Supplementary materials to: "Learning Graphical Factor Models with Riemannian Optimization"

Alexandre Hippert-Ferrer$^{1[0000-0002-7740-5415]}$, Florent Bouchard$^{2[0000-0003-3003-7317]}$, Ammar Mian$^{3,[0000-0003-1796-8707]}$, Titouan Vayer$^{4[0000-0002-8115-572X]}$, and Arnaud Breloy$^{5[0000-0002-3802-9015]}$

[1] Univ Gustave Eiffel, IGN, ENSG, LASTIG, Saint-Mandé, France
alexandre.hippert-ferrer@ign.fr
[2] Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, Gif-sur-Yvette, France
[3] LISTIC, Université Savoie Mont Blanc, Annecy, France
[4] Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, Lyon, France
[5] LEME, Université Paris-Nanterre, Ville d'Avray, France

1 Outline

Section 2 explains potential ethical issues regarding the collection of the ground truth of the *concepts* data set and underlines the limits of the graph predictions.

Section 3 presents validation experiments on synthetic data:

- Subsection 3.1 describes the experimental setup, as well as the graph and data generation parameters.
- Subsection 3.2 displays performance comparisons between the proposed algorithms and other state of the art graph learning methods.
- Subsection 3.3 presents a study of robustness regarding the selection of the parameters of the proposed methods (namely, the rank k in factor models, and the regularization parameter λ for the sparsity promoting penalty).

The code for reproducing these experiments is made available in the following repository:

https://github.com/ahippert/graphfactormodel

Section 4 presents an application of GGFM algorithm on the large scale "$n \ll p$" concepts dataset.

2 Ethical statement

2.1 Data set collection

A part of our numerical experiments are conducted on data sets representing related "entities" of the real world (*animals* and *concepts* data sets). We (scientists of the data/machine learning/signal processing communities) are conscious that both data sets are limited, non-exhaustive and partial. For example, the *concepts* data set has been labelled by humans using a survey conducted on the Amazon Mechanical Turk (AMT) platform. As recently documented in [3], these data sets might incorporate sociological and cultural biases, embedding a particular view of the world they intend to fit. Furthermore, AMT, amongst other micro-working platforms, employs precarious *crowdworkers* around the world at low pay rates [6] for annotating, labelling and correcting data that help to train and test AI models. This, somehow, maintains the illusion that AI systems are autonomous and intelligent by hiding real workers [1] and the labor-intensive process they require [9].

2.2 Limits of the graph prediction

Our proposed algorithms have been designed to learn graph connections between related entities. As discussed above, the *concepts* data set might incorporate various forms of biases because each semantic feature are questions which require a subjective answer. As no ground truth is available for the considered data sets, we advise users to carefully interpret the predicted graph structures. Indeed, spurious graph links might be learned, which would highlight the aforementioned biases. Any predicted graph is no immutable truth and should be questioned in a critical manner. We strongly encourage the development of strategies to mitigate spurious correlations, as reliance on expert advice (*i.e.*, a volcanologist for the GNSS data set) or the design of more open, transparent and representative training data.

3 Numerical experiments on synthetic graphs

In the following, we present extensive validation experiments on synthetic graphs to evaluate the performance of the proposed algorithms, *i.e.*, GGM, EGM, GGFM and EGFM.

3.1 Experimental settings

Graph and sample generation: For a fixed dimension p, we consider four standard random graphs models that are described in Table 1. Given a sampled support of the graph, weights of the edges then are sampled from $U(2, 5)$ (as in [10]), which yields a symmetric weighted adjacency matrix $\boldsymbol{A}$ for each random graph. Figure 1 shows examples of sampled graph structures and their adjacency

Graph model	Edge weights	Probability	Neighbors	Radius
Barabási-Albert	$U(2,5)$	–	–	–
Erdős–Rényi	$U(2,5)$	0.1	–	–
Watts-Strogatz	$U(2,5)$	0.1	5	–
Random geometric	$U(2,5)$	–	–	0.2

Table 1: Considered graph models in numerical experiments.

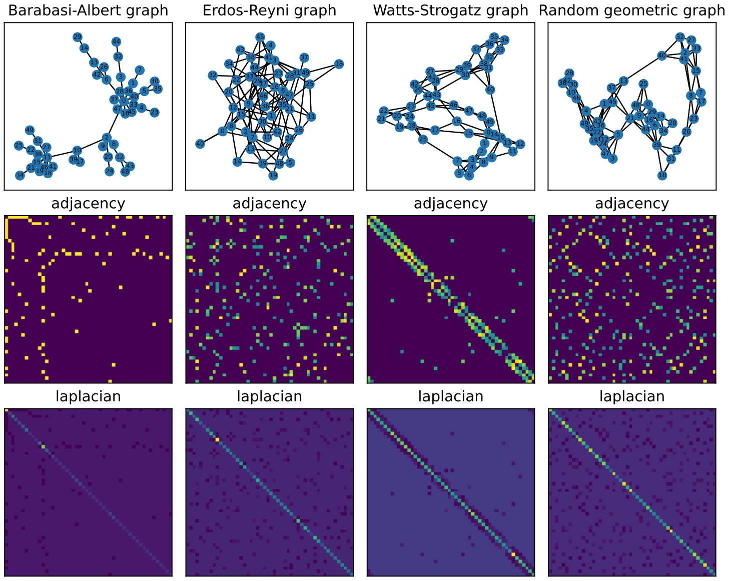


Fig. 1: Considered graph structures (here with $p = 50$ nodes) and associated adjacency and laplacian (precision) matrices.

matrices for each of the four models. The Laplacian matrix of the graph is then used to construct the precision matrix $\boldsymbol{\Theta}$, using the relation $\boldsymbol{\Theta} = \boldsymbol{L} + \kappa \boldsymbol{I}$, where $\boldsymbol{L}$ is the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ (where $\boldsymbol{D}$ is the degree matrix of $\boldsymbol{A}$), and where $\kappa = 1e^{-1}$ is used so that $\boldsymbol{\Theta}$ is non-singular. For the data generation, a total of n samples $\{\boldsymbol{x}_i \in \mathbb{R}^p\}_{i=1}^n$ are drawn from an elliptical graphical model, parameterized by its covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Theta}^{-1}$ and density generator g, *i.e.*, $\boldsymbol{x}_i \sim \mathcal{ES}(\mathbf{0}, \boldsymbol{\Theta}^{-1}, g)$. For the density generator g, we use a Student's t-distribution with degrees of freedom ν.

Performance evaluation: For a given parameter setup, we sample the graph-data pair $(\boldsymbol{A}, \boldsymbol{X})$ as described above. Graph learning algorithms are then applied to the input data $\boldsymbol{X}$. These algorithms can provide different types of estimate structures with various inherent normalization, which are not always comparable. For a meaningful comparison, we consider evaluating the receiver operating characteristic (ROC) curves obtained from the estimated adjacency matrix $\boldsymbol{A}$.

The ROC curves displays the true positive rate (tpr) as function of the false positive rate (fpr): in our case, tpr denotes the capacity of the algorithm to recover actual edges of the algorithm, whereas fpr accounts for the false discovery of non-existing edges. For each curve, the area under curve (AUC) is computed. The AUC takes values in $[0, 1]$, with 1 indicating a perfect recovery of the true edges. These ROC curves are obtained as follows: First, we compute the conditional correlation coefficients

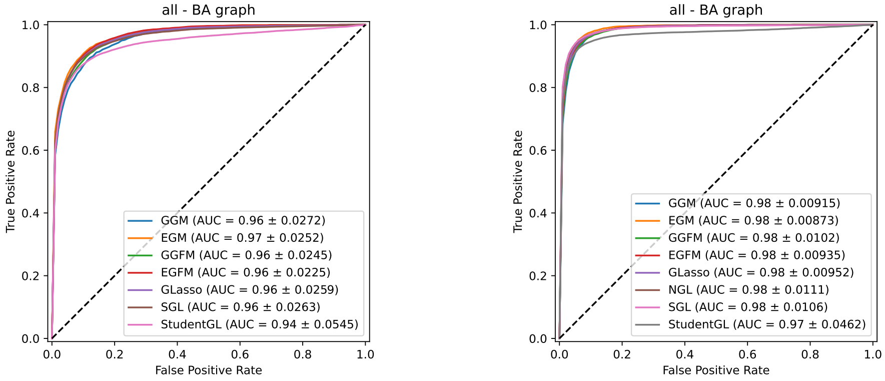
$$\tilde{\boldsymbol{\Theta}}_{ij} = \mathrm{corr}\left[x_q x_\ell | \boldsymbol{x}_{[\![1,p]\!]\setminus\{q,\ell\}}\right] = -\boldsymbol{\Theta}_{q\ell}/\sqrt{\boldsymbol{\Theta}_{qq}\boldsymbol{\Theta}_{\ell\ell}}. \tag{1}$$

obtained from any output estimate of the precision matrix $\boldsymbol{\Theta}$ (respectively, $\boldsymbol{\Sigma}^+$ if the algorithm provides an estimate of the covariance matrix). Given these coefficients estimates, a ROC curve is obtained by varying a threshold tol, i.e., the edge (i, j) is considered active if $\tilde{\boldsymbol{\Theta}}_{ij} > \mathrm{tol}$. The displayed ROC curves are finally obtained from the average of 50 Monte-Carlo experiments.

Compared methods: The proposed algorithms (GGM/GGFM/EGM/EGFM) are compared with state-of-the-art approaches to learn unstructured graphs: GLasso [5], which uses a Gaussian model and ℓ_1-norm as a sparse-promoting penalty; NGL and SGL [10, 7], which use a Gaussian Laplacian-constrained model with a concave penalty regularization (without prior knowledge on the number of graph components NGL and SGL solve a similar problem with a slightly different implementation); StGL [4], which generalizes the above to t-distributed data. For a fair comparison, the regularization parameters of all tested algorithms (including competing methods) are tuned to the best of our ability in order to display the best results. Notably, we notice that all these methods are not critically sensitive to a slight change of their regularization parameter λ (given a reasonable order of magnitude). Furthermore, a robustness analysis concerning this point (performance versus change in parameters) is also performed subsequently for the proposed methods.

3.2 Results: ROC curves comparisons in different setups

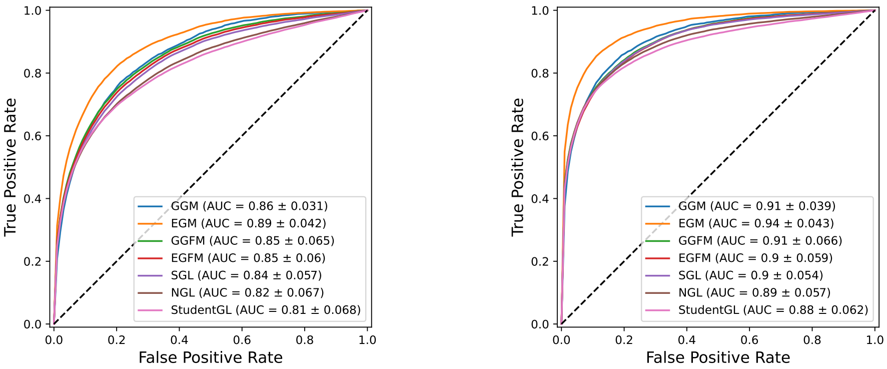
Figure 2, 3, 4, 5 show the results with $p = 50$ and $\nu = 3.5$ (non-Gaussian distributed data) for Barabási-Albert, Erdos-Rényi, Watts-Strogatz and random geometric models, respectively. Two sample support settings are considered, *i.e.*, $n = 2p$ and $n = 5p$. Overall, the proposed approaches outperform all compared methods (*i.e.*, larger AUC), except for Barabási-Albert where no significant differences are observed. Notable gains are obtained when n is low (i.e., when less samples are available). When n increases, the gap in performances between proposed and compared methods gets thinner. We also notice that EGM outperforms GGM, which was to be expected since the underlying data distribution is not Gaussian. As the generated graphs do not necessarily yield a low-rank structured covariance matrix, EGFM and GGFM do not outperform their full-rank counterparts in the considered setting. Still, these algorithms do not perform significantly worse, which means that one can reasonably reduce the dimension of the model (and benefit from a lower computational complexity).



(a) $n = 2p$

(b) $n = 5p$

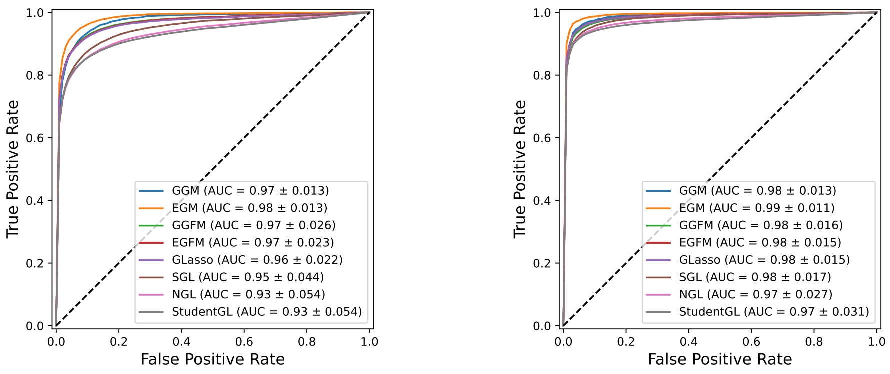
Fig. 2: Mean ROC curves obtained from estimated adjacency matrices $\hat{\mathbf{A}}$ of a **Barabási-Albert** model with GGM/EGM ($\lambda = 0.05$), GGFM/EGFM ($\lambda = 0.01; k = 20$), GLasso ($\alpha = 0.1$), NGL ($\lambda = 0.1$), SGL ($\alpha = 0.1$) and StudentGL (1 component) algorithms. Note that NGL is not displayed for $n = 2p$ because of numerical divergence.



(a) $n = 2p$

(b) $n = 5p$

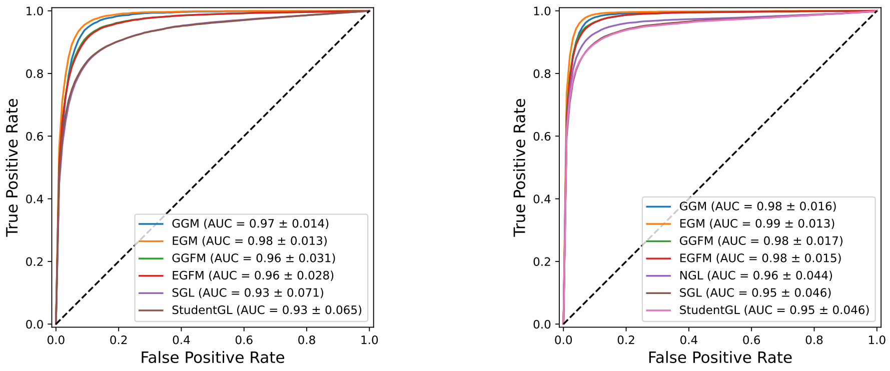
Fig. 3: Mean ROC curves obtained from estimated adjacency matrices $\hat{\mathbf{A}}$ of a **Erdős–Rényi** model with GGM/EGM ($\lambda = 0.05$), GGFM/EGFM ($\lambda = 0.01; k = 20$), GLasso ($\alpha = 0.1$), NGL ($\lambda = 0.1$), SGL ($\alpha = 0.1$) and StudentGL (1 component) algorithms. Note that GLasso is not displayed because of numerical divergence.



(a) $n = 2p$

(b) $n = 5p$

Fig. 4: Mean ROC curves obtained from estimated adjacency matrices $\hat{\mathbf{A}}$ of a **Watts Strogatz** model with GGM/EGM ($\lambda = 0.1$), GGFM/EGFM ($\lambda = 0.01; k = 10$), GLasso ($\alpha = 0.05$), NGL ($\lambda = 0.1$), SGL ($\alpha = 0.1$) and StudentGL (1 component) algorithms.



(a) $n = 2p$

(b) $n = 5p$

Fig. 5: Mean ROC curves obtained from estimated adjacency matrices $\hat{\mathbf{A}}$ of a **Random geometric** model with GGM ($\lambda = 0.1$), EGM ($\lambda = 0.1$), GGFM ($\lambda = 0.01; k = 20$), EGFM ($\lambda = 0.05; k = 20$), NGL ($\lambda = 0.02$), SGL ($\alpha = 0.1$) and StudentGL (1 component) algorithms. Note that GLasso is not displayed because of numerical divergence, neither is NGL for $n = 2p$.

3.3 GGM/GGFM/EGM/EGFM: robustness to parameters λ and k

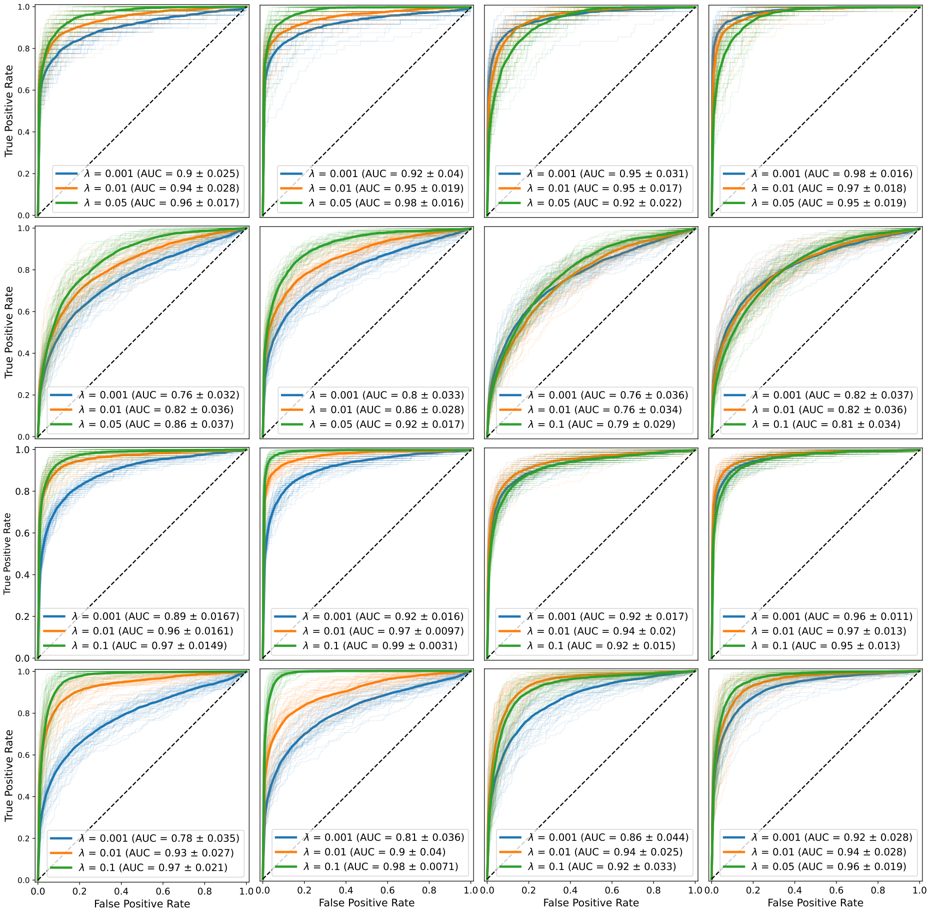
Figure 6 shows the sensitivity of the proposed algorithms to the sparsity parameter λ for each considered graph model, with $p = 50$ and $\nu = 3$. We observe that the performances are slightly different as λ varies. For certain graph models (especially random geometric) GGM/EGM appear to be sensitive to this parameter, so its order of magnitude should be chosen carefully. We also observe that factor-model based approaches GGFM/EGFM have the interesting property of being less sensitive to a changing λ for each graph model.

Figure 7 shows the sensitivity to the rank k, also with $p = 50$ and $\nu = 3$. We observe that acceptable performances are obtained when k is decreased. As the data is not necessarily low-rank, similar results are obtained compared to GGM/EGM, which was already observed in subsection 3.2. However, unlike GGM/EGM, we show that factor-model based approaches are particularly useful for providing interpretable and computable graphs from real-world data, as illustrated in the main body of this paper (section 4.2) and in section 4 below.

4 Concepts data

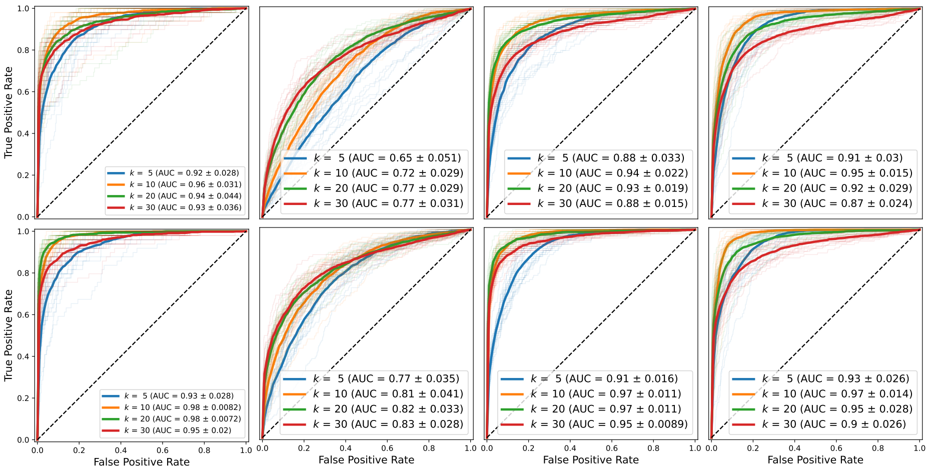
The *concepts* data set [8] collected by Intel Labs, includes $p = 1000$ nodes[6] and $n = 218$ semantic features. As stated in [2], "each node denotes a concept such as 'house', 'coat', and 'whale', and each semantic feature is a question such as 'Can it fly?', 'Is it alive?', and 'Can you use it?' The answers are on a five-point scale from 'definitely no' to 'definitely yes', conducted on Amazon Mechanical Turk". This dataset will only be used to assess the suitability of our factor model approaches, which are expected to be more efficient for such high-dimensional setting. Since GGFM and EGFM lead to very similar results, only the former is shown. In comparison, GGM and EGM did not provide as clean interpretable results, which motivates the use of factor models in practice. Because the "true" number of components is unclear in this data, methods that require prior setting of this variable were also found hard to exploit. The graph learned by the GGFM algorithm is presented in Figure 8. This graph is mostly composed of connected sub-components, each falling into interpretable categories of similar concepts ("tools" in grey, "body parts" in blue, etc.). A closer look on the graph in Figure 9 shows interesting nodes that can be interpreted as links between the clusters of concepts (*e.g.*, the node 'body' linking the cluster of "body parts" to the one of "whole bodies"). Another interpretation is that concepts with ambivalent meaning also act as nodes between these clusters ('gas' can be associated to "drinks" as well as to "natural elements").

[6] For a clearer visualization, only 500 nodes are processed, and isolated nodes are manually removed from the displayed graph.



(a) GGM (b) EGM (c) GGFM ($k = 20$) (d) EGFM ($k = 20$)

Fig. 6: **Sensitivity to λ**: mean ROC curves obtained from the estimated adjacency matrix $\hat{\mathbf{A}}$ of Barabási-Albert (first row), Erdős–Rényi (second row), Watts-Strogatz (third row) and Random geometric (fourth row) graphs with GGM/EGM/GGFM/EGFM algorithms with different values of the sparsity parameter λ.



(a) Barabási-Albert (b) Erdős–Rényi (c) Watts-Strogatz (d) Random geometric

Fig. 7: **Sensitivity to** k: mean ROC curves obtained from the estimated adjacency matrix $\hat{\mathbf{A}}$ with GGFM/EGFM algorithms (first and second row, respectively) with different values of the rank k. For each case, parameters λ are fixed according to Figure 6.

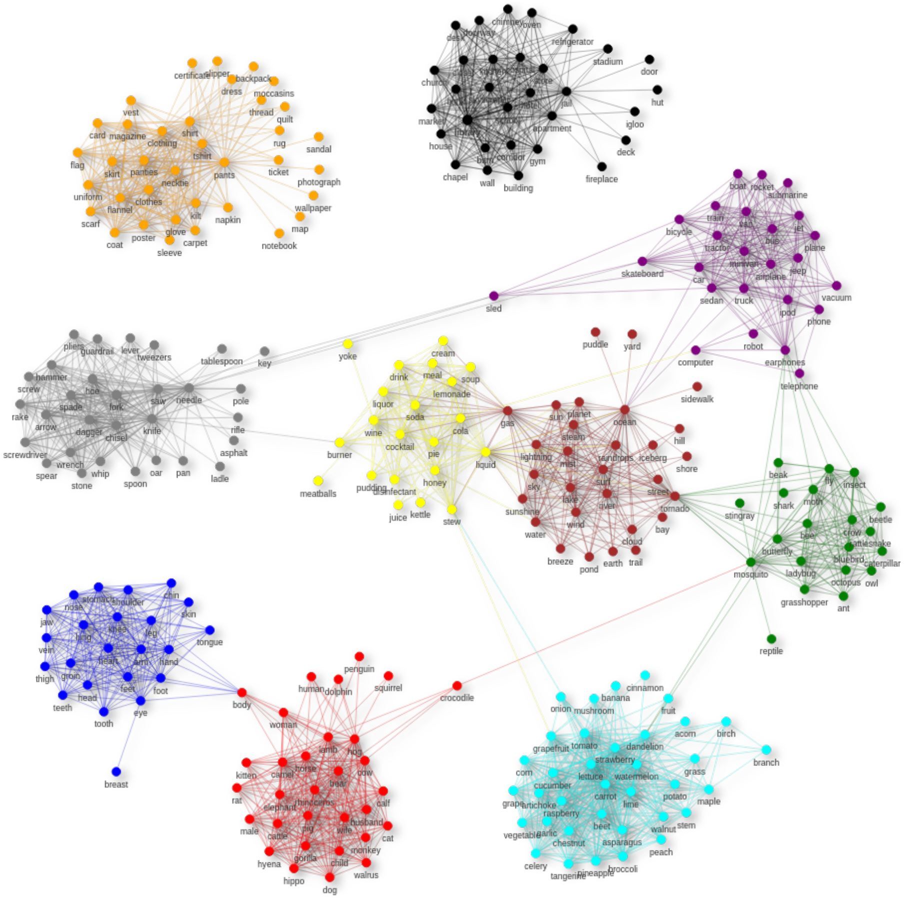


Fig. 8: Learned graph with the GGFM algorithm ($k = 10$ and $\lambda = 3.5$). Note that black and orange graph components are disconnected from the rest. Isolated nodes have been removed by hand for better visualization;

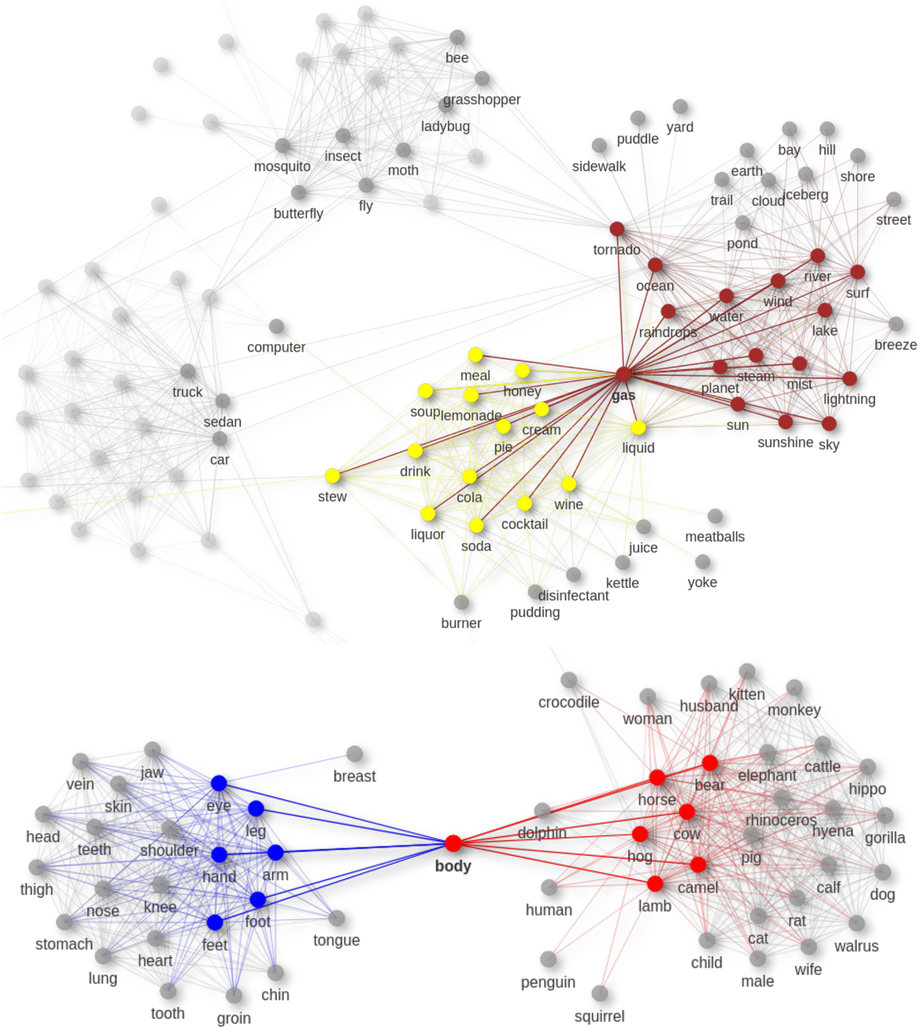


Fig. 9: First (colored) and second (greyed with label) neighbors of nodes ‘body’ and ‘gas’. These concepts connect two graph components each: body parts and whole bodies for ‘body’; drinks and natural elements for ‘gas’.

Bibliography

[1] Aytes, A. (2012). Return of the crowds: Mechanical turk and neoliberal states of exception. In *Digital labor*, pages 87–105. Routledge.

[2] Cai, J.-F., Cardoso, J. V. d. M., Palomar, D. P., and Ying, J. (2022). Fast projected newton-like method for precision matrix estimation under total positivity.

[3] Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

[4] de Miranda Cardoso, J. V., Ying, J., and Palomar, D. (2021). Graphical models in heavy-tailed markets. *Advances in Neural Information Processing Systems*, 34:19989–20001.

[5] Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

[6] Gray, M. L. and Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[7] Kumar, S., Ying, J., de Miranda Cardoso, J. V., and Palomar, D. P. (2020). A unified framework for structured graph learning via spectral constraints. *J. Mach. Learn. Res.*, 21(22):1–60.

[8] Lake, B. and Tenenbaum, J. (2010). Discovering structure by learning sparse graphs.

[9] Tubaro, P. and Casilli, A. A. (2019). Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics*, 46:333–345.

[10] Ying, J., de Miranda Cardoso, J. V., and Palomar, D. (2020). Nonconvex sparse graph learning under laplacian constrained graphical model. *Advances in Neural Information Processing Systems*, 33:7101–7113.