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!
For each topic area, I will

1. Describe a canonical problem(s) through pictures;

2. Shine a “Spotlight on . . .” a key solution(s);

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
Network Topology Inference

A Truly Substantial Literature

Brugere, Gallagher, and Berger-Wolf 2018
Nguyen et al. 2021, “A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data”.
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 \textit{graphical lasso (glasso)} estimator is an $\ell_1$-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

$$\{i, j\} \in \hat{E}^{gl} \text{ iff } \hat{\kappa}^{gl}_{ij} \neq 0.$$
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 \( \lambda \) can be chosen adaptively in various ways (e.g., BIC, stability selection, etc.).

**Illustration: Gene coexpression in Ecoli**

```r
> 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
Network Topology Inference

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)


**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
Community Detection

Spotlight on... Stochastic Block Models

A stochastic block model (SBM) is essentially a mixture of classical random graphs.
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 \( C_1, \ldots, C_Q \). An SBM dictates that

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

\[
Y_{ij} | Z_i = z_i, Z_j = z_j \sim \text{Bernoulli}(\pi_{z_i, z_j})
\]

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
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 $Z$.

The latter corresponds to a model-based version of ‘community detection’.
Community Detection

Profile Likelihood

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

$$\hat{z} = \arg \max_z \ell(y; z, \hat{\pi}(z)),$$

where

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

with

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

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, \(\mathbb{P}(\hat{Z} = Z) \rightarrow 1\), as \(n_v \rightarrow \infty\).

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
Community Detection

Illustration: French blog network

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

Kolaczyk and Csárdi 2020, Ch 6.3
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.

<table>
<thead>
<tr>
<th>Source</th>
<th>Algorithm Type</th>
<th>Recovery Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bui, Chaudhuri, Leighton, Sipser ’84</td>
<td>maxflow-mincut</td>
<td>$A = \Omega(1/n), B = o(n^{-1-4/((A+B)n)})$</td>
</tr>
<tr>
<td>Boppana ’87</td>
<td>spectral meth.</td>
<td>$(A - B)/\sqrt{A + B} = \Omega(\sqrt{\log(n)/n})$</td>
</tr>
<tr>
<td>Dyer, Frieze ’89</td>
<td>min-cut via degrees</td>
<td>$A - B = \Omega(1)$</td>
</tr>
<tr>
<td>Snijders, Nowicki ’97</td>
<td>EM algo.</td>
<td>$A - B = \Omega(1)$</td>
</tr>
<tr>
<td>Jerrum, Sorkin ’98</td>
<td>Metropolis aglo.</td>
<td>$A - B = \Omega(n^{-1/6+\epsilon})$</td>
</tr>
<tr>
<td>Condon, Karp ’99</td>
<td>augmentation algo.</td>
<td>$A - B = \Omega(n^{-1/2+\epsilon})$</td>
</tr>
<tr>
<td>Carson, Impagliazzo ’01</td>
<td>hill-climbing algo.</td>
<td>$A - B = \Omega(n^{-1/2} \log^3(n))$</td>
</tr>
<tr>
<td>McSherry ’01</td>
<td>spectral meth.</td>
<td>$(A - B)/\sqrt{A} \geq \Omega(\sqrt{\log(n)/n})$</td>
</tr>
<tr>
<td>Bickel, Chen ’09</td>
<td>N-G modularity</td>
<td>$(A - B)/\sqrt{A + B} = \Omega(\log(n)/\sqrt{n})$</td>
</tr>
<tr>
<td>Rohe, Chatterjee, Yu ’11</td>
<td>spectral meth.</td>
<td>$A - B = \Omega(1)$</td>
</tr>
</tbody>
</table>

Source: Abbe 2017
Community Detection

**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

<table>
<thead>
<tr>
<th>Female Response</th>
<th>Male Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>50</td>
<td>43</td>
</tr>
<tr>
<td>27</td>
<td>55</td>
</tr>
<tr>
<td>63</td>
<td>28</td>
</tr>
<tr>
<td>74</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Female Average</th>
<th>Male Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>48.8</td>
<td>36</td>
</tr>
</tbody>
</table>

**Question:** What if instead of numbers our ‘data points’ were networks?

Source (right): Kramer et al. 2010
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

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

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. $\text{Rank}(A) = n_v - 1$,
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 $n_v(n_v - 1)/2$. 

See also Kolaczyk et al. 2020
Fréchet Mean

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

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

(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:

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

where $\Sigma := \text{Cov}[\phi(L)]$ 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,

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

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

Implementation

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

For $n \gg O(n^2_V)$, it may be that $S$ is stable, but for $n \ll O(n^2_V)$, 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.
Unlike more naive (aka 'mass univariate') techniques, a Fréchet mean approach detects the difference in sexes at sample sizes relevant to single labs.
Unlike more naive (aka ‘mass univariate’) techniques, a Fréchet mean approach detects the difference in sexes at sample sizes relevant to single labs.
Multiple Networks

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

⇒  ⇒  Density = 0.14 ± ???

Question: What is the uncertainty associated with network summaries we routinely report?
Consider estimating the edge density

\[ \hat{\delta} = \frac{1}{n_v(n_v - 1)} \sum_{i \neq j} A_{ij}, \]

under a ‘signal plus noise’ model, where we observe only

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

with

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

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**... $\alpha, \beta$ and $\delta$ 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**... $\alpha$, $\beta$, and $\delta$ 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 $\beta$, $\delta$, given known $\alpha$, 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** ... \( \alpha, \beta \) and \( \delta \) 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 \( \beta, \delta \), given known \( \alpha \), 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 \( \alpha, \beta \) and \( \delta \) 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 controlled at the 0.05 level through a Bonferroni 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’ $\alpha$**
  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’ $\alpha$**
  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 $\alpha$ and $\beta$**
  Estimating $\alpha$ as well, we obtain $\hat{\alpha} = 0.024$, $\hat{\beta} = 0.232$, and $\hat{\delta} = 0.067$, with an accompanying 95% confidence interval for $\delta$ of (0.06, 0.074).
Clustering Coefficient Changes Substantially!

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

<table>
<thead>
<tr>
<th>Source</th>
<th># 2-Stars</th>
<th># Triangles</th>
<th>Clustering Coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repl 1</td>
<td>19112</td>
<td>3373</td>
<td>0.53</td>
</tr>
<tr>
<td>Repl 2</td>
<td>22952</td>
<td>4814</td>
<td>0.63</td>
</tr>
<tr>
<td>Repl 3</td>
<td>21820</td>
<td>4349</td>
<td>0.60</td>
</tr>
<tr>
<td>Estimate</td>
<td>25248</td>
<td>7243</td>
<td>0.86</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>In:</th>
</tr>
</thead>
</table>
References

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.


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


References


References


References


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.


References  ix


References


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.


References


