Uncertainties in Modeling Psychological Symptom Networks: the case of Suicide

Valeria Epelbaum[1,2,3][0009-0000-6889-1055], Sophie Engels[1][0009-0008-2412-6173], Rory O'Connor[3][0000-0002-3650-4994], Denny Borsboom[1][0000-0001-9720-4162], Derek De Beurs[2,3][0000-0002-0166-6897], and Valeria Krzhizhanovskaya[1,3][0000-0002-8247-129X]

[1] Computational Science Lab, Informatics Institute, Faculty of Science, The University of Amsterdam, Amsterdam, The Netherlands

[2] Faculty of Social and Behavioural Sciences, The University of Amsterdam, Amsterdam, The Netherlands

[3] Centre for Urban Mental Health, The University of Amsterdam, Amsterdam, The Netherlands

[4] Suicidal Behaviour Research Lab, School of Health & Wellbeing, University of Glasgow, Glasgow G12 8TA, UK

`V.Epelbaum@uva.nl`, `S.L.Engels@uva.nl`, `rory.oconnor@glasgow.ac.uk`, `D.Borsboom@uva.nl`, `D.deBeurs@uva.nl`, `V.Krzhizhanovskaya@uva.nl`

Abstract. In psychological research, network models are widely used to study symptoms of mental health disorders. However, these models often fail to account for uncertainty, leading to potentially misleading inferences. To address this issue, this study examines the robustness of psychological networks by analyzing a dataset of risk factors for suicidal behavior with multiple network algorithms. We compare two causal discovery algorithms—Hill Climbing (HC) and TABU search—and the Gaussian Graphical Model (GGM), a widely used statistical network model in psychology. Uncertainty is assessed along two dimensions: (1) the impact of noise, by introducing varying levels of white noise into the dataset, and (2) the effect of sample size reduction, by systematically decreasing the number of observations. Our results indicate that both HC and TABU search are highly sensitive to noise and sample size, with HC slightly outperforming TABU in terms of precision and recall. GGM performance declines gradually with increasing noise and sample size reduction, leading to sparser networks. For all algorithms, recall declined at a faster rate than precision. Finally, we examine the robustness of edges leading to suicidal ideation, finding that the edge from Depression to suicidal ideation remains relatively stable across conditions. This is a promising result, since many suicide interventions are based on treating depressive mood. Our results emphasize the importance of considering uncertainty in network-based psychological research, particularly when applying causal discovery algorithms.

Keywords: Uncertainty Quantification · Suicidal Behaviour · Causal Discovery Algorithms · Gaussian Graphical Model

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

1 Introduction

In the last decade, network models have gained increasing popularity in the social and behavioral sciences. For instance, in psychology, they are particularly valuable for modeling complex interactions between symptoms [3]. The most commonly used approach are undirected network models, where edges represent conditional independence relationships between variables. These are typically estimated using statistical models, such as Gaussian Graphical Models (GGMs), which infer edges based on partial correlations [9]. More recently, there has been a growing interest in causal discovery algorithms [12, 16]. These methods are often used to estimate directed networks, typically as Directed Acyclic Graphs (DAGs). Unlike GGMs, which capture associative patterns, DAGs attempt to infer causal structures using, for instance, score- or constraint-based methods.

Recently, increasing attention has been given to the replicability and robustness of network models, with researchers questioning the extent to which reliable inferences can be drawn from estimated networks [10]. Much of this work has focused on measurement-related uncertainty, such as how different response scales or node aggregation methods influence network structures. For example, [14] showed that using aggregated variables, whether through latent factors or mean scores, improved the performance compared to single-indicator models.

While this line of research addresses uncertainty on a measurement related level, it does not focus on uncertainty arising directly from the data, which can be an especially big issue in psychology [10]. For instance, most psychological data sets have small sample sizes to begin with. Additionally, some participants might misinterpret the asked questions or answer dishonestly, for example due to stigma. As a consequence, the data can become very noisy, which can manifest as spurious or missing relationships in networks. This is particularly concerning in clinical psychology, where network models are increasingly used to identify key symptoms or potential targets for interventions. While multiple studies have investigated the role of sample size [9, 14], other data-related sources of uncertainty, for instance noise, have not been studied in a psychological context yet. Thus, this study aims to address this gap by evaluating the performance of various network algorithms using a real data set. Specifically, two sources of data uncertainty will be examined: noise and sample size reduction.

To examine these effects of uncertainty, this study uses a cross-sectional dataset derived from [19]. The original authors studied risk factors for suicidal ideation through the lens of the Integrated Motivational Volitional theory (IMV) of suicidal behavior. According to that theory, suicidal thoughts arise from the interaction of different factors such as entrapment or defeat [19]. Previously, [7] already used this data to construct various networks using statistical models, such as GGM. Using this data set, the present study systematically introduces two types of data uncertainty—noise, and sample size—and assesses their impact on both directed and undirected network estimation. Additionally, the robustness of the causal edges leading towards suicidal thoughts, under both noise and data reduction, will be examined.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

1.1 Background and related work

Background Most undirected networks in psychology are based on the pairwise Markov random field (PMRF). In these models, when two variables are connected, it implies conditional dependence, while the absence of an edge signifies conditional independence given all other variables in the network. When the data follows a multivariate Gaussian distribution, the appropriate PMRF model is a GGM, described in more detail in 2.2.

A directed acyclcial graph (DAG) is a graph $G = (V, E)$, where $V = V_1, ...V_n$ represents a set of random variables, also referred to as nodes, and E is a set of directed edges. In this context, an edge from variable A to B ($A \rightarrow B$), indicates that variable A is considered a potential cause of variable B. However, to make this kind of conclusions, causal discovery algorithms are usually based on certain assumptions, such as faithfulness. An in-depth review of them is beyond the scope of this paper, but see [13] for a more detailed discussion.

There are multiple approaches to learn a DAG from a data set. For example, score-based algorithms approach this problem from a machine-learning perspective: they search the space of potential DAG structures to determine the one that reduces the score of a loss function.

Related Work In computational science, uncertainty in DAGs has been assessed across multiple levels. For instance, there are many proposed extensions and algorithms that take uncertainty into account [2]. Another strategy is to induce noise directly into the data. For example, [6] compared different DAG algorithms on noisy real-world data. Using multiple empirical networks as different case studies, they induced noise by adding missing/incorrect values or merging variables. They conclude that evaluation based on synthetic data can overestimate performance by anywhere from 10% up to 50%. [1] arrives at a similar conclusion.

Both of these studies compared the noisy graph to a reference network, which was either constructed by the authors themselves (e.g., via simulations) or supplied by experts. However, in social and behavioral sciences, such use of reference networks is rare: graphs are most commonly only learned from the data. Thus, to make our study more realistic, we directly learn the graph from the data, rather than relying on pre-constructed reference graphs.

2 Methodology

For each type of uncertainty, a single experiment will be conducted, resulting in two experiments in total. Each of them will feature all the algorithms described below, and will be run over 500 iterations. Following this, the frequency of edges leading to suicidal ideation identified by the different algorithms will be assessed under two conditions: varying noise levels and sample size. Thus, an edge is considered robust if it appears consistently, even under high noise and reduced sample size.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

The analysis is conducted in R, using the R studio environment [15]. For the undirected networks, the package *bootnet* is used [9], while the DAGs were constructed using *bnlearn* [17]. Further, *tidyverse* was used to clean the data and visualize the results [20].

2.1 Data description

The data comes from a population-based study on wellbeing, and is described in detail elsewhere [7]. Briefly, variables from the IMV theory were measured using various psychological questionnaires in the general population. For example, a question from the Barret Impulsivity scale (BIS), assessing Impulsivity, reads: "I act on impulse", and can be answered on a scale from 1 (not at all) to 4 (always). For a full overview of all the questionnaires and variables see [7].

To conduct further analysis, a common practice is to create sum scores for each questionnaire, i.e. to sum up all the observations for each individual per variable. For example, the BIS questionnaire consists of 30 questions. For each participant, all 30 scores were added up, reducing the amount of data from 30 raw scores to 1 sum score. For the current analysis, only participants experiencing suicidal ideation were used ($n = 333$), as otherwise the distribution of the variables would be non-Gaussian. Further, as the different questions were measured on different scales, all the scores were normalized to a 0 to 100 scale.

2.2 Algorithms for constructing networks from data

Networks are constructed using multiple algorithms. Undirected, statistical graphs were computed using a GGM, one of the most commonly used algorithms in Psychology [9]. The edges in a GGM represent partial correlation coefficients. To enforce sparsity in the network structure, the Graphical Least Absolute Shrinkage and Selection Operator (glasso) was applied, which imposes a penalty on the inverse covariance matrix, driving small partial correlations to zero [9].

For the DAGs, both the Hill Climbing (HC) and TABU algorithm are used. HC is a score-based algorithm, relying on a greedy search strategy. It starts off with an empty graph and makes local moves such as adding/removing an edge that lower a scoring function. Once a move does not improve the score, the algorithm terminates, which can result in it getting stuck in a local minimum. To avoid this problem, an extension of this algorithm was created that allows lower scoring moves. This extension is called TABU search. For both algorithms, the default scoring option of the Bayesian Information Criterion (BIC) will be used.

Each algorithms depends on a set of hyperparameters. For example, for TABU, the number of recent moves can be modified. This kind of memory serves the purpose of not revisiting older moves again. In our experiments, we used all the default settings, as outlined in Table 1.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

Table 1. Software and Hyperparameters for each algorithm.

Algorithm	Software	Default Hyperparameters
Gaussian Graphical Model (GGM)	bootnet	gamma = 0.5, number of tested lambda values = 100, Ratio of lowest lambda value compared to maximal lambda = 0.01
Hill Climbing (HC)	bnlearn	random restarts = 0, scores = BIC
TABU Search	bnlearn	tabu list = 10, escapes = 10, scores = BIC

2.3 Data Augmentation

The data was modified in two different ways. First, white noise was induced in the data set. Using a fixed amplitude α, we add an absolute noise level $\alpha\%$ (in percentage points of the normalized 0–100 scale) by drawing from:

$$\delta_i \sim \text{Uniform}([-\alpha, +\alpha])$$

where δ_i refers to the added noise to the ith observation. Thus, the noisy ith observation x_i^{noisy} is defined as:

$$x_i^{\text{noisy}} = x_i + \delta_i$$

with x_i being the original, i.e. unmodified, observation. For example, if the noise amplitude is $\alpha = 10$, then $\delta_i \in [-10, +10]$, so a point at $x_i = 50$ is perturbed somewhere in $[40, 60]$. Thus, as δ_i has the same range for every observation, the variance of the added noise is constant across all observation (i.e. *homoscedastic*) - a condition necessary for the GGM. Since the range of the data is bounded between 0 and 100, values above or below these thresholds are clipped. Second, data quantity was compromised. Here, rows of the observed data were randomly removed according to a percentage. For example, if this quantity percentage is 20%, then 67 randomly chosen rows would be removed, and the graph would be estimated based on the remaining 266 observations instead of 333.

2.4 Evaluation metrics

Evaluation in this context refers to how well the graph based on modified data performs in comparison to the graph learned from the original non-modified data. It should be noted, that in this sense, there is no "ground truth" network, since the true model is not known when working with an empirical data set. As the main goal is to estimate the impact of noise relative to the original data, the network learned from the non-augmented data will be used as a "reference". Subsequently, the graph based on the non-augmented data will be referred as the original graph. A similar use of reference graphs was also previously done by [13], who evaluated different DAG algorithms on empirical data sets without knowing the true underlying structure. Their study, however, relied on an fMRI data

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

set and one composed of cause–effect pairs—both of which differ substantially from the types of data typically encountered in clinical psychology. Our use of reference graphs also implies that for all three algorithms, the baseline graphs will be different. Three scoring metrics are calculated for each experiment:

1. Precision: Measures the proportion of predicted edges that are actually present in the original graph. It is calculated by dividing the True Positive rate (TP) by the sum of TP and False Positives (FP).
2. Recall: Measures how many true edges from the original graph are successfully recovered by the noisy graph. It is calculated by dividing TP by the sum of TP and False negatives (FN).
3. F1: The F1 score is the harmonic mean between Recall and Precision, and is a commonly used metric in the field of DAGs. It is defined as:

$$\mathrm{F1} = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Even though the F1 score is the mean of both precision and recall, it can still be informative to look at all three scores. The reason is that a low F1 score does not tell us whether it is low due to recall or precision.

3 Results

3.1 Baseline results

Figure 1 displays the baseline graphs. The GGM is very densely connected compared to the HC and the TABU algorithm: the average degree for the GGM is appoximately 9, while for HC and Tabu it is approximatly 2. On the other side, the HC and the TABU algorithms do not differ much. The main differences appear to be the direction of the edges. For instance, in the HC algorithm, feelings of defeat (Def) lead to stress (Str), while in the TABU graph this direction is reversed. The centrality measures degree and betweenness are shown in Table 2. Despite these structural differences, the nodes connected to suicidal ideation appear consistent across the algorithms: suicidal ideation is linked to thwarted belongingness, depression, mental well-being, and feelings of defeat. These associations are supported by prior empirical research [5]. Other empirically supported links, such as those between depression and perceived burdensomeness or impulsivity, are also present [7].

3.2 Uncertainty: Data Noise

Figure 2 shows the results for noisy data. Overall, the impact of noise is quite significant: 20% noise already results in a F1 value of less than 0.8 . HC slightly outperforms TABU search, both in regards to precision and recall. The added noise also had a higher impact on recall compared to precision, meaning that the noisy graph is less prone to false positive edges, but misses many edges that were present in the original graph (high false negative).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

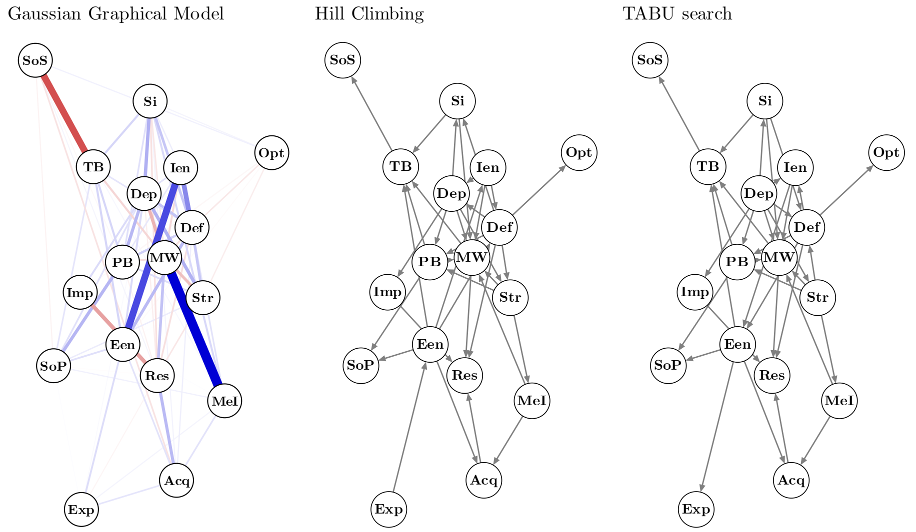

Fig. 1. Baseline graphs for the three algorithms. In the Gaussian graphical model, edge thickness reflects the strength of the partial correlation. Blue edges represent positive relationships, while the red ones the negative. Since Hill Climbing and TABU are structure learning algorithms, no edge weights are assigned. The symptoms are described in Table 2.

For the GGM, the performance gradually decreases up until 50% noise. After that, a slight increase in precision can be observed. A potential reason for this could be that the graph is becoming more sparse: the original graph contains 75 edges on average, which decreases to 59 edges for 50% noise, and further drops to 30 edges for 80% noise. As the number of the detected edges declines, so does the number of false positives. This decrease is happening at a faster rate compared to the decline in true positives, leading to an increase in precision[5].

This pattern is also visualized in Figure 3. The edges detected under higher levels of noise are usually around nodes with high centrality, such as feelings of defeat (Def) or Perceived Burdensomeness (PB) in the example of Figure 3.

3.3 Uncertainty: Data Quantity

Figure 2 shows the results for reducing the data quantity. As the sample size decreases, both precision and recall decline, with recall exhibiting a higher decrease across all three algorithms. HC and TABU perform similarly, though HC consistently outperforms TABU by a small margin. This is not surprising, since TABU can be viewed as an extension of HC. The performance of the GMM

[5] From 50% to 80% noise, the true positives drop by $\approx 46\%$, while the false positives drop by $\approx 63\%$.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

Table 2. Degree centrality (C_D) and betweenness centrality (C_B) of nodes for GGM, Hill Climbing, and TABU graphs. The term in the bracket of the Node column refers to the abbreviation used in the plotted graphs.

Node	**C_D GGM**	**C_D HC**	**C_D Tabu**	**C_B GGM**	**C_B HC**	**C_B Tabu**
Depression (Dep)	12	7	7	40	7	0
Perceived Burdensomeness (PB)	12	6	6	18	7	6
Defeat (Def)	11	8	8	46	18	15
Resilience (Res)	11	4	4	34	0	0
Suicidal Thoughts (Si)	10	4	4	4	2	1
External Entrapment (Een)	10	6	6	34	15	13
Stress (Str)	10	5	5	4	8	1
Mental Well-being (MW)	9	8	8	54	10	9
Social Perfectionism (SoP)	9	2	2	0	0	0
Mental Imagery (MeI)	9	3	3	0	4	1
Thwarted Belongingness (TB)	8	5	5	26	10	9
Internal Entrapment (Ien)	8	4	4	30	4	2
Optimism (Opt)	7	1	1	0	0	0
Acquired Capabilities (Acq)	7	3	3	4	2	2
Social Support (SoS)	7	1	1	4	0	0
Exposure to Suicide (Exp)	6	1	1	0	0	0
Impulsivity (Imp)	4	2	2	0	1	0

gradually decreases as well, but at a smaller rate compared to the causal discovery algorithms. Similar to the noise experiment, the GGM becomes sparser as the sample size decreases. While the retained edges are largely correct, many true edges from the original graph are omitted, likely due to the loss of power (see Figure 4). This results in a relatively high precision and low recall. Notably, when 80% of the data is missing, the graph is estimated to be empty in approximately 8% of the cases. When 90% of the data is missing, the graph is empty every time.

3.4 Uncertainty: Causal Edges leading to Suicidal Ideation

In the baseline HC graph two edges lead to suicidal ideation: depression and defeat, while in the TABU graph only depression leads to suicidal ideation. Figure 5 depicts the frequency of the edges across various noise levels and sample

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

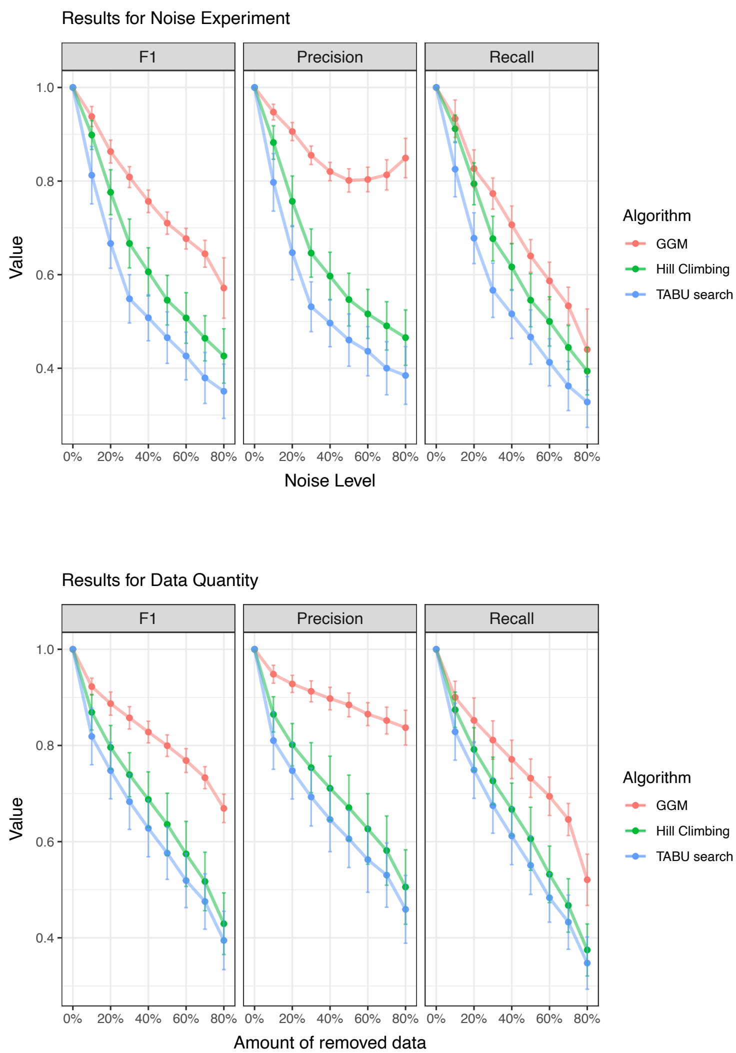

Fig. 2. Simulation results for all the evaluated uncertainty types across the causal discovery algorithms. The first plot depicts the results when adding noise and the second one for the reduction of the sample size. The errorbars correspond to the IQR range of observed values.

sizes (over 500 iterations). Depression → Suicidal ideation appears to be rela-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

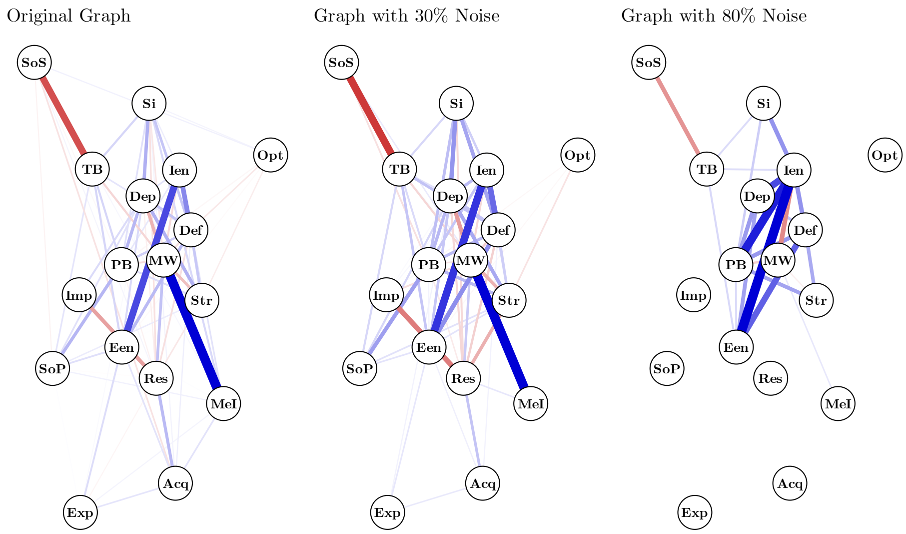

Fig. 3. GGM graph examples with noise. The graph on the left is the original graph, learned from the unmodified data, while the graph in the middle was learned with 30% noise and the one on the right with 80% noise. For node legend see Table 2.

tively robust across both noise and sample size reductions. At 40% noise, the edge was found in approximately 75% of the cases in both HC and TABU algorithms, while for 80% noise, it was slightly below 50%. A similar pattern is observed when reducing the sample size.

The edge Defeat → Suicidal ideation is also relatively stable in the HC algorithm, although less than Depression → Suicidal ideation. From 20% noise onward, the frequency of this edge declines faster than for Depression → Suicidal ideation, so that at 60% noise, it is only present in half of the cases. When reducing the sample size, a similar trend is observed: up until 40% of the data is removed, both edges are present in over 70% of the cases, but when 60% of the data is removed, Defeat → Suicidal is present in only half of the cases. For the TABU algorithm, this edge was not examined, since it was not present in the baseline graph in Figure 1.

4 Conclusion

This study examined the impact of data-related uncertainty—specifically noise and data reduction—on the estimation of psychological networks. Overall, we found that the network structures were highly sensitive. Both HC and TABU search in particular were especially affected, with HC slightly outperforming

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

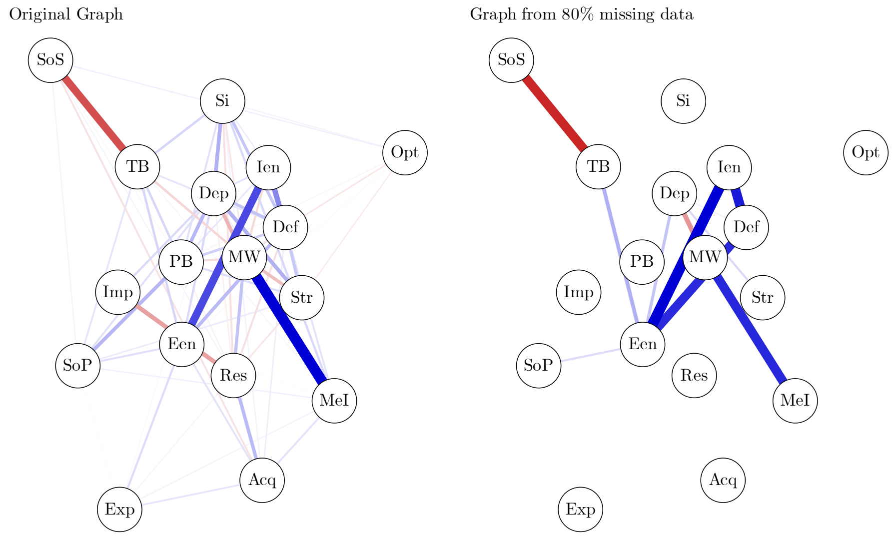

Fig. 4. GGM graph examples with compromised sample size. The graph on the left is the original graph, learned from the unmodified data, while the graph on the right was learned with data with 80% missing observations, meaning only 67 rows of observations were used instead of the full 333. For node legend see Table 2.

TABU search. The GGM graphs mostly became very sparse, with high noise or data reduction resulting only in few estimated edges.

Among the edges leading to suicidal ideation, the edge from depression to suicidal thoughts remained consistently detectable across both noise and sample size. This is a promising result, implying that intervening on depressive mood could lead to changes in suicidal thinking. Indeed, there are multiple trials suggesting that targeting depressive symptoms can lead to lower suicidal ideation [11, 18]. There are also specific suicide interventions, such as PROSPECT-an intervention for elderly people- whose main goal is to target depressive symptoms [4]. For a review of depression-based suicide interventions see [8].

Lastly, although this study is one of the first to examine data-related uncertainty in a psychological data set, there are a few limitations. For instance, we examined only one data set, meaning that for slightly different data, the results could be very different. Additionally, we examined only the network structure and did not analyze the parameter weights. Since determining the structure of a graph is typically the first step before estimating parameters, it is arguably a more critical aspect. Nevertheless, future research should further investigate the influence of data-related uncertainty on the edge weights. Despite these limitations, our study represents an important first step in exploring the impact of data-related uncertainties within a real psychological dataset. By highlighting the sensitivity of network structures to noise and data reduction, it lays the

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

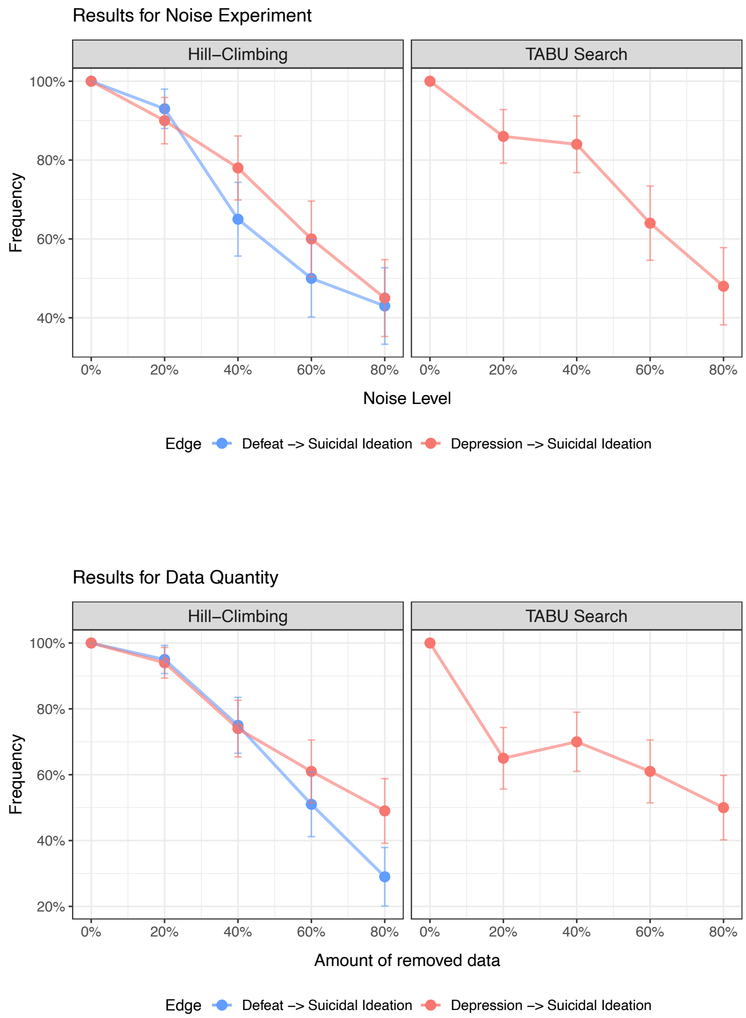

Fig. 5. Results for the experiments of the robustness of the edges leading to suicidal ideation for the Hill Climbing and TABU algorithm. The top plot depicts the results for the noise experiment and the bottom plot for the sample size. The y-axis refers to the frequency in percentage of the observed edges, averaged across all the observations. The error bars represent 95% confidence intervals.

groundwork for future research aimed at improving network estimation methods

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

in the context of psychological data.

Acknowledgments. This study was supported by the Research priority area Urban Mental Health of the University of Amsterdam, within the project "Understanding the Dynamics of Suicide: An Agent-Based Modeling Approach to Inform Intervention Strategies in an Urban Context".

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Beretta, S., Castelli, M., Goncalves, I., Henriques, R., Ramazzotti, D.: Learning the structure of Bayesian Networks: A quantitative assessment of the effect of different algorithmic schemes (Aug 2018). https://doi.org/10.48550/arXiv.1704.08676, http://arxiv.org/abs/1704.08676, arXiv:1704.08676 [cs]
2. Bhattacharya, R., Nabi, R., Shpitser, I., Robins, J.M.: Identification In Missing Data Models Represented By Directed Acyclic Graphs. In: Proceedings of The 35th Uncertainty in Artificial Intelligence Conference. pp. 1149–1158. PMLR (Aug 2020), https://proceedings.mlr.press/v115/bhattacharya20b.html, iSSN: 2640-3498
3. Borsboom, D.: A network theory of mental disorders. World Psychiatry **16**(1), 5–13 (2017). https://doi.org/10.1002/wps.20375, https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.20375, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wps.20375
4. Bruce, M.L., Pearson, J.L.: Designing an intervention to prevent suicide: PROSPECT (Prevention of Suicide in Primary Care Elderly: Collaborative Trial). Dialogues in Clinical Neuroscience **1**(2), 100–112 (Sep 1999), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181574/
5. Cleare, S., Wetherall, K., Eschle, S., Forrester, R., Drummond, J., O'Connor, R.C.: Using the integrated motivational-volitional (imv) model of suicidal behaviour to differentiate those with and without suicidal intent in hospital treated self-harm. Preventive Medicine **152**, 106592 (2021)
6. Constantinou, A.C., Liu, Y., Chobtham, K., Guo, Z., Kitson, N.K.: Large-scale empirical validation of Bayesian Network structure learning algorithms with noisy data. International Journal of Approximate Reasoning **131**, 151–188 (Apr 2021). https://doi.org/10.1016/j.ijar.2021.01.001, https://www.sciencedirect.com/science/article/pii/S0888613X21000025
7. De Beurs, D., Fried, E.I., Wetherall, K., Cleare, S., O' Connor, D.B., Ferguson, E., O'Carroll, R.E., O' Connor, R.C.: Exploring the psychology of suicidal ideation: A theory driven network analysis. Behaviour Research and Therapy **120**, 103419 (Sep 2019). https://doi.org/10.1016/j.brat.2019.103419
8. Dueweke, A.R., Bridges, A.J.: Suicide interventions in primary care: A selective review of the evidence. Families, Systems, & Health **36**(3), 289–302 (Sep 2018). https://doi.org/10.1037/fsh0000349, https://doi.apa.org/doi/10.1037/fsh0000349
9. Epskamp, S., Borsboom, D., Fried, E.I.: Estimating psychological networks and their accuracy: A tutorial paper. Behavior Research Methods **50**(1), 195–212 (Feb 2018). https://doi.org/10.3758/s13428-017-0862-1

10. Huth, K.B.S., Keetelaar, S., Sekulovski, N., van den Bergh, D., Marsman, M.: Simplifying Bayesian analysis of graphical models for the social sciences with easybgm: A user-friendly R-package. advances.in/psychology **2**, e66366 (Mar 2024). https://doi.org/10.56296/aip00010, https://advances.in/psychology/10.56296/aip00010/, publisher: Advances.in Section: Methods
11. Lapierre, S., Erlangsen, A., Waern, M., De Leo, D., Oyama, H., Scocco, P., Gallo, J., Szanto, K., Conwell, Y., Draper, B., Quinnett, P.: A Systematic Review of Elderly Suicide Prevention Programs. Crisis **32**(2), 88–98 (Mar 2011). https://doi.org/10.1027/0227-5910/a000076, https://econtent.hogrefe.com/doi/10.1027/0227-5910/a000076
12. McNally, R.J., Heeren, A., Robinaugh, D.J.: A Bayesian network analysis of posttraumatic stress disorder symptoms in adults reporting childhood sexual abuse. European Journal of Psychotraumatology **8**(sup3), 1341276 (Jul 2017). https://doi.org/10.1080/20008198.2017.1341276, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5632780/
13. Niu, W., Gao, Z., Song, L., Li, L.: Comprehensive Review and Empirical Evaluation of Causal Discovery Algorithms for Numerical Data (Sep 2024). https://doi.org/10.48550/arXiv.2407.13054, http://arxiv.org/abs/2407.13054, arXiv:2407.13054 [cs] version: 2
14. de Ron, J., Robinaugh, D.J., Fried, E.I., Pedrelli, P., Jain, F.A., Mischoulon, D., Epskamp, S.: Quantifying and addressing the impact of measurement error in network models. Behaviour Research and Therapy **157**, 104163 (Oct 2022). https://doi.org/10.1016/j.brat.2022.104163, https://www.sciencedirect.com/science/article/pii/S0005796722001346
15. RStudio Team: R: A Language and Environment for Statistical Computing (2020), http://www.rstudio.com/
16. Schutzeichel, F., Waldorp, L.J., Aan Het Rot, M., Glashouwer, K.A., Frey, M.I., Wiers, R.W., de Jong, P.J.: Life meaning and feelings of ineffectiveness as transdiagnostic factors in eating disorder and comorbid internalizing symptomatology - A combined undirected and causal network approach. Behaviour Research and Therapy **172**, 104439 (Jan 2024). https://doi.org/10.1016/j.brat.2023.104439
17. Scutari, M.: Learning Bayesian Networks with the **bnlearn** *R* Package. Journal of Statistical Software **35**(3) (2010). https://doi.org/10.18637/jss.v035.i03, http://www.jstatsoft.org/v35/i03/
18. Sufrate-Sorzano, T., Santolalla-Arnedo, I., Garrote-Cámara, M.E., Angulo-Nalda, B., Cotelo-Sáenz, R., Pastells-Peiró, R., Bellon, F., Blanco-Blanco, J., Juárez-Vela, R., Molina-Luque, F.: Interventions of choice for the prevention and treatment of suicidal behaviours: An umbrella review. Nursing Open **10**(8), 4959–4970 (Aug 2023). https://doi.org/10.1002/nop2.1820, https://onlinelibrary.wiley.com/doi/10.1002/nop2.1820
19. Wetherall, K., Cleare, S., Eschle, S., Ferguson, E., O'Connor, D.B., O'Carroll, R.E., O'Connor, R.C.: From ideation to action: Differentiating between those who think about suicide and those who attempt suicide in a national study of young adults. Journal of Affective Disorders **241**, 475–483 (Dec 2018). https://doi.org/10.1016/j.jad.2018.07.074, https://linkinghub.elsevier.com/retrieve/pii/S0165032718312898
20. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D.,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10

Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H.: Welcome to the Tidyverse. Journal of Open Source Software **4**(43), 1686 (Nov 2019). https://doi.org/10.21105/joss.01686, https://joss.theoj.org/papers/10.21105/joss.01686

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97573-8_10