Graphical Models

With a focus towards interimly missing data

Dominic DiSanto

Harvard University, Department of Biostatistics

November 23, 2023

Outline

1. Introduction to Graphical Models
2. Estimation for Complete Data
   - Neighborhood Selection
   - Graphical Lasso
   - Further Notes
3. Applications with Missingness

Disclaimers

- Historical coverage is to the best of my ability and time constraint, please correct me with additional information
- Interrupt with any questions, clarification, confusion, etc.
- This is far from a comprehensive treatment, but I attempt to be holistic in my coverage

Outline (Redux)

Graph Theory Origins [1, 2]

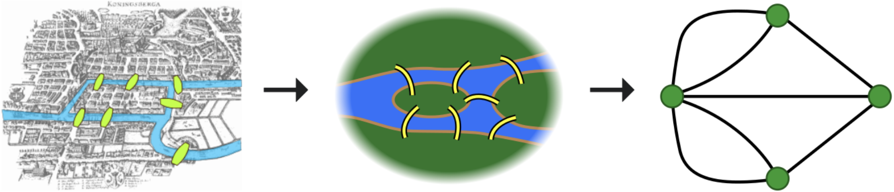


Figure: Euler's Bridges Conceptualization (Recreation)

[1]

---

[1] Image taken from Wikipedia (`https://en.wikipedia.org/wiki/Seven_Bridges_of_Konigsberg`)

Downloadable Slides

Early Applications of Graphs in Mathematics

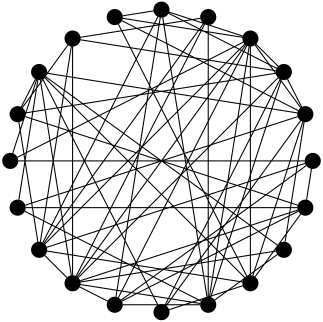
- Graph theory attributed to begin with Euler and the "Seven Bridges of Königsberg" ($\sim$1736)
- Random graph theory began developing in $\sim$1940's (Moreno and Jennings) but most notably with the Erdös-Rényi random graph (1958)
- Ising model ($\sim$1920's) - proposed graphical model of interactions of atomic spin
- Statistical "beginnings"[2] as a subset of methods for contingency tables and log-linear models ($\sim$1970's)
- Judea Pearl $\sim$1980's for causal intepretation of Bayesian networks
- Modern interest in related regularized M-estimation problems and graphical neural networks

[2] "Beginnings" as earlier applications in physics were probabilistic by construction, but this may be seen as the earliest application purely from a statistical point of view, i.e. without considering physical phenomena

Graphical Model Motivation

Suppose you have 20 random variables*, how do you model their interrelationship?
*Consider any of the following:

- General -omic data
- Spatial data
- Computational neuroscience data
- Clinical language (see: EHR LLM[a]) data
- Time-series data



[a] Electronic Healthcare Record Large Language Model

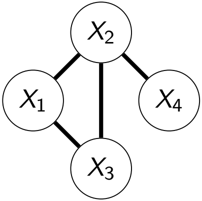
Graphs

- Graphs are a natural way to represent interrelationships among our data!
- Present nice properties for estimation of joint distributions
  - Can avail existing graphical algorithms
  - Ability to characterize conditional (in)dependencies
- Probabilistic graphical modelling provide a general formalism of many existing methods in statistics (e.g. Bayesian hierarchical modelling, Hidden Markov Models, Kalman filter)
- EDIT Reference Wainwright, Jordan for further apps

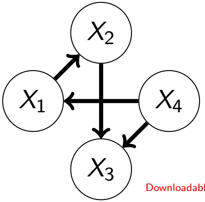
Graphs

- Consider random vector $X \sim N(\boldsymbol{\mu}, \Sigma)$ and precision matrix $\Theta \equiv \Sigma^{-1}$
  - Interested in estimating Σ to characterize joint distribution f_X
- Can construct a resulting graph $\mathcal{G} = (V, E)$, $V = X, E \subseteq V \times V$
- Gaussianity gives us the nice property that $\Theta_{ij} = 0 \Leftrightarrow X_i \perp X_j | X_{-\{i,j\}}$

Undirected Graph



Directed Graph



Notation/Nomenclature

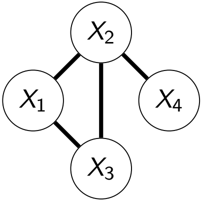
Omitting some philosophical discrepancies

- Directed (Acyclic) Graph $\Leftrightarrow$ Bayesian network
- Undirected graph $\Leftrightarrow$ Markov network / Markov random field

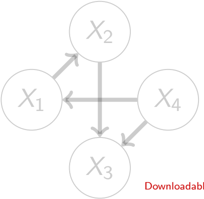
Graphs

- Consider random vector $X \sim N(\boldsymbol{\mu}, \Sigma)$ and precision matrix $\Theta \equiv \Sigma^{-1}$
  - Interested in estimating Σ to characterize joint distribution f_X
- Can construct a resulting graph $\mathcal{G} = (V, E)$, $V = X, E \subseteq V \times V$
- Gaussianity gives us the nice property that $\Theta_{ij} = 0 \Leftrightarrow X_i \perp X_j | X_{-\{i,j\}}$

Undirected Graph



Directed Graph



Outline (Redux)

Neighborhood Selection

Graphical Lasso

Further Notes

- Thus far we have assume Gaussianity, extensions exist (cite Witten paper on mixed graphs, among others(?))

Outline (Redux)

Erose Data

- Cite [3]

Conclusion

- Graphs are a powerful representation of your multivariate data (intuitively and algorithmically)
- Useful, theoretical extensions may follow more immediately under the graphical model formalism
- Extensions beyond Gaussianity substantially increase (theoretical) complexity

References I

- Some diagrams generated in conjunction with ChatGPT 3.5

[1] Imperatorskaia akademia nauk (Russia). *Commentarii Academiae scientiarum imperialis Petropolitanae*. lat. Petropolis, Typis Academiae, 1726.

[2] Rob Shields. "Cultural Topology: The Seven Bridges of Königsburg, 1736". en. In: *Theory, Culture & Society* 29.4-5 (July 2012). Publisher: SAGE Publications Ltd, pp. 43–57.

[3] Lili Zheng. *GI-JOE: Graph Inference when Joint Observations are Erose*. Mar. 2023.

Appendix Slides

Time-Series Data

- Consider that our repeated observations are time-indexed:
  - $\{X_j(t), t \in \mathcal{T}, j = 1, ..., N\}, X_j \in \mathbb{R}^d$
- Graphical perspective of vector auto-regressive models
  - $X_d(t) = \varepsilon_d(t) + \sum_{j \neq d} \sum_{t \in \mathcal{T}} \alpha_t X_j(t)$
- Can infer "Granger causal" relationships
  - Causal relationships for some time-series using prior data from a *different time series*

See Michael Eichler's "*Granger-causality graphs for multivariate time series*" (2007) and Dahlhaus's and Eichler's (2003) "*Causality and graphical models in time series*" for further discussion