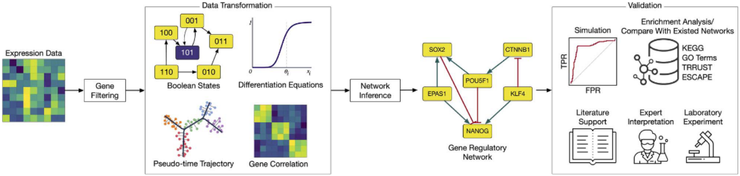
Kolaczyk 2009, Ch 7

A Truly Substantial Literature



Brugere, Gallagher, and Berger-Wolf 2018

Rich Development Even Within Many (Sub)Domains



Nguyen et al. 2021, "A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data".

Spotlight on . . . Gaussian Graphical Models

Let $X = (X_v)_{v \in V}$ represent a vector of continuous measurements indexed by vertices in an undirected graph $G = (V, E)$.

Suppose $X \sim N(0, \Sigma)$, and define $K = [\kappa_{ij}] = \Sigma^{-1}$.

A *Gaussian graphical model (GGM)* for X w.r.t G specifies that $\kappa_{ij} \neq 0$ when $\{i, j\} \in E$.

G is called a *conditional independence graph* or a *concentration graph*.

See Drton and Maathuis 2017 for a comprehensive survey on structure learning in graphical models generally.

GGM Inference

Let $S = [S_{ij}]$ be a sample covariance based on n observations.

The *graphical lasso (glasso)* estimator is an ℓ_1-penalized MLE:

$$\hat{K}^{gl} = \arg\min_{K} \{-\log\det(K) + tr(SK) + \lambda||K||_1\}$$

The estimate $\hat{G}^{gl} = (V, \hat{E}^{gl})$ follows through the rule

$$\{i,j\} \in \hat{E}^{gl} \quad \text{iff} \quad \hat{\kappa}_{ij}^{gl} \neq 0 \ .$$

Yuan and Lin 2007; Banerjee, El Ghaoui, and d'Aspremont 2008

Properties & Implementation

Theorem (Yuan and Lin 2007)

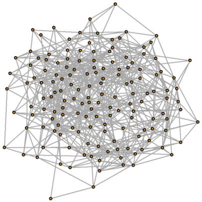
Under appropriate conditions, Glasso selects the correct graph G with probability tending to 1 (and, at the same time, yields a root-n consistent estimate of K).

Implementations utilizing variations on coordinate descent and theory-driven initialization scale to millions of vertices.

Penalty parameter λ can be chosen adaptively in various ways (e.g., BIC, stability selection, etc.).

See, e.g., Friedman, Hastie, and Tibshirani 2008; Witten, Friedman, and Simon 2011; Hsieh et al. 2013; Gao et al. 2012; Liu, Roeder, and Wasserman 2010.

Illustration: Gene coexpression in Ecoli



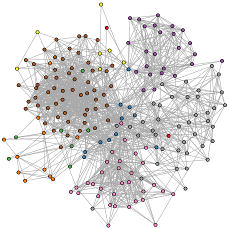
```
> library(igraph)
> library(sand)
> library(huge)
> huge.out <- huge(Ecoli.expr)
> huge.opt <- huge.select(huge.out, criterion="stars")
> g.huge <- graph_from_adjacency_matrix(huge.opt$refit, "undirected")
> summary(g.huge)
IGRAPH a04cd27 U--- 153 623 --
> plot(g.huge,vertex.size=3,vertex.label=NA)
```

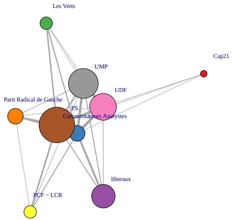
Kolaczyk and Csárdi 2020, Ch 7.3

Shout Out to ...

- Neighborhood selection
  (E.g., Meinshausen and Bühlmann 2006; Ravikumar et al. 2011)
- Hypothesis testing
  (E.g., Drton and Perlman 2007)
- Robust extensions of GGMs
  (E.g., Finegold and Drton 2011; Vogel and Fried 2011; Liu, Han, and Zhang 2012; Bilodeau 2014)
- Semi-parametric extensions of GGMs
  (E.g., Liu, Lafferty, and Wasserman 2009; Liu et al. 2012)
- Joint estimation of multiple GGMs
  (E.g., Guo et al. 2011; Danaher, Wang, and Witten 2014; Ma and Michailidis 2016)
- GGMs with covariates
  (E.g., Yin and Li 2011; Li, Chun, and Zhao 2012; Cai et al. 2013; Chen et al. 2016; Zhang and Li 2022)

Problem in Pictures: Community Detection





Question: Can we infer the vertex labels on the left, and hence the communities on the right, given only the graph topology?

For general surveys on community detection see, e.g., Fortunato 2010; Fortunato and Newman 2022

Spotlight on... Stochastic Block Models

A stochastic block model (SBM) is essentially a mixture of classical random graphs.

Spotlight on. . . Stochastic Block Models

A stochastic block model (SBM) is essentially a mixture of classical random graphs.

Suppose each vertex $i \in V$ of a graph $G = (V, E)$ can belong to one of Q classes, say $\mathcal{C}_1, \ldots, \mathcal{C}_Q$. An SBM dictates that

$$Z_i \overset{i.i.d}{\sim} \text{Multinomial}(1, \alpha)$$

$$Y_{ij} \mid Z_i = z_i, Z_j = z_j \sim \text{Bernoulli}(\pi_{z_i, z_j}) \quad ,$$

for $\alpha = (\alpha_1, \ldots, \alpha_Q)$ and $\pi = [\pi_{qr}]$, where $Y_{ij} = Y_{ji}$, $Y_{ii} \equiv 0$, with $1 \leq i, j \leq n_v$.

Sources: Kolaczyk 2017; Zhao 2017; Abbe 2017

SBM Inference

Inference can focus on

- Parameter inference, i.e., for $\theta = (\alpha, \pi)$, where $\alpha = (\alpha_1, \ldots, \alpha_Q)$ and $\pi = [\pi_{qr}]$; and
- Class label inference, i.e., for $\mathbf{Z}$.

The latter corresponds to a model-based version of 'community detection'.

Profile Likelihood

For the purposes of community detection, the $Q \times Q$ parameter matrix π can be treated as a nuisance parameter. Given observed adjacency matrix $Y = y = [y_{ij}]$, this motivates definition of the estimator

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} \ell\left(y; \mathbf{z}, \hat{\pi}(\mathbf{z})\right) \quad ,$$

where

$$\hat{\pi}(\mathbf{z}) = \arg\max_{\pi} \ell\left(y; \mathbf{z}, \pi\right) \quad ,$$

with

$$\hat{\pi}_{qr}(\mathbf{z}) = \frac{1}{n_{qr}} \sum_{i<j} y_{ij} I\left(z_{iq} = 1, z_{jr} = 1\right) \quad ,$$

for each q and r, where n_{qr} is the maximum number of possible edges between classes q and r.

Properties and Implementation

Theorem (Bickel and Chen 2009)

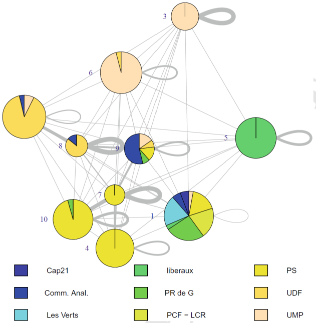
Assume some regularity conditions and sufficiently dense networks (expected average degree grows faster than $\log n_v$). Then up to permutation, $\mathbb{P}\left(\hat{\mathbf{Z}} = \mathbf{Z}\right) \longrightarrow 1$, as $n_v \longrightarrow \infty$.

Global optimization in this context is NP-hard.

In practice, approximate solution of the underlying expectation-maximization (EM) problem has been approached in various ways, with variational methods and belief propagation being popular.

See, e.g., Daudin, Picard, and Robin 2008; Hastings 2006

Illustration: French blog network




```
> library(blockmodels)
> A.fblog <- as.matrix(as_adjacency_matrix(fblog))
> fblog.sbm <- BM_bernoulli("SBM_sym", A.fblog,
+                           verbosity=0, plotting='')
> fblog.sbm$estimate()
```

Kolaczyk and Csárdi 2020, Ch 6.3

Caveat Emptor

Community detection, even under just the SBM and with only $K = 2$ symmetric classes, is decidedly complex, with subtleties abounding around phase transitions between types of recovery (exact, almost exact, partial, weak, distinguishable) as a function of model assumptions and choice of algorithm.

Bui, Chaudhuri, Leighton, Sipser '84	maxflow-mincut	$A = \Omega(1/n), B = o(n^{-1-4/((A+B)n)})$
Boppana '87	spectral meth.	$(A-B)/\sqrt{A+B} = \Omega(\sqrt{\log(n)/n})$
Dyer, Frieze '89	min-cut via degrees	$A - B = \Omega(1)$
Snijders, Nowicki '97	EM algo.	$A - B = \Omega(1)$
Jerrum, Sorkin '98	Metropolis aglo.	$A - B = \Omega(n^{-1/6+\epsilon})$
Condon, Karp '99	augmentation algo.	$A - B = \Omega(n^{-1/2+\epsilon})$
Carson, Impagliazzo '01	hill-climbing algo.	$A - B = \Omega(n^{-1/2}\log^4(n))$
McSherry '01	spectral meth.	$(A-B)/\sqrt{A} \geq \Omega(\sqrt{\log(n)/n})$
Bickel, Chen '09	N-G modularity	$(A-B)/\sqrt{A+B} = \Omega(\log(n)/\sqrt{n})$
Rohe, Chatterjee, Yu '11	spectral meth.	$A - B = \Omega(1)$

Source: Abbe 2017

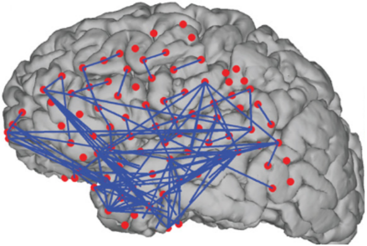
Shout Out to . . .

- Spectral clustering (E.g., Rohe, Chatterjee, and Yu 2011; Sussman et al. 2012; Jin 2015)
- Mixed-membership SBM (E.g., Airoldi et al. 2008)
- Degree-corrected SBM (E.g., Karrer and Newman 2011; Zhao, Levina, and Zhu 2012)
- Dynamic SBM (E.g., Yang et al. 2011; Xu and Hero 2014; Matias and Miele 2017)
- Multilayer SBM (E.g., Valles-Catala et al. 2016; Paul and Chen 2016)
- SBM/covariates (E.g., Binkiewicz, Vogelstein, and Rohe 2017; Zhang, Levina, and Zhu 2016)
- Weighted SBM
  (E.g., Mariadassou, Robin, and Vacher 2010; Zanghi et al. 2010; Aicher, Jacobs, and Clauset 2015)
- Number of communities & goodness of fit
  (E.g., Daudin, Picard, and Robin 2008; Zhao, Levina, and Zhu 2011; Bickel and Sarkar 2016; Lei 2016)

Where We Might Next Invest

Problem in Pictures: Multiple Networks

Female Response	Male Response
30	23
50	43
27	55
63	28
74	31
Female Average	**Male Average**
48.8	36



Question: What if instead of numbers our 'data points' were networks?

Source (right): Kramer et al. 2010

Spotlight on ... Fréchet means

Let $G = (V, E, W)$ be a *weighted* undirected graph, that is

- simple (i.e., no self-loops or multi-edges)
- connected (i.e., only one component)

and define the (combinatorial) graph Laplacian

$$L = D(W) - W \quad ,$$

where D is a diagonal matrix of weighted degrees,
i.e., $D_{jj} = d_j(W) = \sum_{i \neq j} w_{ij}$.

The Fréchet mean generalizes the notion of an 'average' to arbitrary metric spaces.

The Space of Network Graph Laplacians

Theorem (Ginestet et al. 2017)

Let the set $\mathcal{L}_{n_v}$ consist of $n_v \times n_v$ matrices A, satisfying:

(1) $Rank(A) = n_v - 1$,

(2) Symmetry, $A^T = A$,

(3) Positive semi-definiteness, $A \geq 0$,

(4) The entries in each row sum to 0,

(5) The off-diagonal entries are non-positive, $a_{ij} \leq 0$.

Then $\mathcal{L}_{n_v}$ is a manifold with corners, of dimension $n_v(n_v - 1)/2$. Furthermore, $\mathcal{L}_{n_v}$ is a convex subset of an affine space in $\mathbb{R}^{n_v^2}$ of dimension $n_v(n_v - 1)/2$.

Fréchet Mean

For $L_1, \ldots, L_n$ IID wrt some distribution Q,

and ρ_F the Frobenius norm, define the population

$$\mathbb{E}_Q[L] := \arg\min_{L \in \mathcal{L}_d} \int_{\mathcal{L}_d} \rho_F^2(L, \tilde{L}) Q(d\tilde{L})$$

and empirical

$$\widehat{L}_n := \arg\min_{L \in \mathcal{L}_d} \frac{1}{n} \sum_{i=1}^{n} \rho_F^2(L, L_i)$$

(Fréchet) means.

A Central Limit Theorem

Theorem (Ginestet et al. 2017)

If the expectation, $\Lambda := \mathbb{E}_Q[L]$, does not lie on the boundary of $\mathcal{L}_d$, and $\mathbb{P}_Q[U] > 0$, where U is an open subset of $\mathcal{L}_d$ with $\Lambda \in U$, then (under some further regularity conditions) we obtain the following convergence in distribution:

$$n^{1/2}(\phi(\widehat{L}_n) - \phi(\Lambda)) \longrightarrow N(0, \Sigma),$$

where $\Sigma := \mathbb{C}ov[\phi(L)]$ and $\phi(\cdot)$ denotes the half-vectorization of its matrix argument.

Hypothesis Testing

Corollary

Under the null hypothesis $H_0 : \mathbb{E}[L] = \Lambda_0$, we have,

$$T_1 := n\big(\phi(\widehat{L}) - \phi(\Lambda_0)\big)^T \widehat{\Sigma}^{-1}\big(\phi(\widehat{L}) - \phi(\Lambda_0)\big) \longrightarrow \chi^2_m,$$

with $m := \binom{d}{2}$ degrees of freedom, and where $\widehat{\Sigma}$ is the sample covariance.

Implementation

In order to use these results in practice, we require knowledge of Σ or, more realistically, for the sample covariance S to be stable.

For $n \gg O(n_v^2)$, it may be that S is stable, but for $n \ll O(n_v^2)$, we face a "large n, small p" problem.

The extensive literature on estimation of large, structured covariance/precision matrices from limited data can be exploited in this context.[1]

[1] Our applied work uses the approach of Schäfer and Strimmer 2005.

Illustration – 1000 Functional Connectomes

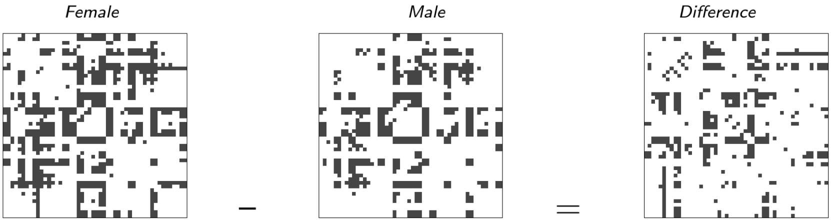
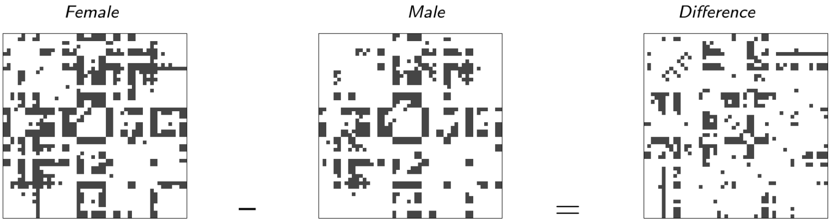


Illustration – 1000 Functional Connectomes



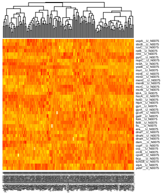
Unlike more naive (aka 'mass univariate') techniques, a Fréchet mean approach detects the difference in sexes *at sample sizes relevant to single labs*.

Shout Out to . . .

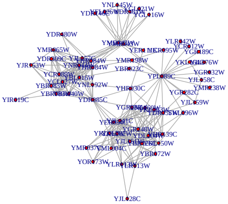
- Smooth manifolds (E.g., Bhattacharya and Patrangenaru 2003; Bhattacharya and Patrangenaru 2005)
- Tree spaces (E.g., Billera, Holmes, and Vogtmann 2001; Barden, Le, and Owen 2013; Wang and Marron 2007; Aydin et al. 2009)
- Symmetric PSD cone (E.g., Bonnabel and Sepulchre 2010; Krishnamachari and Varanasi 2013)
- Fréchet
  - ANOVA (E.G., Dubey and Müller 2019)
  - Regression (E.g., Petersen and Müller 2019; Tucker, Wu, and Müller 2021)
  - Mean for Unlabeled Networks (E.g., Kolaczyk et al. 2020)
- Graph embedding (See Cai, Zheng, and Chang 2018 for a survey.)

Problem in Pictures: Noisy Networks

Noisy Data | **Network from Data** | **Summary**



$\Longrightarrow$



$\Longrightarrow$ Density $= 0.14 \pm$???

Question: What is the uncertainty associated with network summaries we routinely report?

Spotlight on . . . Estimation of Subgraph Counts

Consider estimating the edge density

$$\delta = \frac{1}{n_v(n_v - 1)} \sum_{i \neq j} A_{ij} \quad ,$$

under a 'signal plus noise' model, where we observe only

$$Y_{ij} = A_{ij} I(\varepsilon_{ij} = 0) + I(\varepsilon_{ij} = 1)$$

with

$$\mathbb{P}(\varepsilon_{ij} = 1) = \alpha,\ \mathbb{P}(\varepsilon_{ij} = 0) = 1 - \alpha - \beta, \text{ and } \mathbb{P}(\varepsilon_{ij} = -1) = \beta,$$

where A is the adjacency matrix for G.

Chang, Kolaczyk, and Yao 2022

A Method-of-Moments Approach

In recent work, we have shown...

Impossibility Theorem ... α, β and δ cannot all be estimated from a single noisy network [i.e., analogous to a two-component mixture problem – true generally!]

A Method-of-Moments Approach

In recent work, we have shown...

Impossibility Theorem ... α, β and δ cannot all be estimated from a single noisy network [i.e., analogous to a two-component mixture problem – true generally!]

Somewhat possible with two networks ... method-of-moments estimation for β, δ, given known α, well-defined/behaved with two noisy versions of the same network [i.e., based on network average and first-order differences]

A Method-of-Moments Approach

In recent work, we have shown...

Impossibility Theorem ... α, β and δ cannot all be estimated from a single noisy network [i.e., analogous to a two-component mixture problem – true generally!]

Somewhat possible with two networks ... method-of-moments estimation for β, δ, given known α, well-defined/behaved with two noisy versions of the same network [i.e., based on network average and first-order differences]

Entirely possible with three networks ... method-of-moments estimation for α, β and δ well-defined/behaved with three noisy versions of the same network [i.e., augment above with second order differences]

Properties and Implementation

Theorem (Chang, Kolaczyk, and Yao 2022)

Let $N = n_v(n_v - 1)$ and assume iid errors. If $N_1 = N\delta \to \infty$ and $N_2 = N(1-\delta) \to \infty$, it holds that

$$\sqrt{N}(\hat{\alpha} - \alpha, \hat{\beta} - \beta, \hat{\delta} - \delta)^T \to_d \text{Normal}(0, \Sigma_2)\,,$$

provided that $\delta(1-\delta)(1-\alpha-\beta)^4 \geq c > 0$.

These results extend to the case of estimating density of arbitrary subgraphs and smooth functions thereof.

Complicated expressions for asymptotic (co)variances necessitates development of a novel bootstrap algorithm.

Illustration – Gene Coexpression Network

It is a standard exercise in computational biology to construct and analyze networks from gene expression data.

We illustrate[2] with generic correlation networks of 153 genes, deriving from 40 experiments (each replicated 3 times) in the bacteria *Escherichia coli (E. coli)*.

Family-wise error rate controled at the 0.05 level through a Bonferonni correction.

[2] Constructed as in Kolaczyk and Csárdi 2020, Ch 7.3.1.

Network Density Little Affected by Noise

- **Empirical Edge Densities**

  Approximately $0.073, 0.075$, and 0.074.

Network Density Little Affected by Noise

- **Empirical Edge Densities**
  Approximately $0.073, 0.075$, and 0.074.
- **Estimation with 'Known' α**
  Bonferroni control at 0.05 based on $11,628$ hypothesis tests yields a nominal $\alpha \approx 4.3 \times 10^{-6}$, which in turn yields estimates $\hat{\beta} = 0.456$ and $\hat{\delta} = 0.135$, with a 95% CI of $(0.131, 0.139)$ for the latter.

Network Density Little Affected by Noise

- **Empirical Edge Densities**
  Approximately $0.073, 0.075$, and 0.074.
- **Estimation with 'Known'** α
  Bonferroni control at 0.05 based on $11,628$ hypothesis tests yields a nominal $\alpha \approx 4.3 \times 10^{-6}$, which in turn yields estimates $\hat{\beta} = 0.456$ and $\hat{\delta} = 0.135$, with a 95% CI of $(0.131, 0.139)$ for the latter.
- **Estimation with Unknown** α **and** β
  Estimating α as well, we obtain $\hat{\alpha} = 0.024$, $\hat{\beta} = 0.232$, and $\hat{\delta} = 0.067$, with an accompanying 95% confidence interval for δ of $(0.06, 0.074)$.

Clustering Coefficient Changes Substantially!

- **Estimation of 2-stars, Triangles, and Clustering**

Source	# 2-Stars	# Triangles	Clustering Coeff.
Repl 1	19112	3373	0.53
Repl 2	22952	4814	0.63
Repl 3	21820	4349	0.60
Estimate	25248	7243	0.86

The accompanying 95% confidence interval for the clustering coefficient is $(0.81, 0.91)$.

Shout Out to ...

- Empirical studies (E.g., Hart, Ramani, and Marcotte 2006; Almquist 2012)
- Network denoising (E.g., Chatterjee 2015)
- Vertex classification (E.g., Priebe et al. 2015)
- Graph matching (E.g., Lyzinski 2018; Arroyo et al. 2021)
- Epidemic branching factors (E.g., Li, Sussman, and Kolaczyk 2020)

In addition, there is an increasingly active literature on the related (and still quite hard!) problem of uncertainty quantification from single networks drawn from random ensembles $\Pr(G)$, using extensions of bootstrapping, jackknifing, and the like.

[See P. Sarkar's talk next!]

Wrapping Up

Thank you!

Questions?

References i

Abbe, Emmanuel (2017). "Community detection and stochastic block models: recent developments". In: *The Journal of Machine Learning Research* 18.1, pp. 6446–6531.

Aicher, Christopher, Abigail Z Jacobs, and Aaron Clauset (2015). "Learning latent block structure in weighted networks". In: *Journal of Complex Networks* 3.2, pp. 221–248.

Airoldi, Edo M et al. (2008). "Mixed membership stochastic blockmodels". In: *Advances in neural information processing systems* 21.

Almquist, Zack W (2012). "Random errors in egocentric networks". In: *Social networks* 34.4, pp. 493–505.

Arroyo, Jesús et al. (2021). "Maximum likelihood estimation and graph matching in errorfully observed networks". In: *Journal of Computational and Graphical Statistics* 30.4, pp. 1111–1123.

Aydin, Burcu et al. (2009). "A principal component analysis for trees". In: *The Annals of Applied Statistics*, pp. 1597–1615.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont (2008). "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data". In: *The Journal of Machine Learning Research* 9, pp. 485–516.

References ii

Barden, Dennis, Huiling Le, and Megan Owen (2013). "Central limit theorems for Fréchet means in the space of phylogenetic trees". In: *Electronic journal of probability* 18, pp. 1–25.

Bhattacharya, Rabi and Vic Patrangenaru (2003). "Large sample theory of intrinsic and extrinsic sample means on manifolds". In: *The Annals of Statistics* 31.1, pp. 1–29.

— (2005). "Large sample theory of intrinsic and extrinsic sample means on manifolds—II". In: *The Annals of Statistics* 33.3, pp. 1225–1259.

Bickel, Peter J and Aiyou Chen (2009). "A nonparametric view of network models and Newman–Girvan and other modularities". In: *Proceedings of the National Academy of Sciences* 106.50, pp. 21068–21073.

Bickel, Peter J and Purnamrita Sarkar (2016). "Hypothesis testing for automated community detection in networks". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78.1, pp. 253–273.

Billera, Louis J, Susan P Holmes, and Karen Vogtmann (2001). "Geometry of the space of phylogenetic trees". In: *Advances in Applied Mathematics* 27.4, pp. 733–767.

Bilodeau, Martin (2014). "Graphical lassos for meta-elliptical distributions". In: *Canadian Journal of Statistics* 42.2, pp. 185–203.

Binkiewicz, Norbert, Joshua T Vogelstein, and Karl Rohe (2017). "Covariate-assisted spectral clustering". In: *Biometrika* 104.2, pp. 361–377.

References iii

Bonnabel, Silvere and Rodolphe Sepulchre (2010). "Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank". In: *SIAM Journal on Matrix Analysis and Applications* 31.3, pp. 1055–1070.

Brugere, Ivan, Brian Gallagher, and Tanya Y Berger-Wolf (2018). "Network structure inference, a survey: Motivations, methods, and applications". In: *ACM Computing Surveys (CSUR)* 51.2, pp. 1–39.

Cai, Hongyun, Vincent W Zheng, and Kevin Chen-Chuan Chang (2018). "A comprehensive survey of graph embedding: Problems, techniques, and applications". In: *IEEE Transactions on Knowledge and Data Engineering* 30.9, pp. 1616–1637.

Cai, T Tony et al. (2013). "Covariate-adjusted precision matrix estimation with an application in genetical genomics". In: *Biometrika* 100.1, pp. 139–156.

Chang, Jinyuan, Eric D Kolaczyk, and Qiwei Yao (2022). "Estimation of subgraph densities in noisy networks". In: *Journal of the American Statistical Association* 117.537, pp. 361–374.

Chatterjee, Sourav (2015). "Matrix estimation by universal singular value thresholding". In: *The Annals of Statistics* 43.1, pp. 177–214.

Chen, Mengjie et al. (2016). "Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model". In: *Journal of the American Statistical Association* 111.513, pp. 394–406.

References iv

Danaher, Patrick, Pei Wang, and Daniela M Witten (2014). "The joint graphical lasso for inverse covariance estimation across multiple classes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.2, pp. 373–397.

Daudin, J-J, Franck Picard, and Stéphane Robin (2008). "A mixture model for random graphs". In: *Statistics and computing* 18.2, pp. 173–183.

Drton, Mathias and Marloes H Maathuis (2017). "Structure learning in graphical modeling". In: *Annual Review of Statistics and Its Application* 4, pp. 365–393.

Drton, Mathias and Michael D Perlman (2007). "Multiple testing and error control in Gaussian graphical model selection". In: *Statistical Science* 22.3, pp. 430–449.

Dubey, Paromita and Hans-Georg Müller (2019). "Fréchet analysis of variance for random objects". In: *Biometrika* 106.4, pp. 803–821.

Finegold, Michael and Mathias Drton (2011). "Robust graphical modeling of gene networks using classical and alternative t-distributions". In: *The Annals of Applied Statistics* 5.2A, pp. 1057–1080.

Fortunato, Santo (2010). "Community detection in graphs". In: *Physics reports* 486.3-5, pp. 75–174.

References v

Fortunato, Santo and Mark EJ Newman (2022). "20 years of network community detection". In: *Nature Physics*, pp. 1–3.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2008). "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3, pp. 432–441.

Gao, Xin et al. (2012). "Tuning parameter selection for penalized likelihood estimation of Gaussian graphical model". In: *Statistica Sinica*, pp. 1123–1146.

Ginestet, Cedric E et al. (2017). "Hypothesis testing for network data in functional neuroimaging". In: *The Annals of Applied Statistics*, pp. 725–750.

Guo, Jian et al. (2011). "Joint estimation of multiple graphical models". In: *Biometrika* 98.1, pp. 1–15.

Hart, G Traver, Arun K Ramani, and Edward M Marcotte (2006). "How complete are current yeast and human protein-interaction networks?" In: *Genome biology* 7.11, pp. 1–9.

Hastings, Matthew B (2006). "Community detection as an inference problem". In: *Physical Review E* 74.3, p. 035102.

Hsieh, Cho-Jui et al. (2013). "BIG & QUIC: Sparse inverse covariance estimation for a million variables". In: *Advances in neural information processing systems* 26.

References vi

Jin, Jiashun (2015). "Fast community detection by score". In: *The Annals of Statistics* 43.1, pp. 57–89.

Karrer, Brian and Mark EJ Newman (2011). "Stochastic blockmodels and community structure in networks". In: *Physical review E* 83.1, p. 016107.

Kolaczyk, Eric D (2009). *Statistical Analysis of Network Data*. Springer.

— (2017). *Topics at the Frontier of Statistics and Network Analysis:(re) visiting the Foundations*. Cambridge University Press.

Kolaczyk, Eric D and Gábor Csárdi (2020). *Statistical analysis of network data with R, 2nd Edition*. Springer.

Kolaczyk, Eric D et al. (2020). "Averages of unlabeled networks: Geometric characterization and asymptotic behavior". In: *The Annals of Statistics* 48.1, pp. 514–538.

Kramer, Mark A et al. (2010). "Coalescence and fragmentation of cortical networks during focal seizures". In: *Journal of Neuroscience* 30.30, pp. 10076–10085.

Krishnamachari, Rajesh T and Mahesh K Varanasi (2013). "On the geometry and quantization of manifolds of positive semi-definite matrices". In: *IEEE Transactions on signal processing* 61.18, pp. 4587–4599.

References vii

Lei, Jing (2016). "A goodness-of-fit test for stochastic block models". In: *The Annals of Statistics* 44.1, pp. 401–424.

Li, Bing, Hyonho Chun, and Hongyu Zhao (2012). "Sparse estimation of conditional graphical models with application to gene networks". In: *Journal of the American Statistical Association* 107.497, pp. 152–167.

Li, Wenrui, Daniel L Sussman, and Eric D Kolaczyk (2020). "Estimation of the epidemic branching factor in noisy contact networks". In: *arXiv preprint arXiv:2002.05763*.

Liu, Han, Fang Han, and Cun-hui Zhang (2012). "Transelliptical graphical models". In: *Advances in neural information processing systems* 25.

Liu, Han, John Lafferty, and Larry Wasserman (2009). "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs.". In: *Journal of Machine Learning Research* 10.10.

Liu, Han, Kathryn Roeder, and Larry Wasserman (2010). "Stability approach to regularization selection (stars) for high dimensional graphical models". In: *Advances in neural information processing systems* 23.

Liu, Han et al. (2012). "High-dimensional semiparametric Gaussian copula graphical models". In: *The Annals of Statistics* 40.4, pp. 2293–2326.

Lyzinski, Vince (2018). "Information recovery in shuffled graphs via graph matching". In: *IEEE Transactions on Information Theory* 64.5, pp. 3254–3273.

References viii

Ma, Jing and George Michailidis (2016). "Joint structural estimation of multiple graphical models". In: *The Journal of Machine Learning Research* 17.1, pp. 5777–5824.

Mariadassou, Mahendra, Stéphane Robin, and Corinne Vacher (2010). "Uncovering latent structure in valued graphs: a variational approach". In: *The Annals of Applied Statistics* 4.2, pp. 715–742.

Matias, Catherine and Vincent Miele (2017). "Statistical clustering of temporal networks through a dynamic stochastic block model". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.4, pp. 1119–1141.

Meinshausen, Nicolai and Peter Bühlmann (2006). "High-dimensional graphs and variable selection with the lasso". In: *The annals of statistics* 34.3, pp. 1436–1462.

Nguyen, Hung et al. (2021). "A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data". In: *Briefings in bioinformatics* 22.3, bbaa190.

Paul, Subhadeep and Yuguo Chen (2016). "Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel". In: *Electronic Journal of Statistics* 10.2, pp. 3807–3870.

Petersen, Alexander and Hans-Georg Müller (2019). "Fréchet regression for random objects with Euclidean predictors". In: *The Annals of Statistics* 47.2, pp. 691–719.

References ix

Priebe, Carey E et al. (2015). "Statistical inference on errorfully observed graphs". In: *Journal of Computational and Graphical Statistics* 24.4, pp. 930–953.

Ravikumar, Pradeep et al. (2011). "High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence". In: *Electronic Journal of Statistics* 5, pp. 935–980.

Rohe, Karl, Sourav Chatterjee, and Bin Yu (2011). "Spectral clustering and the high-dimensional stochastic blockmodel". In: *The Annals of Statistics* 39.4, pp. 1878–1915.

Schäfer, Juliane and Korbinian Strimmer (2005). "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics". In: *Statistical applications in genetics and molecular biology* 4.1.

Sussman, Daniel L et al. (2012). "A consistent adjacency spectral embedding for stochastic blockmodel graphs". In: *Journal of the American Statistical Association* 107.499, pp. 1119–1128.

Tucker, Danielle C, Yichao Wu, and Hans-Georg Müller (2021). "Variable Selection for Global Fréchet Regression". In: *Journal of the American Statistical Association*, pp. 1–15.

Valles-Catala, Toni et al. (2016). "Multilayer stochastic block models reveal the multilayer structure of complex networks". In: *Physical Review X* 6.1, p. 011036.

Vogel, Daniel and Roland Fried (2011). "Elliptical graphical modelling". In: *Biometrika* 98.4, pp. 935–951.

References x

Wang, Haonan and JS Marron (2007). "Object oriented data analysis: Sets of trees". In: *The Annals of Statistics* 35.5, pp. 1849–1873.

Witten, Daniela M, Jerome H Friedman, and Noah Simon (2011). "New insights and faster computations for the graphical lasso". In: *Journal of Computational and Graphical Statistics* 20.4, pp. 892–900.

Xu, Kevin S and Alfred O Hero (2014). "Dynamic stochastic blockmodels for time-evolving social networks". In: *IEEE Journal of Selected Topics in Signal Processing* 8.4, pp. 552–562.

Yang, Tianbao et al. (2011). "Detecting communities and their evolutions in dynamic social networks—a Bayesian approach". In: *Machine learning* 82.2, pp. 157–189.

Yin, Jianxin and Hongzhe Li (2011). "A sparse conditional Gaussian graphical model for analysis of genetical genomics data". In: *The annals of applied statistics* 5.4, p. 2630.

Yuan, Ming and Yi Lin (2007). "Model selection and estimation in the Gaussian graphical model". In: *Biometrika* 94.1, pp. 19–35.

Zanghi, Hugo et al. (2010). "Strategies for online inference of model-based clustering in large and growing networks". In: *The Annals of Applied Statistics* 4.2, pp. 687–714.

Zhang, Jingfei and Yi Li (2022). "High-Dimensional Gaussian Graphical Regression Models with Covariates". In: *Journal of the American Statistical Association*, pp. 1–13.

References xi

Zhang, Yuan, Elizaveta Levina, and Ji Zhu (2016). "Community detection in networks with node features". In: *Electronic Journal of Statistics* 10.2, pp. 3153–3178.

Zhao, Yunpeng (2017). "A survey on theoretical advances of community detection in networks". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 9.5, e1403.

Zhao, Yunpeng, Elizaveta Levina, and Ji Zhu (2011). "Community extraction for social networks". In: *Proceedings of the National Academy of Sciences* 108.18, pp. 7321–7326.

— (2012). "Consistency of community detection in networks under degree-corrected stochastic block models". In: *The Annals of Statistics* 40.4, pp. 2266–2292.