Thresholded FOCI for Undirected Graph Structure Learning with Binary Variables

Konrad Furmańczyk[1[0000−0002−7683−4787]]

Institute of Information Technology, Warsaw University of Life Sciences, Warsaw, Poland konrad_furmanczyk@sggw.edu.pl

Abstract. We propose a new method for undirected graph structure learning for binary variables, based on a principled extension of the FOCI selection procedure applied node-wise to recover the Markov blanket of each variable. The final undirected dependency graph is obtained as the union of these Markov blankets.
To adapt FOCI to the binary setting, we introduce an empirical version of a conditional dependence coefficient and incorporate a thresholding step to control false discoveries. Theoretical results establish exact recovery of the true graph with high probability in finite samples.
The performance of the method is evaluated in a simulation study on artificial data generated from binary structural equation models based on logistic regression.

Keywords: Thresholded FOCI · Undirected Binary Graphs · Markov Blanket · SEM.

1 Introduction

Graphical dependency models have been intensively studied over the last two decades [8,6,10]. The core of this research has focused on undirected Gaussian graphical models based on the multivariate normal distribution, Bayesian networks, vine copulas [3], and more recently, undirected graphical models for nonparanormal distributions [9]. Graphical models provide a convenient and interpretable framework for exploring multivariate dependence structures, with particular emphasis on conditional independence relationships.

Several extensions of graphical models beyond the Gaussian setting have been proposed, including models for elliptical distributions [7,15,13]. In such models, the graph structure is typically determined by zeros in the inverse covariance matrix or, equivalently, by vanishing partial correlations. However, outside the Gaussian case, partial correlation graphs generally lack a direct interpretation in terms of conditional independence [13].

In contrast, relatively few methods are available for modeling dependence structures among binary variables. A popular approach is based on the autologistic model [4] for undirected graphs, with extensions to directed graphical models discussed in [11].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

This motivates the development of alternative, model-free approaches for recovering dependency structures in binary data.

In this work, we propose a new procedure for recovering an undirected dependency graph for binary random variables. The method is based on a modification of the FOCI algorithm [2].

The key idea is to apply the procedure separately to each node in order to estimate its Markov blanket, and then to construct the final graph skeleton by aggregating the estimated Markov blankets across all nodes.

The original FOCI procedure relies on rank-based estimators and is therefore not suitable for binary data. To address this limitation, we replace the rank-based dependence measure with an empirical version of the Azadkia–Chatterjee coefficient of conditional dependence [2], [1], which is well defined for binary variables. Next, to reduce false discovery we put some threshold $\tau > 0$ for the FOCI method. Our main theoretical results are the result of proper recovery FOCI procedure for oracle situation, when we have full knowledge of the random distribution (Theorem 1) and recovery theorem for undirected graphs (Theorem 3). The paper is organized as follows. In Section 2, we introduce all necessary definitions and formulate the oracle recovery result for the population version of the forward FOCI selection procedure for a single target variable Y (Theorem 1). In the same section, we state Theorem 2, which establishes the correctness of the FOCI selection procedure for a single target variable based on an empirical version of the conditional dependence coefficient.

In Section 3, we present a consistency theorem for undirected graph recovery using the modified node-wise FOCI procedure. Specifically, Theorem 3 shows that the true dependency graph can be recovered as the union of the Markov blankets obtained for all vertices.

Finally, in Section 4, we describe a thresholded modification of the FOCI algorithm designed for skeleton recovery in undirected graphs.

In a Simulation Study, we are checking the quality of the proposed selection procedure for the the simulated SEM binary models. In this section, we introduce a simple cross-validation procedure for selecting the threshold and provide a comparison with the PC-stable algorithm for skeleton recovery [5]. In the Discussion and Conclusion section, we provide an overview of the obtained results and discuss directions for future work. All proofs are given in the Appendix. The R code used for the simulation experiments is available upon request and will be made publicly accessible in a future repository.

2 Markov Blanket Selection Consistency of FOCI

2.1 Oracle recovery for FOCI

In this subsection, we formulate equivalent conditions $(B0)$–$(B1)$ under which the oracle FOCI procedure correctly identifies the true active set of predictors of a single target variable Y.

Let Y be the response, $X = (X_1, \dots, X_p)$ the predictors, and $S^\star \subset \{1, \dots, p\}$ the Markov blanket of Y.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

By definition:

- **Markov Blanket=Sufficiency:**

$$Y \perp\!\!\!\perp X_{-S^\star} \mid X_{S^\star}.$$

  This means that Y and $X_{-S^\star}$ are conditionally independent given $X_{S^\star}$.
- **Markov boundary** if Markov blanket and minimal.
- **Minimality:**

$$Y \not\perp\!\!\!\perp X_j \mid X_{S^\star \setminus \{j\}}, \quad \forall j \in S^\star.$$

Let $\widehat{S}_n$ be the set selected by FOCI.

Key property of the Azadkia–Chatterjee coefficient[2] Define the population coefficient for random variable Y and random vectors X, Z:

$$T(Y, Z \mid X) = \frac{\mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid X)]}{\mathbb{E}[\mathrm{Var}(Y \mid X)]}.$$

We now introduce several technical lemmas used to prove the oracle recovery result for the FOCI procedure (Theorem 1). This result is essential for establishing our main theorem (Theorem 3) on undirected graph recovery.

Lemma 1. ***(Exact characterization):*** *Suppose that Y is not almost surely equal to a measurable function of vector X*

$$T(Y, Z \mid X) = 0 \iff Y \perp\!\!\!\perp Z \mid X.$$

This equivalence holds without smoothness, parametric, or linearity assumptions.
Markov blanket $S^\star$ is minimal if for some $\delta > 0$ we have that for any $S \subsetneq S^\star$, we have
(B0)

$$\max_{j \in S^\star \setminus S} T(Y, X_j \mid X_S) \;\geq\; \delta \;>\; \sup_{k \notin S^\star} T(Y, X_k \mid X_S).$$

Define the population functional

$$Q(Y, X_S) := \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X_S])]\,,$$

Since

$$T(Y, X_j \mid X_S) = \frac{Q(Y, X_{S \cup \{j\}}) - Q(Y, X_S)}{\mathbb{E}[\mathrm{Var}(Y \mid X_S)]}.$$

(B0) is equivalent to,
(B1) for some $\delta > 0$

$$\max_{j \in S^\star \setminus S} \big(Q(Y, X_{S \cup \{j\}}) - Q(Y, X_S)\big) \;\geq\; \delta \;>\; \max_{k \notin S^\star} \big(Q(Y, X_{S \cup \{k\}}) - Q(Y, X_S)\big).$$

We define the signal gap

$$\delta := \inf_{S \text{ not sufficient}} \; \sup_{j \notin S} \Big(Q(Y, X_{S \cup \{j\}}) - Q(Y, X_S)\Big).$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

Population-level forward selection (oracle behavior) Consider the idealized algorithm that uses T instead of T_n.

Lemma 2. ***Population ordering property*** *Assume (B1). Let $S \subsetneq S^\star$. Then:*

- *There exists $j \in S^\star \setminus S$ such that*

$$T(Y, X_j \mid X_S) \geq \delta > 0.$$

- *For all $k \notin S^\star$,*

$$T(Y, X_k \mid X_S) \leq \delta.$$

Lemma 3. *(Correct stopping): If $S = S^\star$, then*

$$\max_{j \notin S} T(Y, X_j \mid X_S) = 0,$$

so the stopping rule triggers if and only if the Markov blanket is complete.

From Lemmas 1-3 we have

Theorem 1. *(Oracle recovery): The population forward selection recovers $S^\star$ exactly and uniquely.*

2.2 Computation of the Conditional Dependence Coefficient for Binary Variables

In this subsection, we construct an empirical version of the conditional dependence coefficient for binary data. Let $(Y_i, X_i, Z_i), \quad i = 1, \ldots, n$, be i.i.d. observations of the random vector (Y, X, Z), where $Y, X, Z \in \{0, 1\}$. Our goal is to construct a consistent estimator $T_n(Y, Z \mid X)$ of $T(Y, Z \mid X)$ by estimating the numerator and denominator separately. First, we consider population version of $\mathbb{E}[\mathrm{Var}(Y \mid X)]$. Since Y is binary, for each $x \in \{0, 1\}$,

$$\mathrm{Var}(Y \mid X = x) = p_x(1 - p_x), \quad p_x := \mathbb{P}(Y = 1 \mid X = x).$$

Therefore,

$$\mathbb{E}[\mathrm{Var}(Y \mid X)] = \sum_{x \in \{0,1\}} \mathbb{P}(X = x)\, p_x(1 - p_x).$$

Define the empirical probabilities

$$\widehat{\mathbb{P}}(X = x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i = x\},$$

$$\hat{p}_x := \frac{\sum_{i=1}^{n} \mathbf{1}\{Y_i = 1,\, X_i = x\}}{\sum_{i=1}^{n} \mathbf{1}\{X_i = x\}}.$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

The denominator estimator is then

$$\widehat{D}_n := \sum_{x\in\{0,1\}} \widehat{\mathbb{P}}(X=x)\,\hat{p}_x(1-\hat{p}_x).$$

Next, we consider population version of the numerator $\mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid X)]$.

For each $x, z \in \{0,1\}$, define $p_{xz} := \mathbb{P}(Y=1 \mid X=x, Z=z)$. Then

$$\mathbb{E}[Y \mid X=x, Z=z] = p_{xz}, \quad \mathbb{E}[Y \mid X=x] = p_x.$$

Hence,

$$\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid X=x) = \sum_{z\in\{0,1\}} \mathbb{P}(Z=z \mid X=x)\big(p_{xz}-p_x\big)^2.$$

Taking expectation over X,

$$\mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid X)] = \sum_{x\in\{0,1\}} \mathbb{P}(X=x) \sum_{z\in\{0,1\}} \mathbb{P}(Z=z \mid X=x)\big(p_{xz}-p_x\big)^2.$$

Define the empirical conditional probabilities

$$\hat{p}_{xz} := \frac{\sum_{i=1}^n \mathbf{1}\{Y_i=1,\, X_i=x,\, Z_i=z\}}{\sum_{i=1}^n \mathbf{1}\{X_i=x,\, Z_i=z\}},$$

$$\widehat{\mathbb{P}}(Z=z \mid X=x) := \frac{\sum_{i=1}^n \mathbf{1}\{X_i=x,\, Z_i=z\}}{\sum_{i=1}^n \mathbf{1}\{X_i=x\}}.$$

The numerator estimator is

$$\widehat{N}_n := \sum_{x\in\{0,1\}} \widehat{\mathbb{P}}(X=x) \sum_{z\in\{0,1\}} \widehat{\mathbb{P}}(Z=z \mid X=x)\big(\hat{p}_{xz}-\hat{p}_x\big)^2.$$

We define the empirical conditional dependence coefficient as

$$T_n(Y, Z \mid X) = \frac{\widehat{N}_n}{\widehat{D}_n}.$$

In the fully discrete binary setting, the nearest-neighbor construction of Azadkia and Chatterjee becomes degenerate due to ties. The plug-in estimator defined above is the natural discrete analogue of their functional and should be preferred in this case.

From the Law of Large Numbers,

$$\widehat{N}_n \xrightarrow{a.s.} \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid X)], \quad \widehat{D}_n \xrightarrow{a.s.} \mathbb{E}[\mathrm{Var}(Y \mid X)].$$

Therefore,

Lemma 4.

$$T_n(Y, Z \mid X) \xrightarrow{a.s.} T(Y, Z \mid X).$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

2.3 Binary version of Theorem 6.1 Azadkia and Chatterjee [2]

Now, we present the correctness of the FOCI selection procedure for a single target variable based on an empirical version of the conditional dependence coefficient. This result is crucial in the proof of our main Theorem 2. Assume that all random variables

$$Y \in \{0,1\}, \qquad X = (X_1, \dots, X_p) \in \{0,1\}^p$$

are binary. Let $S^\star \subset \{1,\dots,p\}$ be a minimal sufficient set, i.e.

$$Y \perp X_{-S^\star} \mid X_{S^\star}.$$

Define the population functional

$$Q(Y, X_S) := \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X_S])],$$

and its empirical plug-in estimator

$$Q_n(Y, X_S) := \sum_{x_S} \widehat{\mathbb{P}}(X_S = x_S)\big(\hat{p}_{x_S} - \hat{p}\big)^2,$$

where

$$\hat{p}_{x_S} = \mathbb{P}_n(Y = 1 \mid X_S = x_S), \qquad \hat{p} = \mathbb{P}_n(Y = 1).$$

Let $\widehat{S}$ be the set selected by the FOCI algorithm using Q_n.

Theorem 2. *Assume*
(B2) for every subset $S \subset \{1,\dots,p\}$ with

$$|S| \le \left\lfloor \frac{1}{\delta} \right\rfloor + 1$$

and
(B3)

$$\mathbb{E}[\mathrm{Var}(Y \mid X_S)] > 0 \quad \textit{for all } |S| \le \left\lfloor \frac{1}{\delta} \right\rfloor + 1.$$

Then there exist constants $L_1, L_2, L_3 > 0$, depending only on δ, such that

$$\mathbb{P}(\widehat{S} \textit{ is sufficient}) \;\ge\; 1 - L_1 p^{L_2} e^{-L_3 n}.$$

The proof is very similar as Theorem 6.1 [2]. We only need new version of the following

Lemma 5. *Assume (B1)-(B3). Let*

$$E' := \left\{ \max_{1 \le k \le K} \left| Q_n(Y, X_{S_k}) - Q(Y, X_{S_k}) \right| \le \delta/8 \right\},$$

There exist constants $L_1, L_2, L_3 > 0$ such that

$$\mathbb{P}(E') \ge 1 - L_1 p^{L_2} e^{-L_3 n}.$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

3 Setting for skeleton recovery

In this section, we present a result for undirected graph recovery using the modified node-wise FOCI procedure. Specifically, Theorem 3 shows that the true dependency graph can be recovered with high probability as the union of the Markov blankets obtained for all vertices. Let

$$Y = (Y_1, \ldots, Y_p), \qquad Y_i \in \{0, 1\},$$

be a collection of binary random variables. Assume that the joint distribution of Y is Markov with respect to an undirected graph $G = (V, E)$, where $V = \{1, \ldots, p\}$.

For each node i, let

$$\mathrm{MB}(i) \subset V \setminus \{i\}$$

denote the (unique and minimal-for each i consider (B1)) Markov blanket of Y_i, i.e.

$$Y_i \perp Y_{V \setminus (\{i\} \cup \mathrm{MB}(i))} \mid Y_{\mathrm{MB}(i)}.$$

Estimation procedure

For each $i \in \{1, \ldots, p\}$, apply the (binary) FOCI algorithm with target Y_i and predictors Y_{-i}, obtaining an estimated sufficient set

$$\widehat{S}_i \subset \{1, \ldots, p\} \setminus \{i\}.$$

Define the estimated skeleton

$$\widehat{E} := \big\{\{i, j\} : j \in \widehat{S}_i \text{ or } i \in \widehat{S}_j\big\}.$$

Signal gap assumptions

For each node i, define

$$\delta_i := \inf_{S \text{ not sufficient for } Y_i} \sup_{j \notin S} \Big(Q(Y_i, Y_j \mid Y_S) - Q(Y_i \mid Y_S)\Big).$$

Because all variables are binary and the state space is finite,

$$\delta_i > 0 \qquad \text{for all } i.$$

Let

$$\delta := \min_{1 \le i \le p} \delta_i > 0.$$

Theorem 3. *(Skeleton recovery for binary variables). Assume the Signal Gap Assumptions and conditions (B2)–(B3). There exist constants $L_1, L_2, L_3 > 0$ such that*

$$\mathbb{P}\Big(\widehat{E} = E\Big) \ \ge\ 1 - L_1 p^{L_2} e^{-L_3 n}.$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

In Theorem 3 we recover the exact skeleton, not just supersets. No faithfulness or linearity is assumed. Binary variables eliminate all smoothness and NN technicalities. Errors decay exponentially in n and polynomially in p. The proof is a direct lift of Azadkia–Chatterjee Theorem 6.1 [2] (see Appendix) one FOCI run per node, a union bound across nodes.

4 Markov Blanket Recovery via thresholded FOCI

In this section, we introduce a thresholded modification of the FOCI algorithm tailored to skeleton recovery in undirected graphs. The proposed modification incorporates an explicit thresholding step to control spurious selections and reduce false discoveries, while preserving the ability of the original FOCI procedure to identify relevant dependencies.
For each node $i = 1, \ldots, p$:

1. Initialize $S_i^{(0)} = \varnothing$.
2. At step $k \geq 1$, select
$$j_k = \arg\max_{j \notin S_i^{(k-1)}} T_n(Y_i, Y_j \mid Y_{S_i^{(k-1)}}).$$
3. Update
$$S_i^{(k)} = S_i^{(k-1)} \cup \{j_k\}.$$
4. Stop when
$$\max_{j \notin S_i^{(k)}} T_n(Y_i, Y_j \mid Y_{S_i^{(k)}}) \leq \tau_n, \qquad \tau_n \downarrow 0.$$

The skeleton estimator is defined as

$$\widehat{G} = \bigcup_{i=1}^{p} \widehat{\mathrm{MB}}(i),$$

where $\widehat{\mathrm{MB}}(i)$ denotes the estimated Markov blanket of node i obtained using the thresholded FOCI procedure.

4.1 Cross-validation via predictive risk

In practice we should choose the threshold $\tau = \tau_n$ via cross-validation. Since the true graph structure is unknown, the threshold is selected based on predictive performance, which serves as a proxy for the quality of the recovered Markov blankets. A valid Markov blanket $S_i(\tau)$ should provide good predictive power for Y_i. Here is the detail for the procedure:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

Procedure. Fix a grid of thresholds τ and perform K-fold cross-validation. For each fold $k = 1, \ldots, K$ and each node i:

1. Estimate the Markov blanket on the training data:

$$S_i^{(-k)}(\tau).$$

2. Fit a logistic regression model of Y_i on $S_i^{(-k)}(\tau)$.
3. Evaluate the log-loss on the validation fold:

$$\text{LogLoss}_i^{(k)}(\tau) = -\frac{1}{|V_k|} \sum_{t \in V_k} \log \widehat{P}^{(-k)}(Y_i^{(t)} \mid S_i^{(-k)}(\tau)).$$

Selection. Choose

$$\hat{\tau} = \arg\min_{\tau} \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{p} \text{LogLoss}_i^{(k)}(\tau).$$

5 Numerical Study

We will simulate binary SEM on a Random DAG. Let $X_j \in \{0, 1\}$ be the binary outcome for node j, for $j = 1, \ldots, p$. Let $pa(j)$ denote the set of parent nodes of node j in the DAG A, and let β_{ij} be the coefficient associated with the edge $i \to j$ if $A_{ij} = 1$.

Logistic SEM Equation

For each node $j = 1, \ldots, p$, the structural equation is:

$$\text{logit}\Big(\Pr(Y_j = 1 \mid Y_{pa(j)})\Big) = \sum_{i \in pa(j)} \beta_{ij} Y_i,$$

DAG Generation

The DAG A is generated as a random *upper-triangular adjacency matrix* of size $p \times p$ as follows:

1. Initialize A as a $p \times p$ matrix of zeros.
2. For each pair of nodes $i < j$, assign an edge $i \to j$ with probability edge_prob:

$$A_{ij} = \begin{cases} 1 & \text{with probability edge\_prob,} \\ 0 & \text{otherwise.} \end{cases}$$

3. By construction, A is strictly upper-triangular, which ensures there are no cycles (acyclic).

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

Once the DAG is generated, the SEM is simulated using the logistic model for each node, where each node depends only on its parents. The goal of the experiment is to recover the skeleton of the underlying causal graph. Edge coefficients β are independently drawn from the uniform distribution on $[0.8, 1.5]$. In our simulation study, we consider sample sizes $n = 300$ and $n = 3000$, and numbers of nodes $p = 15$ and $p = 30$. The underlying data-generating process is based on a random DAG with edge probability equal to 0.05 or 0.15. The target of inference is the recovery of the corresponding undirected dependency graph, which follows an Erdős–Rényi model.

The threshold is set to

$$\tau_n = c\sqrt{\frac{\log p}{n}},$$

with $c \in \{0.5, 1, 2\}$. The reported results are based on $R = 300$ independent Monte Carlo repetitions. For each setting, we report the empirical means and standard deviations of the True Positive Rate (TPR) and the False Positive Rate (FPR) for correct edge recovery from the simulated skeleton, as summarized in Table 1.

The results confirm the theoretical scaling $\tau_n \asymp \sqrt{\log(p)/n}$. For sparse graphs (edge_prob = 0.05), FOCI achieves near-perfect skeleton recovery already for $n = 3000$, while denser graphs require larger samples. Smaller constants $c < 1$ increase TPR at the cost of higher FPR, whereas larger c produce overly conservative estimators.

In Table 2, panel (A), we present results for $c = 0$ from $R = 300$ Monte Carlo repetitions. For $c = 0$ we have $\tau_n = 0$, which corresponds to the original FOCI procedure without thresholding. In this case, all selected variables are included in the Markov blanket, which typically leads to very high true positive rates but also to extremely large false positive rates. These results demonstrate that thresholding is essential for consistent graph recovery, since the unthresholded FOCI procedure produces nearly complete graphs even for large sample sizes.

In practice, the constant c may be selected using stability-based criteria or validation on subsamples. Table 2, panel (B), presents cross-validation results for the selection of c in the threshold τ_n (see Subsection 4.1) and compares the proposed method with the PC-stable algorithm based on mutual information. We extract the skeleton by symmetrizing the adjacency matrix returned by the PC-stable algorithm implemented in the `bnlearn` package in R ([14]). Reported results are based on $R = 50$ Monte Carlo repetitions. FOCI attains high TPR, particularly for large n, indicating strong recovery of true edges. This improvement is accompanied by an increase in FPR, especially in denser graph settings.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

Table 1. Skeleton recovery performance of binary FOCI under Erdős–Rényi graphs. The threshold is set to $\tau_n = c\sqrt{\log(p)/n}$.

p	n	edge_prob	c	TPR (SD)	FPR (SD)
Baseline threshold ($c = 1$)					
15	300	0.05	1	0.223 (0.222)	0 (0)
15	3000	0.05	1	0.970 (0.090)	0 (0)
15	300	0.15	1	0.164 (0.116)	0.0013 (0.011)
15	3000	0.15	1	0.859 (0.121)	0.0052 (0.022)
30	300	0.05	1	0.121 (0.078)	8.1×10^{-5} (0.0010)
30	3000	0.05	1	0.864 (0.100)	0.0011 (0.0068)
30	300	0.15	1	0.048 (0.035)	0.0021 (0.006)
30	3000	0.15	1	0.430 (0.094)	0.0136 (0.024)
Effect of threshold scaling					
15	300	0.05	0.5	0.729 (0.237)	0.023 (0.054)
15	3000	0.05	0.5	0.996 (0.058)	3.4×10^{-5} (0.0006)
15	300	0.05	2	0.0012 (0.015)	0 (0)
15	3000	0.05	2	0.590 (0.256)	0 (0)
15	300	0.15	0.5	0.667 (0.164)	0.169 (0.135)
15	3000	0.15	0.5	0.996 (0.022)	0.075 (0.089)
15	300	0.15	2	0.0030 (0.014)	0 (0)
15	3000	0.15	2	0.362 (0.139)	0 (0)
30	300	0.05	0.5	0.609 (0.133)	0.035 (0.030)
30	3000	0.05	0.5	0.994 (0.024)	0.0116 (0.022)
30	300	0.05	2	0.00055 (0.0048)	0 (0)
30	3000	0.05	2	0.329 (0.115)	0 (0)
30	300	0.15	0.5	0.396 (0.097)	0.144 (0.065)
30	3000	0.15	0.5	0.810 (0.104)	0.305 (0.106)
30	300	0.15	2	0.00043 (0.0027)	6.4×10^{-5} (0.0006)
30	3000	0.15	2	0.099 (0.043)	0 (0)

Table 2. Skeleton recovery performance under Erdős–Rényi graphs for $p = 15$. Panel (A): original FOCI ($c = 0$). Panel (B): FOCI with 5-fold CV and PC-stable.

(A) Original FOCI ($c = 0$)				
n	edge_prob	Method	TPR (SD)	FPR (SD)
300	0.05	FOCI ($c = 0$)	0.9997 (0.0048)	0.9813 (0.0110)
3000	0.05	FOCI ($c = 0$)	1.0000 (0.0000)	1.0000 (0.0000)
300	0.15	FOCI ($c = 0$)	0.9913 (0.0229)	0.9807 (0.0137)
3000	0.15	FOCI ($c = 0$)	1.0000 (0.0000)	1.0000 (0.0000)
(B) FOCI (CV) vs PC-stable				
n	edge_prob	Method	TPR (SD)	FPR (SD)
300	0.05	FOCI (CV)	0.310 (0.267)	0.0116 (0.0421)
		PC	0.518 (0.210)	0.0020 (0.0041)
3000	0.05	FOCI (CV)	0.993 (0.033)	0.0060 (0.0241)
		PC	1.000 (0.000)	0.0046 (0.0073)
300	0.15	FOCI (CV)	0.479 (0.262)	0.179 (0.246)
		PC	0.406 (0.104)	0.0013 (0.0037)
3000	0.15	FOCI (CV)	0.975 (0.053)	0.174 (0.143)
		PC	0.958 (0.064)	0.0009 (0.0030)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

6 Discussion and Conclusion

In this work, we introduced a natural modification of the FOCI selection procedure for recovering the skeleton of an undirected dependency graph in the case of binary data. The proposed method is based on node-wise Markov blanket estimation and aggregates these local structures to obtain a global graph representation. We also established theoretical guarantees ensuring consistent graph recovery under suitable assumptions, and introduced a thresholding mechanism to stabilize the selection procedure and control false discoveries.

Although our primary focus was on binary random variables, the proposed framework is not restricted to this setting. In particular, the thresholded node-wise FOCI approach can be naturally extended to nonparametric graphical models with continuous distributions, similarly to the original formulation in [2]. Moreover, the oracle-type theoretical results derived in this paper are not limited to binary variables; rather, they clarify the general behavior of variable selection in the FOCI algorithm and provide insights that can be directly transferred to the problem of undirected random graph reconstruction.

While our experiments focus on controlled synthetic settings, the method is model-agnostic and applicable to real binary data. Extending empirical validation to real-world networks constitutes an important direction for future research.

The developed methodology also opens the door to structure learning for Bayesian networks. Analogously to the classical PC algorithm, the recovered skeleton may serve as a first step, followed by orientation of edges using standard rules such as Meek's orientation rules [12]. This suggests a promising direction for future research, where the full reconstruction of Bayesian networks using the proposed FOCI-based framework will be investigated for binary data. Importantly, the methodology can be extended in a straightforward manner to other types of random variables beyond the binary case.

A natural limitation of the proposed methodology is that the recovered structure is restricted to the undirected skeleton of the underlying graph. While this is sufficient for identifying conditional dependencies, it does not directly provide causal directions. Moreover, the performance of the thresholded FOCI procedure depends on the choice of the threshold parameter τ_n, which in practice may require calibration. Although we provide theoretical guidance of the form $\tau_n = c\sqrt{\log(p)/n}$, the optimal constant c may depend on the unknown data-generating mechanism. In Subsection 4.1, we propose a cross-validation procedure for selecting the parameter c and compare this approach with the PC-stable algorithm for skeleton recovery. The obtained results are promising, but a more comprehensive study of this choice will be provided in future work.

Finally, as with most nonparametric dependence-based methods, the procedure may suffer from reduced power in very high-dimensional settings with small sample sizes.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

7 Appendix

Proof of Lemma 1 See Theorem 2.1 [2]
Proof of Lemma 2 Since S does not block all information paths, at least one missing true variable still affects $\mathbb{E}[Y \mid X]$. Irrelevant variables are conditionally independent by the Markov blanket property. By Lemma 1, this translates exactly into positivity vs. zero of T. Also see condition (B0).
Proof of Lemma 3 Obvious.
Proof of Lemma 5 Fix a subset $S \subseteq \{1, \dots, p\}$ with $|S| \leq K$. For each configuration $x_S \in \{0,1\}^{|S|}$ define

$$p_{x_S} = \mathbb{P}(Y = 1 \mid X_S = x_S), \qquad \hat{p}_{x_S} = \mathbb{P}_n(Y = 1 \mid X_S = x_S).$$

Let

$$\pi_{x_S} = \mathbb{P}(X_S = x_S), \qquad \hat{\pi}_{x_S} = \mathbb{P}_n(X_S = x_S).$$

Then the population functional is

$$Q(Y, X_S) = \sum_{x_S} \pi_{x_S} (p_{x_S} - p)^2,$$

and its empirical version is

$$Q_n(Y, X_S) = \sum_{x_S} \hat{\pi}_{x_S} (\hat{p}_{x_S} - \hat{p})^2.$$

Step 1: Concentration of empirical probabilities
For each fixed (x_S, y) the indicator $1\{Y = y, X_S = x_S\}$
is Bernoulli with mean $\mathbb{P}(Y = y, X_S = x_S)$.
By Hoeffding's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\big(\big|\hat{\pi}_{x_S} - \pi_{x_S}\big| > \varepsilon\big) \leq 2e^{-2n\varepsilon^2},$$

and

$$\mathbb{P}(|\hat{p}_{x_S} - p_{x_S}| > \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

Similarly,

$$\mathbb{P}(|\hat{p} - p| > \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

Step 2: Uniform control over all cells of S
There are at most $2^{|S|} \leq 2^K$ configurations x_S. Applying a union bound gives

$$\mathbb{P}\left(\max_{x_S} \big\{|\hat{p}_{x_S} - p_{x_S}|, |\hat{\pi}_{x_S} - \pi_{x_S}|, |\hat{p} - p|\big\} > \varepsilon\right) \leq C_1 2^K e^{-2n\varepsilon^2}.$$

Step 3: Control of $Q_n(Y, X_S) - Q(Y, X_S)$
Since all probabilities lie in $[0, 1]$, the mapping

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

$$(\pi_{x_S}, p_{x_S}, p) \mapsto \sum_{x_S} \pi_{x_S}(p_{x_S} - p)^2$$

is Lipschitz. Hence there exists a constant $C > 0$ such that

$$\big|Q_n(Y, X_S) - Q(Y, X_S)\big| \le C\, 2^{|S|} \max_{x_S} \Big(|\hat{\pi}_{x_S} - \pi_{x_S}| + |\hat{p}_{x_S} - p_{x_S}| + |\hat{p} - p|\Big).$$

Therefore

$$\mathbb{P}(|Q_n(Y, X_S) - Q(Y, X_S)| > \delta/8) \le C_1 2^K e^{-C_3 n \delta^2}.$$

Step 4: Uniformity over all subsets

The number of subsets S with $|S| \le K$ is bounded by

$$\sum_{m=0}^{K} \binom{p}{m} \le C p^K.$$

Applying another union bound yields

$$\mathbb{P}(E'^c) \le L_1 p^{L_2} e^{-L_3 n},$$

for suitable constants $L_1, L_2, L_3 > 0$.

Proof of Theorem 3 Proof is parallel to Theorem 6.1 from [2].

Step 1: Correct recovery of each Markov blanket

Fix i. By Theorem 2(the binary version of Theorem 6.1), applied with target Y_i,

$$\mathbb{P}\Big(\widehat{S}_i = \mathrm{MB}(i)\Big) \;\ge\; 1 - L_1' p^{L_2'} e^{-L_3' n}.$$

Step 2: Uniform control over all nodes

Define the event $\mathcal{E} := \bigcap_{i=1}^{p} \left\{\widehat{S}_i = \mathrm{MB}(i)\right\}$.

By the union bound,

$$\mathbb{P}(\mathcal{E}^c) \le \sum_{i=1}^{p} \mathbb{P}\Big(\widehat{S}_i \ne \mathrm{MB}(i)\Big) \le p \cdot L_1' p^{L_2'} e^{-L_3' n}.$$

Absorbing constants,

$$\mathbb{P}(\mathcal{E}) \ge 1 - L_1 p^{L_2} e^{-L_3 n}.$$

Step 3: From local Markov blankets to the skeleton On the event $\mathcal{E}$, for every pair $i \ne j$,

$$j \in \widehat{S}_i \quad\Longleftrightarrow\quad j \in \mathrm{MB}(i) \quad\Longleftrightarrow\quad \{i, j\} \in E.$$

Therefore, $\widehat{E} = E \quad \text{on } \mathcal{E}$.

Step 4: Conclusion Combining Steps 2 and 3 yields

$$\mathbb{P}\Big(\widehat{E} = E\Big) \;\ge\; 1 - L_1 p^{L_2} e^{-L_3 n}. \qquad \square$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24

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

References

1. Azadkia M.: FOCI- R package (2020) https://cran.r-project.org/web/packages/FOCI/FOCI.pdf
2. Azadkia M., Chatterjee S.: A simple measure of conditional dependence. Ann. Stat. 49, 3070-3102 (2021)
3. Bedford T and Cooke R.M.: Vines-a new graphical model for dependent random variables. Ann. Stat. 30, 1031-1068 (2002)
4. Besag J., E.: Nearest-Neighbour Systems and the Auto-Logistic Model for Binary Data. J. R. Stat. B: Stat. Methodol, 34(1), 75–83 (1972)
5. Colombo D. and Maathuis, M.H.: Order-independent constraint based causal structure learning. Journal of Machine Learning Research, 15:3921–3962, (2014)
6. Drton M. and Perlman M.D.: Multiple Testing and Error Control in Gaussian Graphical Model Selection. Stat.Scien. 22, 430-449(2007)
7. Finegold, M. A, and Drton, M.: Robust graphical modeling with tdistributions. Proceedings of the Twenty-Fifth Conference on Uncertainty in Articial Intelligence., pp. 169-176 (2009)
8. Lauritzen, S. Graphical Models. Clarendon Press, Oxford (1996)
9. Liu, H, Lafferty J and Wasserman, L.: The nonparanormal: Semiparametric Estimation of High Dimensional Undirected Graphs. Journ. Machine Learn Res. 10, 2295-2328 (2009)
10. Maathuis, M., Drton, M., Lauritzen, S., Wainwright, M.: Handbook of Graphical Models (1st ed). Chapman and Hall/CRC, (2020)
11. Madigan D., York J., Allard D.: Bayesian graphical models for discrete data. Int Stat Rev 63(2), 215–232 (1995)
12. Meek, C.: Causal inference and causal explanation with background knowledge. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI), pp. 403–410 (1995)
13. Rossell D and Zwiernik P.: Dependence in Elliptical Partial Correlation Graphs. Electron. J. Statist. 15(2): 4236-4263 (2021)
14. Scutari M . Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, 35(3), 1–22 (2010) doi:10.18637/jss.v035.i03.
15. Vogel, D., and Fried, R.: 2011. Elliptical graphical modelling. Biometrika, 98(4), 935-951 (2011)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_24