Neuman *et al. Applied Network Science* (2024) 9:6
https://doi.org/10.1007/s41109-024-00612-8

Applied Network Science

RESEARCH Open Access

Module-based regularization improves Gaussian graphical models when observing noisy data

Magnus Neuman[1*], Joaquín Calatayud[2], Viktor Tasselius[1,3] and Martin Rosvall[1]

*Correspondence: magnus.neuman@gmail.com

[1] Integrated Science Lab, Department of Physics, Umeå University, Umeå, Sweden
[2] Department of Biology, Geology, Physics and Inorganic Chemistry, King Juan Carlos University, Madrid, Spain
[3] School of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden

Abstract

Inferring relations from correlational data allows researchers across the sciences to uncover complex connections between variables for insights into the underlying mechanisms. The researchers often represent inferred relations using Gaussian graphical models, requiring regularization to sparsify the models. Acknowledging that the modular structure of these inferred networks is often studied, we suggest module-based regularization to balance under- and overfitting. Compared with the graphical lasso, a standard approach using the Gaussian log-likelihood for estimating the regularization strength, this approach better recovers and infers modular structure in noisy synthetic and real data. The module-based regularization technique improves the usefulness of Gaussian graphical models in the many applications where they are employed.

Keywords: Gaussian graphical models, Correlational data, Regularization, Model selection, Modular structure, Network communities

Introduction

Inferring relations between observed features from correlational data is a foundational approach to exploring underlying mechanisms in, for example, ecological, genetic and neural systems (Barberán et al. 2012; Wang and Huang 2014; Bullmore and Sporns 2009). The resulting relations are often represented as a network where the features are nodes and their respective relations are links. These networks are dense, making it difficult to discern relevant structures. Field-specific methods to sparsify them suggest soft thresholding (Zhang and Horvath 2005), but hard thresholding is often applied in practice (Barberán et al. 2012; de Vries et al. 2018; Neuman et al. 2022). Gaussian graphical models provide an alternative way of representing correlational data by encoding relations between features through partial correlations. A popular approach to infer a Gaussian graphical model is the graphical lasso (GLASSO) (Friedman et al. 2007; Yuan and Lin 2007), which estimates the precision matrix while ensuring sparsity through l_1-regularization. This method and related methods, such as neighborhood selection (Meinshausen and Bühlmann 2006), elastic net (Zou and Hastie 2005) and Markov networks (Murphy 2012), are widely used in many

disciplines (Harris 2016; Epskamp et al. 2018; Cao et al. 2017; Severson et al. 2019). Despite its widespread application, GLASSO struggles to tackle noise and the high dimensionality that comes with many observed features, often exceeding the number of available samples (Ravikumar et al. 2008; Wainwright 2009; Ravikumar et al. 2011; Liu et al. 2012).

Representing the inferred relations as networks enables studying structures in the data with standard tools from network science. Network modules – groups of tightly connected nodes – are studied across scientific disciplines because they reveal significant patterns and functional relationships in diverse systems, ranging from ecological (Calatayud et al. 2020, 2021) to metabolic networks (Guimera and Nunes Amaral 2005). However, the GLASSO is agnostic to modular structure in the inferred networks, which can obscure network structure and subsequent interpretation and understanding of the studied systems. Simultaneously inferring the network and its modular structure can alleviate this problem, but requires prior knowledge about the dynamical processes on the network (Peixoto 2019), which is rarely applicable or available. Attempts at integrating modular structure with the GLASSO use a predetermined number of modules (Ambroise et al. 2009; Ver Steeg et al. 2019) or other criteria than modular structure when regularizing (Tan et al. 2015; Pircalabelu and Claeskens 2020; Kumar et al. 2020). Manually setting the number of modules or regularizing using criteria that do not account for modular structure risks over- or underfitting the modular structure to data.

Here we propose a novel approach to solve this model selection problem by integrating the two steps from relational data to network modules – network inference and community detection – in an extension of the GLASSO method. We use the network's modular structure to select the regularization strength, which allows us to balance over- and underfitting the modular structure to the data. Using synthetic data, we show that this approach allows us to recover more modular structures in noisy data compared with the standard GLASSO. Applied to country-level daily incidence during the Covid-19 pandemic and gene co-expression data from the plant *Arabidopsis thaliana*, we find that the module-based GLASSO can identify more modular structure in these data compared to the standard GLASSO – highly relevant for researchers studying these systems.

Results

Gaussian graphical models describe relations between observed features. They are derived from the precision matrix Θ that encodes conditional independence between variables, meaning that two observations X_i and X_j are independent, given all other observations, if the corresponding *ij*:th element in Θ is zero. The GLASSO aims at maximizing the Gaussian log-likelihood of the precision matrix given the data while ensuring a sparse solution by imposing an l_1-regularization term $\lambda||\Theta||_1$, with the regularization parameter λ. The best precision matrix Θ^λ for a specific value of λ is thus

$$\Theta^\lambda = \underset{\Theta}{\operatorname{argmax}} \left(\log\det(\Theta) - \operatorname{tr}(\Theta\hat{\Sigma}) - \lambda||\Theta||_1 \right), \tag{1}$$

where $\hat{\Sigma}$ is the covariance matrix calculated from the observed data. The parameter λ determines the regularization strength and thereby the sparsity of the inferred precision matrix. The regularization parameter λ is often determined using cross-validation where

the best value λ_{ll}^* is the one that has the largest log-likelihood of the test data $\hat{\Sigma}^{test}$ given the model $\Theta^{\lambda,train}$ inferred from the training data $\hat{\Sigma}^{train}$ such that

$$\lambda_{ll}^* = \underset{\lambda}{\operatorname{argmax}} \left(\log\det(\Theta^{\lambda,train}) - \operatorname{tr}(\Theta^{\lambda,train}\hat{\Sigma}^{test}) \right). \tag{2}$$

The resulting regularization strength conserves relations with support in both the training and test data, without considering any conserved structures in the data.

To take the modular structure into account when selecting the regularization strength, we suggest using the map equation framework and its search algorithm Infomap (Rosvall and Bergstrom 2008; Rosvall et al. 2009; Edler et al. 2017). The map equation encodes a random walk on a network and measures the codelength $L(M)$ of the random walk given a partition M of the network into modules. Infomap uses a greedy approach to find the partition M^* that minimizes the codelength,

$$M^* = \underset{M}{\operatorname{argmin}}\, L(M), \tag{3}$$

such that M^* is the best partition of the network according to the minimum description length principle. This popular approach is widely recognized as one of the best methods for detecting network communities (Lancichinetti and Fortunato 2009; Aldecoa and Marín 2013). To connect Infomap with the GLASSO regularization, we suggest maximizing the signal of modular structure present in both the training and test sets when cross-validating the regularization parameter λ. We measure this signal using the codelength savings in the test data given the optimal partition of the training data, such that

$$\lambda_{cl}^* = \underset{\lambda}{\operatorname{argmax}} \frac{L^{test}(1) - L^{test}(M^{\lambda,train})}{L^{test}(1)}, \tag{4}$$

where $M^{\lambda,train}$ is the optimal partition of the training data and $L^{test}(1)$ is the one-level codelength of the test data with all nodes in the same module. The codelength savings are positive if the modular structure in the training data is present also in the test data and has its maximum when this shared modular structure is most prominent. This peak is associated with the λ that best captures modular structure in the data without over- or underfitting, analogous to the log-likelihood in Eq. 2. In previous work (Neuman et al. 2022), we explored this module-based approach for hard thresholding of correlation networks and showed that a too low threshold gives a highly connected network with little modular structure in both the training and test networks, while a too high threshold gives a highly modular structure in the training network that is not present in the test network. The same reasoning applies to the GLASSO when selecting regularization strength. The approach we suggest finds the best compromise between these two extremes.

To derive the network $\mathcal{G}(\Theta^*)$ that balances under- and overfitting, we use GLASSO to estimate the precision matrix Θ^* corresponding to λ^*. We use the relation between a partition matrix element and the partial correlation so that the link e_{ij} between nodes i and j is given by

$$e_{ij} = |-\theta_{ij}/\sqrt{\theta_{ii}\theta_{jj}}|, \tag{5}$$

where θ_{ij} is elements of Θ^*, and the link weight is thus the absolute value of the partial correlation.

Synthetic data

To test the module-based regularization, we generate synthetic data by sampling a covariance matrix S from a Wishart distribution such that

$$S \sim W_p(n, \Sigma), \tag{6}$$

where Σ is the block-diagonal covariance matrix of the planted (oracle) modular structure, p is the dimension (number of features or nodes) and n is the number of degrees of freedom. We plant a modular structure by imposing a block-diagonal structure:

$$\Sigma_{i,j} = \begin{cases} 1, \ i = j \\ c, \ M(p_i) = M(p_j) \\ 0, \ M(p_i) \neq M(p_j), \end{cases} \tag{7}$$

where $M(p_i)$ denotes the module of node p_i. In this way, both the planted covariance matrix Σ and the sampled matrix S are positive definite. To change the signal-to-noise ratio, we can vary the the planted within-module covariance c and the degrees of freedom n in the Wishart distribution. Using this setup, we sample the observed data X from a p-variate normal distribution such that $X_i \sim N_p(0, S)$ and $X \in \mathbb{R}^{p \times q}$, where q denotes the number of samples. The objective is to infer the planted modular structure using these data.

We use ten planted modules with ten nodes in each module to illustrate our approach, as shown in Fig. 1. The sampled covariance matrix is shown in Fig. 1a, where the number of degrees of freedom is $n = 100$ and the planted covariance is $c = 0.4$, and we see that the matrix is noisy but with discernible modular structure. Using this covariance matrix we draw $q = 100$ samples to obtain the synthetic data. We see that the log-likelihood-based GLASSO (hereafter Standard GLASSO) gives a lower optimal λ value and hence regularizes less than the module-based GLASSO (hereafter Modular GLASSO), since their respective quality functions peak at different λ-values (Fig. 1b). This leads to the Standard GLASSO including a lot of noisy links, as the network representation shows (Fig. 1c). In contrast, Modular GLASSO increases the regularization to maximize the modular structure common to the test and training data, enabling the method to correctly recover the planted modular structure (Fig. 1d).

To explore this result, we vary the covariance c and the number of samples q. We quantify how well the methods recover the planted partition by calculating the adjusted mutual information (AMI) between the planted partition and the recovered partition, which is the partition found by Infomap given the network $\mathcal{G}(\Theta^*)$ (Fig. 2). The Standard GLASSO recovers the planted partition when the number of samples is small and the covariance is large, but not for many samples (Fig. 2a). This tendency to recover more modular structure with fewer samples exemplifies the "blessing of dimensionality" (Ver Steeg et al. 2019). In contrast, the Modular GLASSO

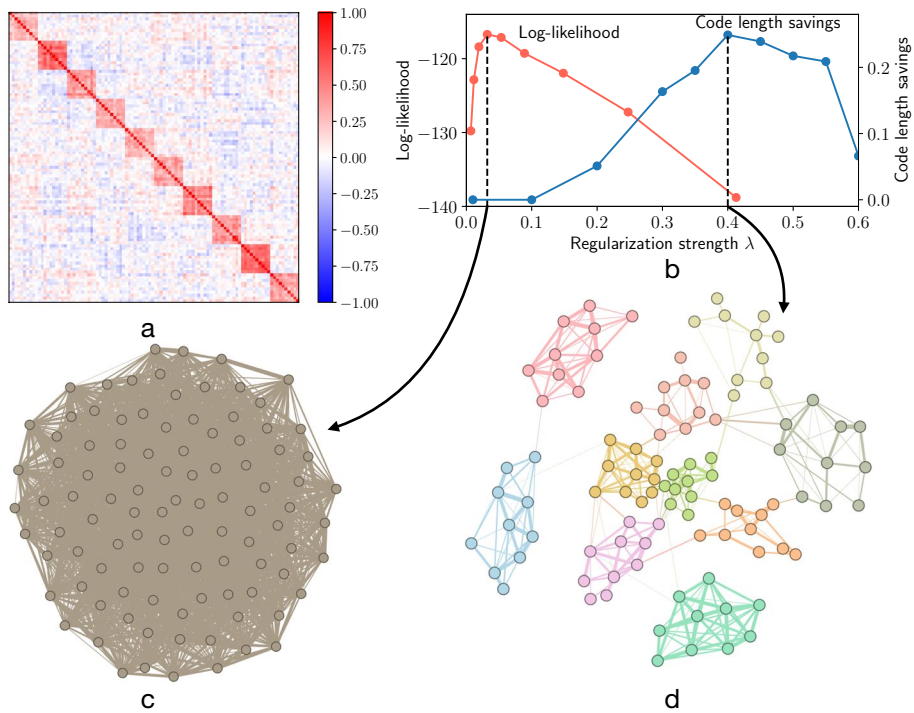

Fig. 1 Comparing Standard and Modular GLASSO methods in detecting planted modular structure. The covariance matrix sampled from the Wishart distribution is noisy but with modular structure (**a**). With data sampled using this matrix, the GLASSO based on log-likelihood (Standard GLASSO) regularizes less than the GLASSO based on modular structure through Infomap's codelength (Modular GLASSO) (**b**), which leads to the Standard GLASSO's failure to identify any modular structure (**c**) while the Modular GLASSO successfully recovers the planted modular structure (**d**)

recovers the modular structure when the number of samples and the covariance are large–increasing the number of samples is always beneficial until all modular structure is recovered (Fig. 2b).

When we decrease the noise level by using $n = 1000$ degrees of freedom in the Wishart distribution, the methods show similar performance, with a slight advantage for the Standard GLASSO, and recover the modular structure for sufficiently large covariance and number of samples (Fig. 2cd). This result indicates that Standard GLASSO's performance is sensitive to the presence of noise in the data.

To compare the methods more closely, we plot the optimal λ-value as a function of the number of samples for a fixed value $c = 0.6$ of the within-module covariance (Fig. 3). The Standard GLASSO's optimal λ decreases for more samples, while it increases for the Modular GLASSO. The AMI approaches zero for the Standard GLASSO for more samples because it regularizes less. The Modular GLASSO's regularization increases with the number of samples. The AMI reaches 1 since the method captures the signal of the modular structure and adapts the regularization. In contrast, the Standard GLASSO does not regularize at all when there are many samples, retaining all spurious relations between the observed features and obscuring the modular structure.

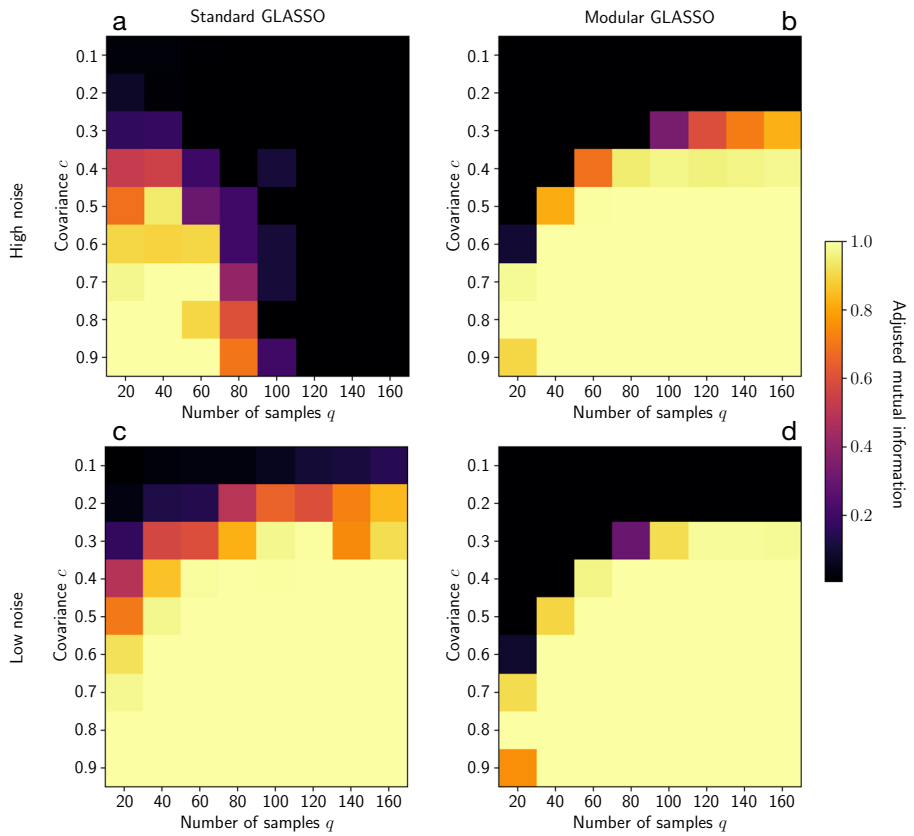

Fig. 2 Performance comparison of the Standard and the Modular GLASSO in detecting planted partitions under low and high noise conditions. The adjusted mutual information (AMI) between recovered and planted partitions shows that the Standard GLASSO finds the planted partition only if the samples are few when the noise level is high, but when samples and within-module covariance are sufficient for low noise. In contrast, the Modular GLASSO finds the planted partition also in high noise when samples and covariance are sufficient

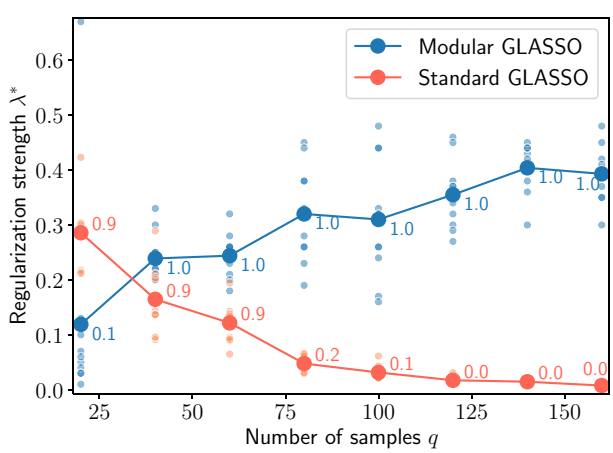

Fig. 3 The optimal regularization strength as a function of the number of samples for Standard and Modular GLASSO. The Standard GLASSO regularizes less, resulting in the inclusion of many noisy correlations. The Modular GLASSO regularizes based on the modular structure, leading to a stronger regularization as the number of samples increases and the recovery of the planted modular structure. The large points represent averages over ten runs, with individual runs shown as small points. The AMI between recovered and planted partitions is displayed as a number next to each point

Real-world data

Covid-19 data. We analyze the global Covid-19 data ("Our World in Data", 2022) with the daily incidence of Covid-19 in 192 countries over 777 days, from 2020/01/01 to 2022/02/15. The observed features are the world's countries and the samples are the 777 days with Covid-19 incidence, making it a sample-rich data set. The signal-to-noise ratio is high because the distribution of correlations significantly deviates from what would be expected from spurious correlations (the Kolmogorov-Smirnov statistic is 0.72). For these data, Standard and Modular GLASSO suggest vastly different λ-values (Fig. 4a). For the Standard GLASSO, $\lambda^* \sim 0.001$ and $\lambda \lesssim 0.1$ results in only one module in the corresponding network, providing no information about modular structure in the Covid-19 data. Excluding the edge-case peak for a disintegrated network with many singletons, $\lambda^* \sim 0.36$ for the Modular GLASSO resulting in 14 modules (Fig. 4b). The modules spread on the world map exhibit a geographic signal, with neighboring countries often belonging to the same module, as in Eastern Europe and parts of Central America, for example. China, however, forms its own module. In some cases, the connection between countries within the same module is less obvious, such as between the United States and the Iberian Peninsula, leaving it unclear whether a causal connection exists.

While the Standard GLASSO provides no information about modular structure in the global Covid-19 data, the Modular GLASSO unveils intriguing modular patterns. This situation resembles cases with high noise levels and many samples in the analysis of synthetic data when the Standard GLASSO retains many spurious relations, resulting in a dense, module-free network.

Gene co-expression data. We analyze gene co-expression data obtained from the plant *Arabidopsis thaliana* under cold stress with included control samples (see Methods for details). We select the 1,000 genes with the highest variance across the 209 samples. Similar to the Covid-19 data, the correlations deviate significantly from what would be expected from pure noise (the Kolmogorov-Smirnov statistic is 0.39). In this case, however, the number of features exceeds the number of samples.

Cross-validating using the codelength savings to maximize the modular structure common to training and test data regularizes more. The Standard GLASSO applies minimal regularization ($\lambda^* \sim 0.002$) and finds seven modules in the data (Fig. 5a). In contrast, the Modular GLASSO suggests strong regularization ($\lambda^* \sim 0.76$) and disconnects

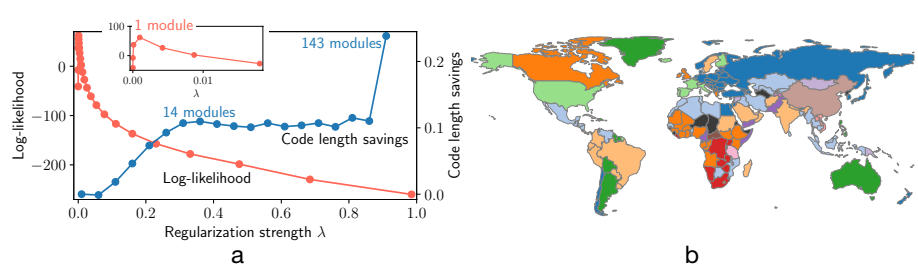

Fig. 4 Application of Standard and Modular GLASSO to Covid-19 incidence data. The Standard GLASSO (log-likelihood) and Modular GLASSO (codelength savings) suggest vastly different regularization strengths for the Covid-19 data (**a**). The Standard GLASSO reveals no modular structure in the resulting network, while the Modular GLASSO uncovers the 14 modules represented by different colors on the world map (**b**). The modules exhibit a geographical signal as adjacent countries tend to belong to the same module, with some interesting exceptions

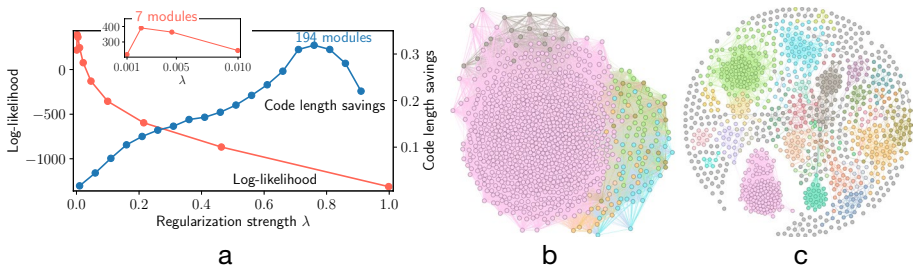

Fig. 5 Application of Standard and Modular GLASSO to gene co-expression data. The Standard GLASSO (log-likelihood) and Modular GLASSO (codelength savings) suggest vastly different regularization strengths also for the gene co-expression data (**a**). The Standard GLASSO's minimal regularization leads to a network with little modular structure (**b**). In contrast, Modular GLASSO disconnects nodes to maximize the modular structure during cross-validation, revealing more regularities in the underlying system (**c**)

nodes, resulting in distinct network representations of the data (Fig. 5b, c). The stronger regularization can reveal additional structure in the underlying data, offering valuable insights into the gene regulation patterns.

Discussion

Regularizing Gaussian graphical models is challenging due to the presence of noise and the complexity of high-dimensional data. To tackle this issue, we introduce a regularization method that capitalizes on the modular structure inherent in the data. The Modular GLASSO outperforms standard regularization approaches by applying stronger regularization to retain only the connections that contribute significantly to the modular structure. In contrast, the Standard GLASSO, which maximizes the Gaussian log-likelihood, regularizes less when dealing with noisy data, retaining many noisy links and failing to detect modular structure. The differences between these two methods are crucial when analyzing real-world data sets. For example, when analyzing Covid-19 incidence data and gene co-expression data, the Modular GLASSO uncovered more modular structure in the data, providing deeper insights about the underlying system such as identifying groups of countries with similar epidemic patterns and gene clusters with similar functions.

Constructing a network from correlational data requires many samples. When cross-validating modules, both the training and test networks must contain modular structure present in the complete data set. Two-fold splitting offers a reliable approach but requires relatively many samples, leading to suboptimal results with Modular GLASSO for low-noise data (Fig. 2) and relatively large spread (Fig. 3). A potential solution to this data-splitting issue is to eliminate the need for splitting altogether by employing Bayesian methods as an alternative to cross-validation. A Bayesian approach would make Modular GLASSO less data demanding.

Using codelength savings for model selection may result in selecting an overly sparse model when some modules have much larger link weights than others. In such cases, the codelength savings in the test network can be larger if the modules with smaller link weights are completely disintegrated into disconnected nodes. Partly washing out modular structure by excessive regularization in this way can, however, reveal potentially interesting structure through the remaining modules that can be difficult to discern with

less regularization, as for the gene co-expression data (Fig. 5c). Since the disconnected nodes are weakly connected in the unregularized network, disconnecting them is supported in the data and in line with the module-based regularization.

In summary, we find that regularization based on modules effectively uncovers more structure in relational data sets. Because many downstream analysis tasks rely on identifying modular structure, including studying groups, communities, and clusters of observed features, many researchers may find it appealing and intuitive to base also their model selection criterion on modular structure. As we show, this approach is essential for detecting underlying modular structure in correlational data that would otherwise remain obscured by noise.

Methods

Optimization and randomness in the map equation

The relative codelength savings in the test network, $L^{test}(M^{\lambda,train})$, which depend on the regularization parameter λ, are stochastic for two reasons: the two-fold splitting in the cross-validation and the inherent randomness in the search algorithm Infomap optimizing the non-convex map equation objective function (Calatayud et al. 2019). To overcome this stochasticity, we perform two-fold splitting ten times and average the results. To calculate the AMI in Fig. 2, we also perform a sample average approximation with ten runs. For simple partition comparisons, we only look for two-level solutions with Infomap. When selecting the optimal λ to calculate the AMI, we test a set of $\lambda \in \{0.01, 0.06, \ldots, 0.96\}$, and choose the λ corresponding to the first maximum in the codelength savings. To avoid noise around zero at low λ values, we use the additional condition that the codelength savings must exceed 0.01.

Analysis and real-world data

We use the R packages `GLASSO` and `CVGLASSO` throughout the tests. To find the maximum log-likelihood, we increase or decrease the function parameters `nlam` and `lam.min.ratio` if necessary.

Gene co-expression data come from the Sequence Read Archive (SRA), where we identified all available RNA-Seq samples relating to cold stress in the leaf tissue of *Arabidopsis thaliana* ecotype Columbia-0. The selected data include both control and treated samples and were retrieved in April 2021. A full list of included samples can be found in the supplementary material. The data were quantified using salmon version 1.2.1 (Patro et al. 2017) against the Araport 11 release of the *Arabidopsis thaliana* genome. Pre-processing and normalization were done in R using the variance stabilizing transform available in DESeq2 (Love et al. 2014).

To avoid constant values when analyzing the world Covid-19 data, we leave out countries belonging to the 10:th percentile with the lowest variance in daily incidence.

All data are standardized before analysis, and have zero mean and unit standard deviation.

Modular GLASSO algorithm

Algorithm 1 shows the pseudo code for Modular GLASSO. The code uses the R function `GLASSO` to estimate the precision matrix Θ with a given value of the the regularization

parameter λ. The code uses the Infomap algorithm to infer the modules M in a network $\mathcal{G}(\Theta)$ and to calculate the codelength savings l when a network is partitioned into a given modular structure. We have prepared a Jupyter notebook with code and examples ("Modular GLASSO", 2024).

Algorithm 1 Modular GLASSO

Input: $X \in \mathbb{R}^{p \times q}$, p - number of features (nodes), q - number of samples
Output: $\mathcal{G}(\Theta^*)$, best model
$\Lambda \leftarrow \{0.01, 0.06, \ldots, 0.96\}$ // set of λ values to test
$\lambda^* \leftarrow \min(\Lambda)$ // optimal λ
$l^* \leftarrow 0$ // optimal codelength savings
for $\lambda \in \Lambda$ **do** // These steps are repeated 10 times and averaged
  $X^{train} \leftarrow \text{sample}(X, \text{fraction} = 0.5, \text{axis} = 1)$
  $X^{test} \leftarrow X \setminus X^{train}$
  $\mathcal{G}(\Theta^{train}) \leftarrow \text{GLASSO}(X^{train}, \lambda)$
  $M^{train} \leftarrow \text{Infomap}(\mathcal{G}(\Theta^{train}))$
  $\mathcal{G}(\Theta^{test}) \leftarrow \text{GLASSO}(X^{test}, \lambda)$
  $l \leftarrow \text{Infomap}(\mathcal{G}(\Theta^{test}), M^{train})$
  if $l > l^*$ **then**
    $l^* \leftarrow l$
    $\lambda^* \leftarrow \lambda$
  end if
end for
$\mathcal{G}(\Theta^*) \leftarrow \text{GLASSO}(X, \lambda^*)$

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-024-00612-8.

Additional file 1. Sample list.

Acknowledgements
Not applicable

Author contributions
MN an JC conceived the study. MN designed the study, performed the experiments and wrote the first draft. VT assembled the gene expression data. All authors discussed the results and edited the manuscript.

Funding
Open access funding provided by Umea University. MN, VT and MR were supported by the Swedish Foundation for Strategic Research, grant no. SB16-0089. JC was supported by the Spanish Ministry of Science and Innovation through the UNIPER project (PID2020-114851GA-I00). MR was supported by the Swedish Research Council, grant no. 2016-00796.

Availability of data and materials
The datasets analysed during the current study are available in the Sequence Read Archive (SRA), https://www.ncbi.nlm.nih.gov/sra, and from Data on COVID-19 (coronavirus) by Our World in Data, https://github.com/owid/covid-19-data/tree/master/public/data, respectively. A list of samples extracted from SRA can be found in the supplementary material. The datasets assembled from these sources are available from the corresponding author upon reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 5 December 2023 Accepted: 29 February 2024
Published online: 18 March 2024

References
Aldecoa R, Marín I (2013) Exploring the limits of community detection strategies in complex networks. Sci Rep 3:2216

Ambroise C, Chiquet J, Matias C (2009) Inferring sparse Gaussian graphical models with latent structure. Electron J Stat 3:205–238. https://doi.org/10.1214/08-EJS314

Barberán A, Bates ST, Casamayor EO, Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6(2):343–351. https://doi.org/10.1038/ismej.2011.119

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10(3):186–198. https://doi.org/10.1038/nrn2575

Calatayud J, Andivia E, Escudero A, Melián CJ, Bernardo-Madrid R, Stoffel M, Aponte C, Medina NG, Molina-Venegas R, Arnan X et al (2020) Positive associations among rare species and their persistence in ecological assemblages. Nat Ecol Evol 4(1):40–45

Calatayud J, Bernardo-Madrid R, Neuman M, Rojas A, Rosvall M (2019) Exploring the solution landscape enables more reliable network community detection. Phys Rev E 100:052308. https://doi.org/10.1103/PhysRevE.100.052308

Calatayud J, Neuman M, Rojas A, Eriksson A, Rosvall M (2021) Regularities in species' niches reveal the world's climate regions. eLife 10:58397. https://doi.org/10.7554/eLife.58397

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357(6352):661–667. https://doi.org/10.1126/science.aam8940. (www.science.org/doi/pdf/10.1126/science.aam8940)

de Vries FT, Griffiths RI, Bailey M, Craig H, Girlanda M, Gweon HS, Hallin S, Kaisermann A, Keith AM, Kretzschmar M, Lemanceau P, Lumini E, Mason KE, Oliver A, Ostle N, Prosser JI, Thion C, Thomson B, Bardgett RD (2018) Soil bacterial networks are less stable under drought than fungal networks. Nat Commun 9(1):3033. https://doi.org/10.1038/s41467-018-05516-7

Edler D, Bohlin L, Rosvall M (2017) Mapping higher-order network flows in memory and multilayer networks with infomap. Algorithms. https://doi.org/10.3390/a10040112

Epskamp S, Borsboom D, Fried EI (2018) Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods 50(1):195–212. https://doi.org/10.3758/s13428-017-0862-1

Friedman J, Hastie T, Tibshirani R (2007) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441. https://doi.org/10.1093/biostatistics/kxm045. (**academic.oup.com/biostatistics/article-pdf/9/3/432/17742149/kxm045.pdf**)

Guimera R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433(7028):895–900

Harris DJ (2016) Inferring species interactions from co-occurrence data with markov networks. Ecology 97(12):3308–3314. https://doi.org/10.1002/ecy.1605. (**esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.1605**)

Kumar S, Ying J, de Cardoso MJV, Palomar DP (2020) A unified framework for structured graph learning via spectral constraints. J Mach Learn Res 21:1–60

Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. Phys Rev E 80:056117. https://doi.org/10.1103/PhysRevE.80.056117

Liu H, Han F, Yuan M, Lafferty J, Wasserman L (2012) High-dimensional semiparametric Gaussian copula graphical models. Ann Stat 40(4):2293–2326

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol 15(12):1–21

Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the Lasso. Ann Stat 34(3):1436–1462. https://doi.org/10.1214/009053606000000281

Modular GLASSO (2024) https://github.com/magnusneuman/Modular-GLASSO/

Murphy KP (2012) Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge, Massachusetts, USA

Neuman M, Jonsson V, Calatayud J, Rosvall M (2022) Cross-validation of correlation networks using modular structure. Appl Netw Sci 7(1):75. https://doi.org/10.1007/s41109-022-00516-5

Our World in Data: Data on COVID-19 (coronavirus) (2022). https://github.com/owid/covid-19-data/tree/master/public/data Accessed 16 Feb 2022

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14(4):417–419

Peixoto TP (2019) Network reconstruction and community detection from dynamics. Phys Rev Lett 123:128301. https://doi.org/10.1103/PhysRevLett.123.128301

Pircalabelu E, Claeskens G (2020) Community-based group graphical lasso. J Mach Learn Res 21:1–32

Ravikumar P, Wainwright MJ, Raskutti G, Yu B (2011) High-dimensional covariance estimation by minimizing l1-penalized log-determinant divergence. Electron J Stat 5:935–980

Ravikumar P, Raskutti G, Yu B, Wainwright MJ (2008) Model selection in gaussian graphical models: High-dimensional consistency of $\ell_{\uparrow}$-regularized mle. In: Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (eds.) Advances in Neural Information Processing Systems, vol. 21. Curran Associates, Inc., (2008). https://proceedings.neurips.cc/paper/2008/file/61f2585b0ebcf1f532c4d1ec9a7d51aa-Paper.pdf

Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. Eur Phys J Spec Topics 178(1):13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci 105(4):1118–1123. https://doi.org/10.1073/pnas.0706851105. (www.pnas.org/content/105/4/1118.full.pdf)

Severson KA, Attia PM, Jin N, Perkins N, Jiang B, Yang Z, Chen MH, Aykol M, Herring PK, Fraggedakis D, Bazant MZ, Harris SJ, Chueh WC, Braatz RD (2019) Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 4(5):383–391. https://doi.org/10.1038/s41560-019-0356-8

Tan KM, Witten D, Shojaie A (2015) The cluster graphical lasso for improved estimation of gaussian graphical models. Comput Stat Data Anal 85:23–36

Ver Steeg G, Harutyunyan H, Moyer D, Galstyan A (2019) Fast structure learning with modular regularization. In: Wallach, H., Larochelle, H., Beygelzimer, A., d' Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates, Inc., ??? . https://proceedings.neurips.cc/paper/2019/file/e2e14235335d2c0aa5f6855e339233d9-Paper.pdf

Wainwright MJ (2009) Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). IEEE Trans Inf Theory 55(5):2183–2202. https://doi.org/10.1109/TIT.2009.2016018
Wang YXR, Huang H (2014) Review on statistical methods for gene network reconstruction using expression data. J Theor Biol 362:53–61. https://doi.org/10.1016/j.jtbi.2014.03.040
Yuan M, Lin Y (2007) Model selection and estimation in the Gaussian graphical model. Biometrika 94(1):19–35. https://doi.org/10.1093/biomet/asm018. (**academic.oup.com/biomet/article-pdf/94/1/19/617853/asm018.pdf**)
Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:17. https://doi.org/10.2202/1544-6115.1128
Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B (Statistical Methodology) 67(2):301–320

Publisher's Note