

GGMnonreg: Non-Regularized Gaussian Graphical Models in R

Donald R. Williams^{1, 2}

1 Department of Psychology, University of California, Davis 2 NWEA, Portland, USA

DOI: [10.21105/joss.03308](https://doi.org/10.21105/joss.03308)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Arfon Smith](#) ↗

Reviewers:

- [@AlexChristensen](#)
- [@GiulioCostantini](#)

Submitted: 03 April 2021

Published: 11 November 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Studying complex relations in multivariate datasets is a common task across the sciences. Cognitive neuroscientists model brain connectivity with the goal of unearthing functional and structural associations between cortical regions ([Ortiz et al., 2015](#)). In clinical psychology, researchers wish to better understand the intricate web of symptom interrelations that underlie mental health disorders ([Borsboom et al., 2011](#); [McNally, 2016](#)). To this end, graphical modeling has emerged as an oft-used tool in the chest of scientific inquiry. The basic idea is to characterize multivariate relations by learning the conditional dependence structure. The cortical regions or symptoms are *nodes* and the featured connections linking nodes are *edges* that graphically represent the conditional dependence structure.

Graphical modeling is quite common in fields with wide data, that is, when there are more variables (p) than observations (n). Accordingly, many regularization-based approaches have been developed for those kinds of data. There are key drawbacks of regularization, including, but not limited to, the fact that obtaining a valid measure of parameter uncertainty is very (very) difficult ([Bühlmann et al., 2014](#)) and there can be an inflated false positive rate (see for example, [Donald R. Williams et al., 2019](#)).

Statement of Need

More recently, graphical modeling has emerged in psychology (Epskamp et al. 2018), where the data is typically long or low-dimensional ($p < n$; [Donald R. Williams et al. \(2019\)](#), [Donald R. Williams & Rast \(2019\)](#)). The primary purpose of **GGMnonreg** is to provide methods that were specifically designed for low-dimensional data (e.g., those common in the social-behavioral sciences).

Supported Models

- Gaussian graphical model (GGM). The following data types are supported.
 - Gaussian
 - Ordinal
 - Binary
- Ising model ([Marsman et al., 2017](#))
- Mixed graphical model

Additional methods

The following are also included

- Expected network replicability ([Donald R. Williams, 2020](#))
- Compare Gaussian graphical models
- Measure of parameter uncertainty ([Donald R. Williams, 2021](#))
- Edge inclusion “probabilities” (e.g., Figure 6.4 in [Hastie et al., 2015](#))
- Network visualization
- Constrained precision matrix (the network, given an assumed graph, see p. 631 in [Hastie et al., 2009](#))
- Predictability (variance explained for each node, [Haslbeck & Waldorp, 2018](#))

Gaussian graphical Model

The following estimates a GGM for 5 post-traumatic stress disorder (PTSD) symptoms ([Ar-mour et al., 2017](#)):

```
fit <- ggm_inference(Y = ptsd[,1:5],
                    boot = FALSE)

fit
#>           1           2           3           4           5
#> 1 0.0000000 0.2262934 0.0000000 0.3335737 0.1547986
#> 2 0.2262934 0.0000000 0.4993419 0.0000000 0.0000000
#> 3 0.0000000 0.4993419 0.0000000 0.2205442 0.1841798
#> 4 0.3335737 0.0000000 0.2205442 0.0000000 0.3407634
#> 5 0.1547986 0.0000000 0.1841798 0.3407634 0.0000000
```

Predictability

It is common to then estimate “predictability,” which corresponds to R^2 for each node in the network. In **GGMnonreg**, this is implemented with the following code:

```
predictability(fit)

#>   Estimate Est.Error Ci.lb Ci.ub
#> 1    0.45     0.05  0.35  0.54
#> 2    0.50     0.05  0.41  0.59
#> 3    0.55     0.04  0.47  0.64
#> 4    0.50     0.05  0.41  0.59
#> 5    0.46     0.05  0.37  0.55
```

Ising Model

An Ising model is for binary data. The PTSD symptoms can be binary, indicating the symptom was either present or absent. This network is estimated with:

```
# make binary
Y <- ifelse(ptsd[,1:5] == 0, 0, 1)
```

```
# fit model
fit <- ising_search(Y, IC = "BIC",
                  progress = FALSE)

fit
#>           1           2           3           4           5
#> 1 0.000000 1.439583 0.000000 1.273379 0.000000
#> 2 1.439583 0.000000 1.616511 0.000000 1.182281
#> 3 0.000000 1.616511 0.000000 1.716747 1.077322
#> 4 1.273379 0.000000 1.716747 0.000000 1.662550
#> 5 0.000000 1.182281 1.077322 1.662550 0.000000
```

Network Replicability

Recently, the topic of replicability has captivated the network literature. To this end, I developed an analytic solution to estimate network replicability (Donald R. Williams, 2020).

The first step is to define a “true” partial correlation network. As an example, I generate a synthetic partial correlation matrix, and then compute expected network replicability.

```
# edges between 0.05 and 0.25
main <- gen_net(p = 20,
               lb = 0.05,
               ub = 0.25)

# enr
enr(main$pcors,
     n = 500,
     replications = 4)

#> Average Replicability: 0.53
#> Average Number of Edges: 30 (SD = 2.12)
#>
#> ----
#>
#> Cumulative Probability:
#>
#>   prop.edges  edges  Pr(R > prop.edges)
#>         0.0     0         1.00
#>         0.1     6         1.00
#>         0.2    11         1.00
#>         0.3    17         1.00
#>         0.4    23         1.00
#>         0.5    28         0.78
#>         0.6    34         0.02
#>         0.7    40         0.00
#>         0.8    46         0.00
#>         0.9    51         0.00
#>
#> ----
#> Pr(R > prop.edges):
#> probability of replicating more than the
#> corresponding proportion (and number) of edges
```

On average, we can expect to replicate roughly half of the edges in four replication attempts,

where replication is defined as detecting a given edge in each attempt. Further, the probability of replicating more than 70% of the edges is zero.

Network Visualization

A key aspect of graphical modeling is visualizing the conditional dependence structure. To this end, **GGMnonreg** makes network plots with **ggplot2** (Wickham, 2016).

```
plot(fit,  
     node_names = colnames(Y),  
     edge_magnify = 2)
```

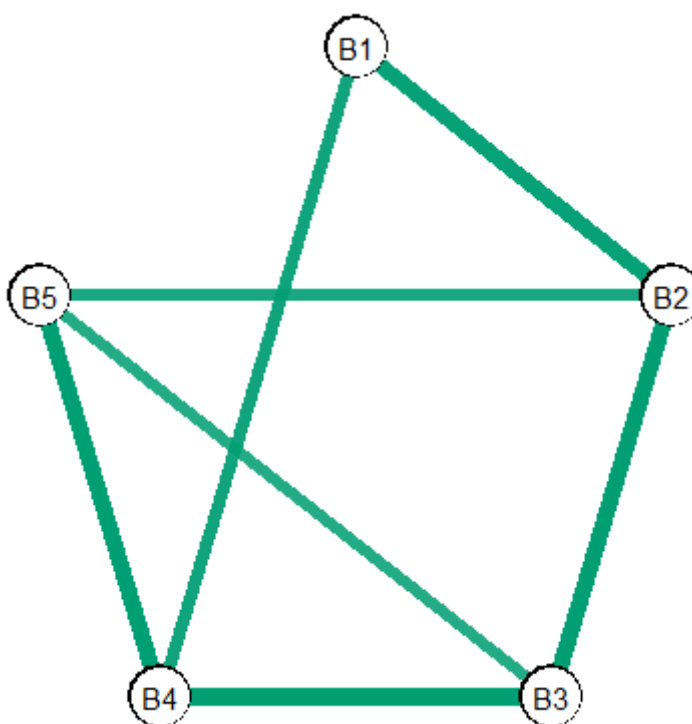


Figure 1: Conditional Dependence Structure

Acknowledgements

DRW was supported by a National Science Foundation Graduate Research Fellowship under Grant No. 1650042

References

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., & Pietrzak, R. H. (2017). A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in US military veterans. *Journal of Anxiety Disorders*, 45, 49–59. <https://doi.org/10.31234/osf.io/p69m7>

- Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The Small World of Psychopathology. *PLoS ONE*, 6(11), e27407. <https://doi.org/10.1371/journal.pone.0027407>
- Bühlmann, P., Kalisch, M., & Meier, L. (2014). High-Dimensional Statistics with a View Toward Applications in Biology. *Annual Review of Statistics and Its Application*, 1(1), 255–278. <https://doi.org/10.1146/annurev-statistics-022513-115545>
- Haslbeck, J. M., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods*, 50(2), 853–861. <https://doi.org/10.3758/s13428-017-0910-x>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations* (p. 362). CRC Press. <https://doi.org/10.1201/b18401>
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., Bork, R. van, Waldorp, L. J., Taylor, B., Waldorp, L., J van der Maas, H. L., & Maris, G. (2017). An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models. *Taylor & Francis*, 53(1), 15–35. <https://doi.org/10.1080/00273171.2017.1379379>
- McNally, R. J. (2016). *Can network analysis transform psychopathology?* (Vol. 86, pp. 95–104). Elsevier Ltd. <https://doi.org/10.1016/j.brat.2016.06.006>
- Ortiz, A., Munilla, J., Álvarez-Illán, I., Górriz, J. M., & Ramírez, J. (2015). Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's disease diagnosis. *Frontiers in Computational Neuroscience*, 9(November), 1–18. <https://doi.org/10.3389/fncom.2015.00132>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4
- Williams, Donald R. (2020). Learning to live with sampling variability: Expected replicability in partial correlation networks. *PsyArXiv*. <https://doi.org/10.31234/osf.io/fb4sa>
- Williams, Donald R. (2021). The confidence interval that wasn't: Bootstrapped “confidence intervals” in L1-regularized partial correlation networks. *PsyArXiv*. <https://doi.org/10.31234/osf.io/kjh2f>
- Williams, Donald R., & Rast, P. (2019). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*. <https://doi.org/10.1111/bmsp.12173>
- Williams, Donald R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On nonregularized estimation of psychological networks. *Multivariate Behavioral Research*, 54(5), 719–750. <https://doi.org/10.1080/00273171.2019.1575716>