

# Resources for Probabilistic Network Models

## 1 Background Readings

The readings below include many classical works in the study of networks, canonical examples of the techniques to be explored throughout the course, and useful review papers. Readings are listed roughly in the order in which they will become relevant and those most likely to be useful for this course are in blue text.

### 1.1 Introduction to networks and random graphs

Stephen E. Fienberg. “A Brief History of Statistical Models for Network Analysis and Open Challenges”. In: *Journal of Computational and Graphical Statistics* 21 (2012), pp. 825–839. URL: <https://api.semanticscholar.org/CorpusID:52232135>

Anna Goldenberg et al. “A survey of statistical network models”. In: *Foundations and Trends in Machine Learning* 2.2 (2010), pp. 129–233

Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. 2010

Tom AB Snijders. “Statistical models for social networks”. In: *Annual Review of Sociology* 37 (2011), pp. 131–153

### 1.2 Stochastic block models

Clement Lee and Darren Wilkinson. “A review of stochastic block models and extensions for graph clustering”. In: *Applied Network Science* 4 (Dec. 2019). DOI: 10.1007/s41109-019-0232-2

Krzysztof Nowicki and Tom A. B Snijders. “Estimation and Prediction for Stochastic Blockstructures”. In: *Journal of the American Statistical Association* 96.455 (2001), pp. 1077–1087. DOI: 10.1198/016214501753208735. eprint: <https://doi.org/10.1198/016214501753208735>. URL: <https://doi.org/10.1198/016214501753208735>

Brian Karrer and M. E. J. Newman. “Stochastic blockmodels and community structure in networks”. In: *Physical Review E* 83.1 (Jan. 2011). DOI: 10.1103/physreve.83.016107. URL: <https://doi.org/10.1103/physreve.83.016107>

Edoardo M. Airol di et al. “Mixed Membership Stochastic Blockmodels”. In: *Journal of Machine Learning Research* 9.65 (2008), pp. 1981–2014. URL: <http://jmlr.org/papers/v9/airoldi08a.html>

Catherine Matias and Vincent Miele. *Statistical Clustering of Temporal Networks Through a Dynamic Stochastic Block Model*. 2016. arXiv: 1506.07464 [stat.ME]

Sirio Legramanti et al. “Extended stochastic block models with application to criminal networks”. In: *The Annals of Applied Statistics* 16.4 (2022), pp. 2369–2395. DOI: 10.1214/21-AOAS1595. URL: <https://doi.org/10.1214/21-AOAS1595>

Sara Wade and Zoubin Ghahramani. “Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion)”. In: *Bayesian Analysis* 13.2 (2018), pp. 559–626. DOI: 10.1214/17-BA1073. URL: <https://doi.org/10.1214/17-BA1073>

### 1.3 Latent space network models

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. “Latent Space Approaches to Social Network Analysis”. In: *Journal of the American Statistical Association* 97.460 (2002), pp. 1090–1098. DOI: 10.1198/016214502388618906. eprint: <https://doi.org/10.1198/016214502388618906>. URL: <https://doi.org/10.1198/016214502388618906>

Peter D. Hoff. *Modeling homophily and stochastic equivalence in symmetric relational data*. 2007. arXiv: 0711.1146 [stat.ME]

A. Bhattacharya and D. B. Dunson. “Sparse Bayesian infinite factor models”. In: *Biometrika* 98.2 (June 2011), pp. 291–306. ISSN: 0006-3444. DOI: 10.1093/biomet/asr013. eprint: <https://academic.oup.com/biomet/article-pdf/98/2/291/46695653/asr013.pdf>. URL: <https://doi.org/10.1093/biomet/asr013>

Daniele Durante et al. “Bayesian Modelling of Networks in Complex Business Intelligence Problems”. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 66.3 (July 2016), pp. 555–580. ISSN: 0035-9254. DOI: 10.1111/rssc.12168. eprint: [https://academic.oup.com/jrsssc/article-pdf/66/3/555/49361259/jrsssc\\_66\\_3\\_555.pdf](https://academic.oup.com/jrsssc/article-pdf/66/3/555/49361259/jrsssc_66_3_555.pdf). URL: <https://doi.org/10.1111/rssc.12168>

Georgia Papadogeorgou et al. “Covariate-Informed Latent Interaction Models: Addressing Geographic and Taxonomic Bias in Predicting Bird–Plant Interactions”. In: *Journal of the American Statistical Association* 0.0 (2023), pp. 1–12. DOI: 10.1080/01621459.2023.2208390. eprint: <https://doi.org/10.1080/01621459.2023.2208390>. URL: <https://doi.org/10.1080/01621459.2023.2208390>

Evan Poworoznek, Federico Ferrari, and David Dunson. “Efficiently resolving rotational ambiguity in Bayesian matrix sampling with matching”. In: *arXiv preprint arXiv:2107.13783* (2021)

D. Durante and D. B. Dunson. “Nonparametric Bayes dynamic modelling of relational data”. In: *Biometrika* 101.4 (Oct. 2014), pp. 883–898. ISSN: 1464-3510. DOI: 10.1093/biomet/asu040. URL: <http://dx.doi.org/10.1093/biomet/asu040>

Daniele Durante, David B. Dunson, and Joshua T. Vogelstein. “Nonparametric Bayes Modeling of Populations of Networks”. In: *Journal of the American Statistical Association* 112.520 (2017), pp. 1516–1530. DOI: 10.1080/01621459.2016.1219260. eprint: <https://doi.org/10.1080/01621459.2016.1219260>. URL: <https://doi.org/10.1080/01621459.2016.1219260>

Daniele Durante and David B. Dunson. “Bayesian Inference and Testing of Group Differences in Brain Networks”. In: *Bayesian Analysis* 13.1 (2018), pp. 29–58. DOI: 10.1214/16-BA1030. URL: <https://doi.org/10.1214/16-BA1030>

### 1.4 MCMC Background

Andrew Gelman, Jessica Hwang, and Aki Vehtari. *Understanding predictive information criteria for Bayesian models*. 2013. arXiv: 1307.5928 [stat.ME]

Michael Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. 2018. arXiv: 1701.02434 [stat.ME]

Matthew D Hoffman, Andrew Gelman, et al. “The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.” In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623

Christian P. Robert et al. *Accelerating MCMC Algorithms*. 2018. arXiv: 1804.02719 [stat.CO]

## 2 Data

Sample and simulated data sets will be provided and created in the course labs, available on Github. The repositories below contain additional data sets which may be of interest:

Stanford Large Network Dataset Collection

Network Data Repository

Duke Network Analysis Center

Kenneth R. French’s Data Library

## References

- [1] Edoardo M. Airoldi et al. “Mixed Membership Stochastic Blockmodels”. In: *Journal of Machine Learning Research* 9.65 (2008), pp. 1981–2014. URL: <http://jmlr.org/papers/v9/airoldi08a.html>.
- [2] Michael Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. 2018. arXiv: 1701.02434 [stat.ME].
- [3] A. Bhattacharya and D. B. Dunson. “Sparse Bayesian infinite factor models”. In: *Biometrika* 98.2 (June 2011), pp. 291–306. ISSN: 0006-3444. DOI: 10.1093/biomet/asr013. eprint: <https://academic.oup.com/biomet/article-pdf/98/2/291/46695653/asr013.pdf>. URL: <https://doi.org/10.1093/biomet/asr013>.
- [4] D. Durante and D. B. Dunson. “Nonparametric Bayes dynamic modelling of relational data”. In: *Biometrika* 101.4 (Oct. 2014), pp. 883–898. ISSN: 1464-3510. DOI: 10.1093/biomet/asu040. URL: <http://dx.doi.org/10.1093/biomet/asu040>.
- [5] Daniele Durante and David B. Dunson. “Bayesian Inference and Testing of Group Differences in Brain Networks”. In: *Bayesian Analysis* 13.1 (2018), pp. 29–58. DOI: 10.1214/16-BA1030. URL: <https://doi.org/10.1214/16-BA1030>.
- [6] Daniele Durante, David B. Dunson, and Joshua T. Vogelstein. “Nonparametric Bayes Modeling of Populations of Networks”. In: *Journal of the American Statistical Association* 112.520 (2017), pp. 1516–1530. DOI: 10.1080/01621459.2016.1219260. eprint: <https://doi.org/10.1080/01621459.2016.1219260>. URL: <https://doi.org/10.1080/01621459.2016.1219260>.
- [7] Daniele Durante et al. “Bayesian Modelling of Networks in Complex Business Intelligence Problems”. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 66.3 (July 2016), pp. 555–580. ISSN: 0035-9254. DOI: 10.1111/rssc.12168. eprint: [https://academic.oup.com/jrsssc/article-pdf/66/3/555/49361259/jrsssc\\_66\\_3\\_555.pdf](https://academic.oup.com/jrsssc/article-pdf/66/3/555/49361259/jrsssc_66_3_555.pdf). URL: <https://doi.org/10.1111/rssc.12168>.
- [8] Stephen E. Fienberg. “A Brief History of Statistical Models for Network Analysis and Open Challenges”. In: *Journal of Computational and Graphical Statistics* 21 (2012), pp. 825–839. URL: <https://api.semanticscholar.org/CorpusID:52232135>.
- [9] Andrew Gelman, Jessica Hwang, and Aki Vehtari. *Understanding predictive information criteria for Bayesian models*. 2013. arXiv: 1307.5928 [stat.ME].

- [10] Anna Goldenberg et al. “A survey of statistical network models”. In: *Foundations and Trends in Machine Learning* 2.2 (2010), pp. 129–233.
- [11] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. “Latent Space Approaches to Social Network Analysis”. In: *Journal of the American Statistical Association* 97.460 (2002), pp. 1090–1098. DOI: 10.1198/016214502388618906. eprint: <https://doi.org/10.1198/016214502388618906>. URL: <https://doi.org/10.1198/016214502388618906>.
- [12] Peter D. Hoff. *Modeling homophily and stochastic equivalence in symmetric relational data*. 2007. arXiv: 0711.1146 [stat.ME].
- [13] Matthew D Hoffman, Andrew Gelman, et al. “The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.” In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623.
- [14] Brian Karrer and M. E. J. Newman. “Stochastic blockmodels and community structure in networks”. In: *Physical Review E* 83.1 (Jan. 2011). DOI: 10.1103/physreve.83.016107. URL: <https://doi.org/10.1103/physreve.83.016107>.
- [15] Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. 2010.
- [16] Clement Lee and Darren Wilkinson. “A review of stochastic block models and extensions for graph clustering”. In: *Applied Network Science* 4 (Dec. 2019). DOI: 10.1007/s41109-019-0232-2.
- [17] Sirio Legramanti et al. “Extended stochastic block models with application to criminal networks”. In: *The Annals of Applied Statistics* 16.4 (2022), pp. 2369–2395. DOI: 10.1214/21-A0AS1595. URL: <https://doi.org/10.1214/21-A0AS1595>.
- [18] Catherine Matias and Vincent Miele. *Statistical Clustering of Temporal Networks Through a Dynamic Stochastic Block Model*. 2016. arXiv: 1506.07464 [stat.ME].
- [19] Krzysztof Nowicki and Tom A. B Snijders. “Estimation and Prediction for Stochastic Blockstructures”. In: *Journal of the American Statistical Association* 96.455 (2001), pp. 1077–1087. DOI: 10.1198/016214501753208735. eprint: <https://doi.org/10.1198/016214501753208735>. URL: <https://doi.org/10.1198/016214501753208735>.
- [20] Georgia Papadogeorgou et al. “Covariate-Informed Latent Interaction Models: Addressing Geographic and Taxonomic Bias in Predicting Bird–Plant Interactions”. In: *Journal of the American Statistical Association* 0.0 (2023), pp. 1–12. DOI: 10.1080/01621459.2023.2208390. eprint: <https://doi.org/10.1080/01621459.2023.2208390>. URL: <https://doi.org/10.1080/01621459.2023.2208390>.
- [21] Evan Poworoznek, Federico Ferrari, and David Dunson. “Efficiently resolving rotational ambiguity in Bayesian matrix sampling with matching”. In: *arXiv preprint arXiv:2107.13783* (2021).
- [22] Christian P. Robert et al. *Accelerating MCMC Algorithms*. 2018. arXiv: 1804.02719 [stat.CO].
- [23] Tom AB Snijders. “Statistical models for social networks”. In: *Annual Review of Sociology* 37 (2011), pp. 131–153.
- [24] Sara Wade and Zoubin Ghahramani. “Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion)”. In: *Bayesian Analysis* 13.2 (2018), pp. 559–626. DOI: 10.1214/17-BA1073. URL: <https://doi.org/10.1214/17-BA1073>.