Lecture 13: High Dimensional Gaussian Graphical Models

October 4, 2024

Gaussian Graphical Model

Consider a set of Gaussian random variables, there are two measures for their dependency:

- Correlation Coefficient
- Partial Correlation Coefficient:

Compared to the partial correlation coefficient, the correlation coefficient is much weaker as marginally, i.e. directly or indirectly, all variables in a system are more or less correlated. The goal of GGM learning is to distinguish direct from indirect dependencies for all variables in a system.

The partial correlation coefficient provides a measure for direct dependence as it will vanish for the indirect case.

Applications

- Gene regulatory networks: molecular mechanism underlying cancer with data available at The Cancer Genome Atlas (TCGA).
- Causal networks: time course data.

Covariance Selection

let $\boldsymbol{X} = (X^{(1)}, \ldots X^{(p)})$ denote a p-dimensional random vector drawn from a multivariate Gaussian distribution $N_p(\boldsymbol{\mu}, \Sigma)$. Denote the concentration matrix by $\boldsymbol{C} = \Sigma^{-1} = (C_{ij})$. Then

$$\rho_{ij|\boldsymbol{V}\setminus\{i,j\}} = -\frac{C_{i,j}}{\sqrt{C_{i,i}C_{j,j}}}, \quad i,j = 1,\ldots,p. \tag{1}$$

The covariance selection method (Dempster, 1972; Lauritzen, 1996) is to identify the non-zero elements in the concentration matrix. However, it is not feasible for high dimensional problems with $p > n$.

Limited order partial correlations

The idea is to use low-order partial correlation as a surrogate of the full-order partial correlation, and it is widely acknowledged that the limited order partial correlation methods can result in something inbetween the full GGM (with correlations conditioned on all $p - 2$ variables) and the correlation graph.

Related work: Spirtes et al. (2000), Magwene and Kim (2004), Wille and Bühlmann (2006), Castelo and Roverato (2006, 2009).

Nodewise Regression

It is to use Lasso (Tibshirani, 1996) as a variable selection method to identify the neighborhood of each variable, and thus the nonzero elements of the concentration matrix.
Consider a linear regression

$$X^{(j)} = \beta_i^{(j)} X^{(i)} + \sum_{r \in \boldsymbol{V} \setminus \{j,i\}} \beta_r^{(j)} X^{(r)} + \epsilon^{(j)}, \tag{2}$$

where $\epsilon^{(j)}$ is a zero-mean Gaussian random error. Let $S^{(j)} = \{r : \beta_r^{(j)} \neq 0\}$ denote the set of explanatory variables identified by Lasso for $X^{(j)}$. The GGM can then be constructed using the "or"-rule: *estimate an edge between vertices i and j* $\Longleftrightarrow i \in S^{(j)}$ or $j \in S^{(i)}$, or the "and"-rule: *estimate an edge between vertices i and j* $\Longleftrightarrow i \in S^{(j)}$ and $j \in S^{(i)}$.

Graphical Lasso

Yuan and Lin (2007) proposed to directly estimate the concentration matrix by minimizing

$$-\log(\det(\boldsymbol{C})) + \text{trace}(\widehat{\Sigma}_{MLE}\boldsymbol{C}) + \lambda\|\boldsymbol{C}\|, \tag{3}$$

where $\widehat{\Sigma}_{MLE}$ denotes the maximum likelihood estimator of Σ, $\|\boldsymbol{C}\|$ denotes the norm of $\boldsymbol{C}$, and λ is the regularization parameter. Later, the algorithm is accelerated by Friedman *et al.* (2008) and Banerjee *et al.* (2008).

Graph Theory

An undirected graph is a pair $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V}$ is the set of vertices and $\boldsymbol{E} = (e_{ij})$ is the adjacency matrix.

- If two vertices $i, j \in \boldsymbol{V}$ forms an edge then we say that i and j are *adjacent* and set $e_{ij} = 1$.
- The *boundary set* of a vertex $v \in \boldsymbol{V}$, denoted by $b_{\boldsymbol{G}}(v)$, is the set of vertices adjacent to v, i.e., $b_{\boldsymbol{G}}(v) = \{j : e_{vj} = 1\}$.
- A *path* of length $l > 0$ from v_0 to v_l is a sequence $v_0, v_1, \ldots, v_l$ of distinct vertices such that $e_{v_{k-1}, v_k} = 1$ for all $k = 1, \ldots, l$.
- The subset $\boldsymbol{U} \subset \boldsymbol{V}$ is said to *separate* $\boldsymbol{I} \subset \boldsymbol{V}$ from $\boldsymbol{J} \subset \boldsymbol{V}$ if for every $i \in \boldsymbol{I}$ and $j \in \boldsymbol{J}$, all paths from i to j have at least one vertex in $\boldsymbol{U}$.
- For a pair of vertices $i \neq j$ with $e_{ij} = 0$, a set $\boldsymbol{U} \subset \boldsymbol{V}$ is called a $\{i, j\}$-*separator* if it separates $\{i\}$ and $\{j\}$ in $\boldsymbol{G}$.
- Let $\boldsymbol{G}_{ij}$ be a reduced graph of $\boldsymbol{G}$ with e_{ij} being set to zero. Then both the boundary sets $b_{\boldsymbol{G}_{ij}}(i)$ and $b_{\boldsymbol{G}_{ij}}(j)$ are $\{i, j\}$-separators in $\boldsymbol{G}_{ij}$.

Graph Theory (continuation)

Definition 1 We say that $P_{\boldsymbol{V}}$ satisfies the *Markov property* with respect to $\boldsymbol{G}$ if for every triple of disjoint sets $\boldsymbol{I}, \boldsymbol{J}, \boldsymbol{U} \subset \boldsymbol{V}$, it holds that $X_{\boldsymbol{I}} \perp X_{\boldsymbol{J}} | X_{\boldsymbol{U}}$ whenever $\boldsymbol{U}$ separates $\boldsymbol{I}$ and $\boldsymbol{J}$ in $\boldsymbol{G}$.

Definition 2
We say that $P_{\boldsymbol{V}}$ satisfies the adjacency faithfulness condition with respect to $\boldsymbol{G}$: If two variables $X^{(i)}$ and $X^{(j)}$ are adjacent in $\boldsymbol{G}$, then they are dependent conditioned on any subset of $X_{\boldsymbol{V} \setminus \{i,j\}}$.

Equivalent Measure

- Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the correlation graph of $X_1, \ldots, X_p$, where $\mathcal{E} = (\tilde{e}_{ij})$ is the edge set, and $\tilde{e}_{ij} = 1$ if $r_{ij} \neq 0$ and 0 otherwise.
- Let $\hat{\mathcal{E}}_{\gamma_i,i,-j} = \{v : |\hat{r}_{iv}| > \gamma_i\} \setminus \{j\}$ denote a reduced neighborhood of node i in $\mathcal{G}_{ij}$, where $\mathcal{G}_{ij}$ denotes a reduced graph of $\mathcal{G}$ with $\tilde{e}_{ij}$ being set to 0, and γ_i denotes a threshold value.
- $\hat{\mathcal{E}}_{\gamma_i,i} = \{v : |\hat{r}_{iv}| > \gamma_i\}$ as a reduced neighborhood of node i in $\mathcal{G}$.

For any pair of vertices i and j, we define the partial correlation coefficient ψ_{ij} by

$$\psi_{ij} = \rho_{ij|S_{ij}}, \tag{4}$$

where $S_{ij} = \hat{\mathcal{E}}_{\gamma_i,i,-j}$ if $|\hat{\mathcal{E}}_{\gamma_i,i,-j}| < |\hat{\mathcal{E}}_{\gamma_j,j,-i}|$ and $S_{ij} = \hat{\mathcal{E}}_{\gamma_i,i,-j}$ otherwise.

Equivalent Measure

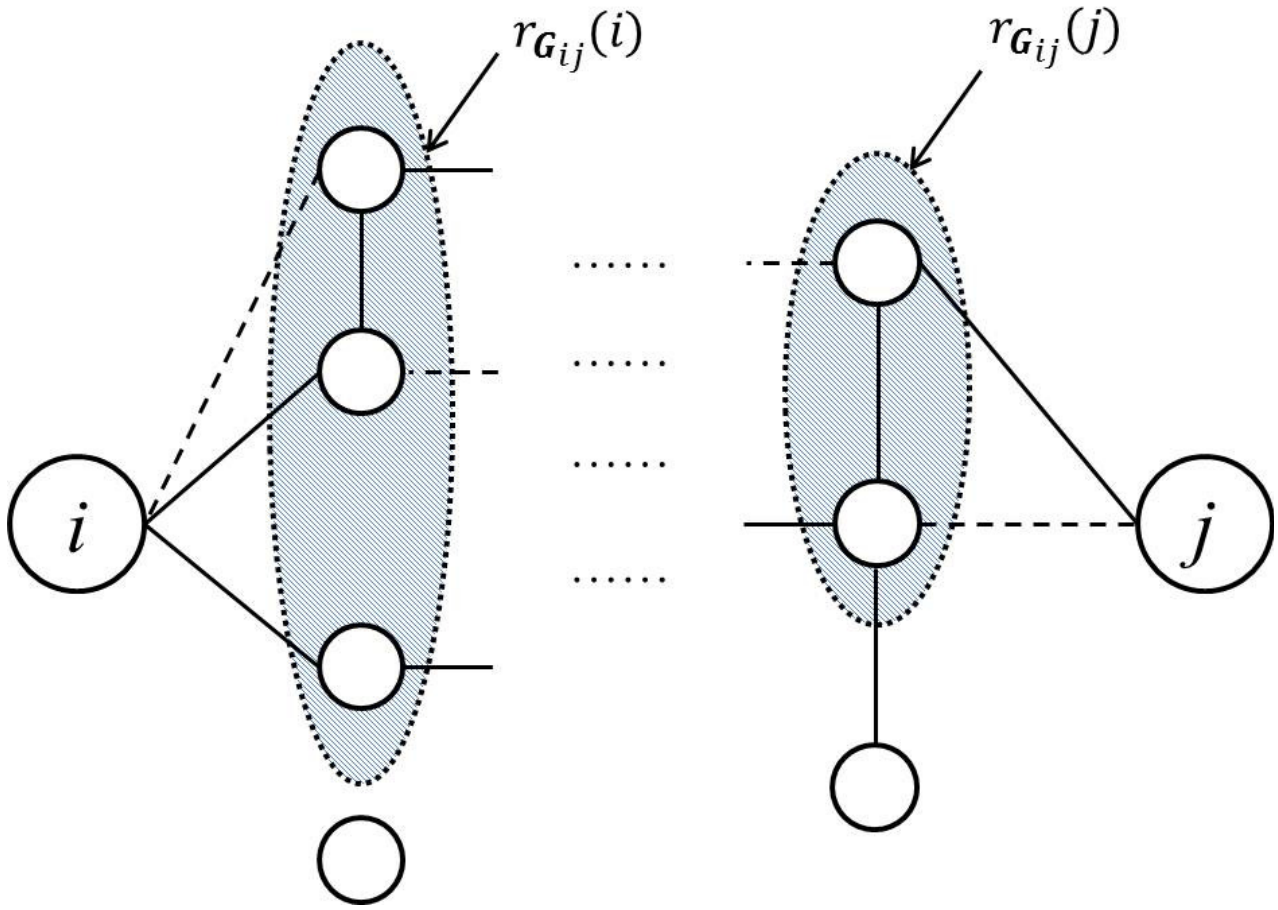

Figure: Illustrative plot for calculation of ψ-partial correlation coefficients, where the solid and dotted edges indicate the direct and indirect associations, respectively.

Equivalent Measure

Theorem 1 Suppose that a GGM $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ satisfies the assumption of faithfulness and $b_{\boldsymbol{G}}(i) \subseteq \hat{\mathcal{E}}_{\gamma_i, i}$ is true for each node i. Then ψ_{ij} defined in (4) is an equivalent measure of the partial correlation coefficient $\rho_{ij|\boldsymbol{V}\setminus\{i,j\}}$ in the sense that

$$\psi_{ij} = 0 \Longleftrightarrow \rho_{ij|\boldsymbol{V}\setminus\{i,j\}} = 0.$$

Equivalent Measure

Remarks:

1. There are many examples where the faithfulness assumption is violated, see Bühlmann and van de Geer (2011, pp.446-448) for some examples. However, as pointed out in Koller and Friedman (2009, p.1041), these examples usually correspond to particular parameter values, which are a set of measure zero within the space of all possible parameterizations.
2. There are many ways to specify the separator S_{ij}. For example, we can set S_{ij} to be $\hat{\mathcal{E}}_{\gamma_i,i,-j}$ or $\hat{\mathcal{E}}_{\gamma_j,j,-i}$ for which the cardinality is larger, or even set $S_{ij} = \hat{\mathcal{E}}_{\gamma_i,i,-j} \cup \hat{\mathcal{E}}_{\gamma_j,j,-i}$. Under these settings, ψ_{ij} is still an equivalent measure of $\rho_{ij|\boldsymbol{V}\setminus\{i,j\}}$. However, since ψ_{ij} has an approximate variance of $1/(n - |S_{ij}| - 3)$, we prefer to set S_{ij} to $\hat{\mathcal{E}}_{\gamma_i,i,-j}$ or $\hat{\mathcal{E}}_{\gamma_j,j,-i}$ for which the cardinality is smaller.

ψ-Learning Algorithm

(a) (*Correlation screening*) Determine the reduced neighborhood $\hat{\mathcal{E}}_{\gamma_i,i}$ for each variable $X^{(i)}$:

  (i) Conduct a multiple hypothesis test to identify the pairs of vertices for which the empirical correlation coefficient is significantly different from zero. This step results in a so-called correlation network.

  (ii) For each variable $X^{(i)}$, identify its neighborhood in the correlation network, and reduce the size of the neighborhood to $O(n/\log(n))$ by removing the variables having lower correlation (in absolute value) with $X^{(i)}$.

(b) (ψ-*calculation*) For each pair of vertices i and j, identify the separator S_{ij} based on the correlation network resulted in step (a) and calculate ψ_{ij} by inverting the subsample covariance matrix indexed by the variables in $S_{ij} \cup \{i, j\}$.

(c) (ψ-*screening*) Conduct a multiple hypothesis test to identify the pairs of vertices for which ψ_{ij} is significantly different from zero, and set the corresponding elements of $\boldsymbol{E}$ to be 1.

Consistency

(A_1) The distribution $P^{(n)}$ satisfies the conditions:

(i) $P^{(n)}$ is multivariate Gaussian;

(ii) $P^{(n)}$ satisfies the Markov property and adjacency faithfulness condition with respect to the undirected graph $\boldsymbol{G}^{(n)}$ for all $n \in \mathbb{N}$.

(A_2) The dimension $p_n = O(\exp(n^{\delta}))$ for some $0 \leq \delta < 1$.

(A_3) The correlation satisfy

$$\min\{|r_{ij}|; r_{ij} \neq 0,\ i, j = 1, 2, \ldots, p_n,\ i \neq j\} \geq c_0 n^{-\kappa}, \quad (5)$$

for some constants $c_0 > 0$ and $0 < \kappa < (1 - \delta)/2$, and

$$\max\{|r_{ij}|; i, j = 1, \ldots, p_n, i \neq j\} \leq M_r < 1, \quad (6)$$

for some constant $0 < M_r < 1$.

Consistency

Define

$$\begin{aligned} \widetilde{\boldsymbol{E}}^{(n)} &= \{(i,j) : \rho_{ij|\boldsymbol{V}\setminus\{i,j\}} \neq 0,\ \ i,j = 1,\ldots,p_n\}, \\ \widetilde{\mathcal{E}}^{(n)} &= \{(i,j) : r_{ij} \neq 0,\ \ i,j = 1,\ldots,p_n\}, \end{aligned} \tag{7}$$

as the edge sets of $\boldsymbol{G}^{(n)}$ and $\mathcal{G}^{(n)}$, respectively.

Since $\widetilde{\boldsymbol{E}}^{(n)} \subset \widetilde{\mathcal{E}}^{(n)}$, there exist constants $c_1 > 0$ and $0 < \kappa' \leq \kappa$ such that

$$\min\{|r_{ij}|; (i,j) \in \widetilde{\boldsymbol{E}}^{(n)}, i,j = 1,\ldots,p_n\} \geq c_1 n^{-\kappa'}. \tag{8}$$

Consistency

Lemma 1 Assume (A_1), (A_2), and (A_3) hold. Let $\gamma_n = 2/3c_1 n^{-\kappa'}$. Then there exist constants c_2 and c_3 such that

$$P(\widetilde{\boldsymbol{E}}^{(n)} \subseteq \hat{\mathcal{E}}_{\gamma_n}) \geq 1 - c_2 \exp(-c_3 n^{1-2\kappa'}),$$

$$P(b_{\boldsymbol{G}^{(n)}}(i) \subseteq \hat{\mathcal{E}}_{\gamma_n,i}) \geq 1 - c_2 \exp(-c_3 n^{1-2\kappa'}).$$

Consistency

(A_4) There exist constants $c_4 > 0$ and $0 \leq \tau < 1 - 2\kappa'$ such that $\lambda_{\max}(\Sigma) \leq c_4 n^{\tau}$, where Σ denotes the covariance matrix of X_i, and $\lambda_{\max}(\Sigma)$ is the largest eigenvalue of Σ.

Lemma 2 Assume (A_1), (A_2), (A_3), and (A_4) hold. Let $\gamma_n = 2/3c_1 n^{-\kappa'}$. Then for each node i,

$$P\left[|\hat{\mathcal{E}}_{\gamma_n,i}| \leq O(n^{2\kappa'+\tau})\right] \geq 1 - c_2 \exp(-c_3 n^{1-2\kappa'}),$$

where c_2 and c_3 are as given in Lemma 1.

Consistency

Lemma 3 Assume (A_1)-(i), (A_2), and (A_3). If $\eta_n = 1/2c_0 n^{-\kappa}$, then

$$P[\hat{\mathcal{E}}_{\eta_n} = \widetilde{\mathcal{E}}^{(n)}] = 1 - o(1), \quad \text{as } n \to \infty.$$

Consistency

The faithfulness implies that $\widetilde{\boldsymbol{E}}^{(n)} \subseteq \widetilde{\mathcal{E}}^{(n)}$. Further, it follows from Lemma 3 that

$$P[\widetilde{\boldsymbol{E}}^{(n)} \subseteq \hat{\mathcal{E}}_{\eta_n}] = 1 - o(1). \tag{9}$$

Based on Lemma 1, Lemma 2 and (9), we propose to restrict the neighborhood size of each node in calculation of ψ-partial correlation coefficients to be

$$\min \left\{ |\hat{\mathcal{E}}_{\eta_n,i}|, \frac{n}{\xi_n \log(n)} \right\}, \tag{10}$$

where η_n can be determined through a multiple hypothesis test.

Consistency

Let ζ_n denote the threshold value of ψ-partial correlation used in the ψ-screening step, and let $\hat{\boldsymbol{E}}_{\zeta_n}$ denote the final network obtained through thresholding ψ-partial correlation. That is, we define

$$\hat{\boldsymbol{E}}_{\zeta_n} = \{(i,j) : |\hat{\psi}_{ij}| > \zeta_n,\ i,j = 1,2,\ldots,p_n\}.$$

(A_5) The ψ-partial correlation coefficients satisfy

$$\inf\{\psi_{ij}; \psi_{ij} \neq 0,\ i,j = 1,\ldots,p_n,\ i \neq j,\ |S_{ij}| \leq q_n\} \geq c_6 n^{-d},$$

where $q_n = O(n^{2\kappa'+\tau})$, $0 < c_6 < \infty$ and $0 < d < (1-\delta)/2$ are some constants. In addition,

$$\sup\{\psi_{ij}; i,j = 1,\ldots,p_n,\ i \neq j, |S_{ij}| \leq q_n\} \leq M_\psi < 1,$$

for some constant $0 < M_\psi < 1$.

Consistency

Theorem 2 Consider a GGM with distribution $P^{(n)}$ and underlying conditional independence graph $\boldsymbol{G}^{(n)}$. Assume (A_1)–(A_5) hold. Then

$$P[\hat{\boldsymbol{E}}_{\zeta_n} = \widetilde{\boldsymbol{E}}^{(n)}] \geq 1 - o(1), \quad \text{as } n \to \infty.$$

An Illustrative Example

Consider an auto-regressive process of order two with the concentration matrix given by

$$C_{ij} = \begin{cases} 0.5, & \text{if } |j-i| = 1,\ i = 2, \ldots, (p-1), \\ 0.25, & \text{if } |j-i| = 1,\ i = 3, \ldots, (p-2), \\ 1, & \text{if } i = j,\ i = 1, \ldots, p, \\ 0, & \text{otherwise.} \end{cases}$$

This example has been used by Yuan and Lin (2007) and Mazumder and Hastie (2012) to illustrate their gLasso algorithms. We are interested in it because all the variables in it are dependent, either directly or indirectly, and thus it can serve as a good example for testing whether or not the proposed ψ-learning algorithm can distinguish direct dependencies from indirect ones.

An Illustrative Example

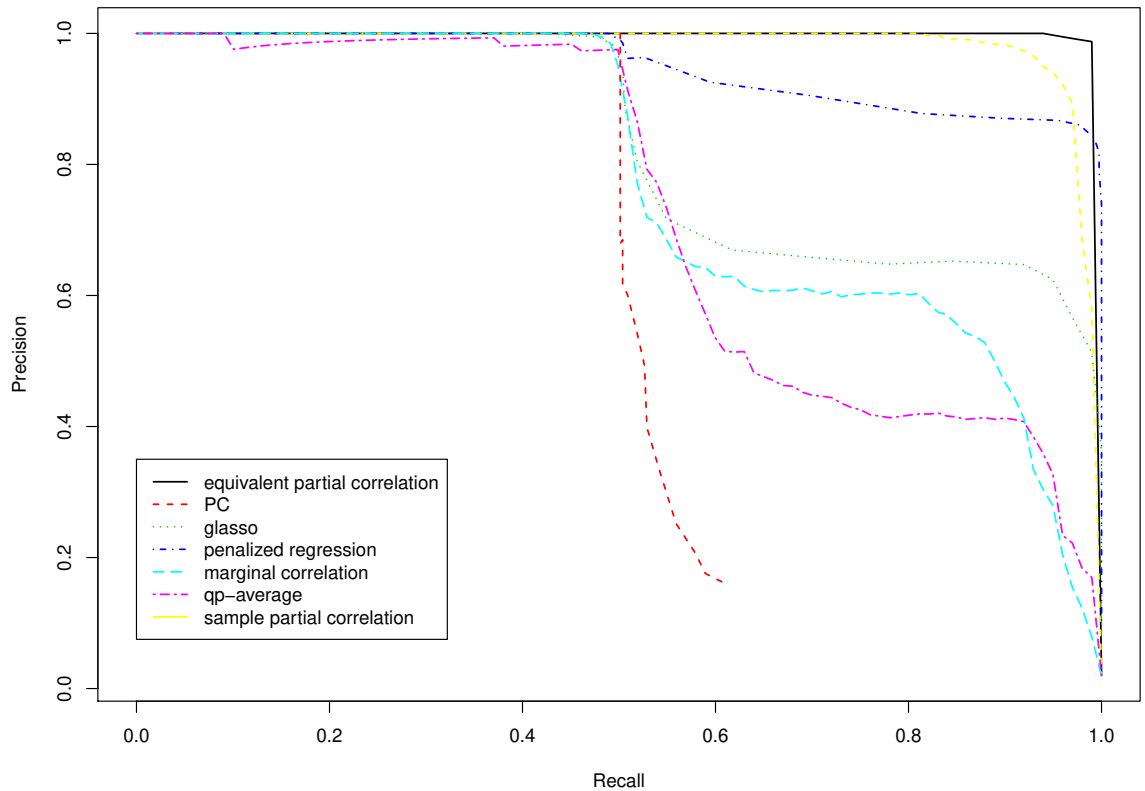

Figure: Precision-recall curves of the ψ-learning, gLasso, nodewise regression, *qp*-average, PC, partial correlation, and correlation methods for one dataset simulated with $(n, p) = (500, 200)$. Under this setting, the true partial correlation is available and thus included for comparison.

An Illustrative Example

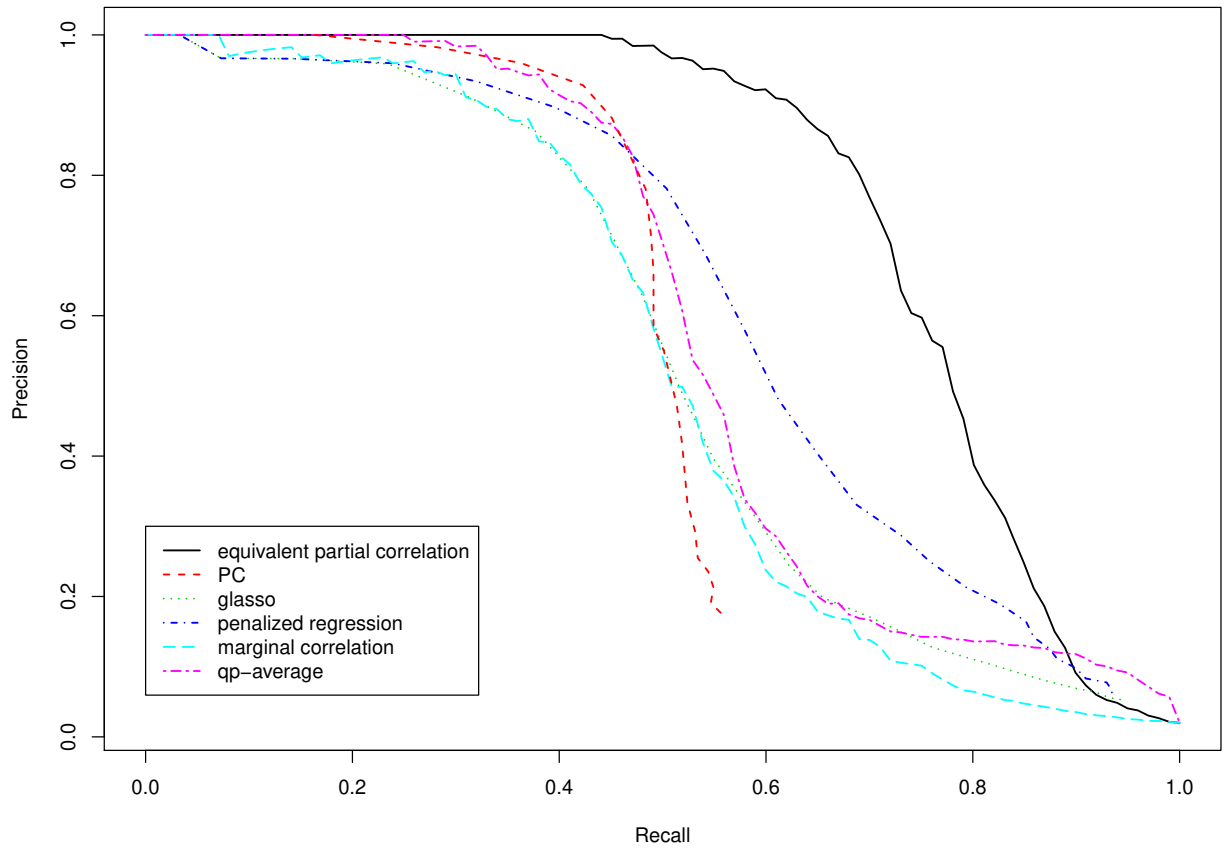

Figure: Precision-recall curves of the ψ-learning, gLasso, nodewise regression, *qp*-average, PC, partial correlation, and correlation methods for one dataset simulated with $(n, p) = (100, 200)$. Under this setting, the true partial correlation is not available.

Partial Correlation Scores

The true partial correlation score is defined by

$$z_{ij}^{*} = \frac{1}{2}\log\left[\frac{1+\rho_{ij|\boldsymbol{V}\setminus\{i,j\}}}{1-\rho_{ij|\boldsymbol{V}\setminus\{i,j\}}}\right], \quad \tilde{z}_{ij}^{*} = \Phi^{-1}\left(2\Phi(\sqrt{n-p-1}|z_{ij}^{*}|)-1\right), \tag{11}$$

where $n-p-1$ is its effective sample size.

The ψ-score is defined by

$$z'_{ij} = \frac{1}{2}\log\left[\frac{1+\psi_{ij}}{1-\psi_{ij}}\right], \quad \tilde{z}'_{ij} = \Phi^{-1}\left(2\Phi(\sqrt{n-|S_{ij}|-3}|z'_{ij}|)-1\right), \tag{12}$$

where S_{ij} denotes the conditional set, and $n-|S_{ij}|-3$ is its effective sample size.

An Illustrative Example

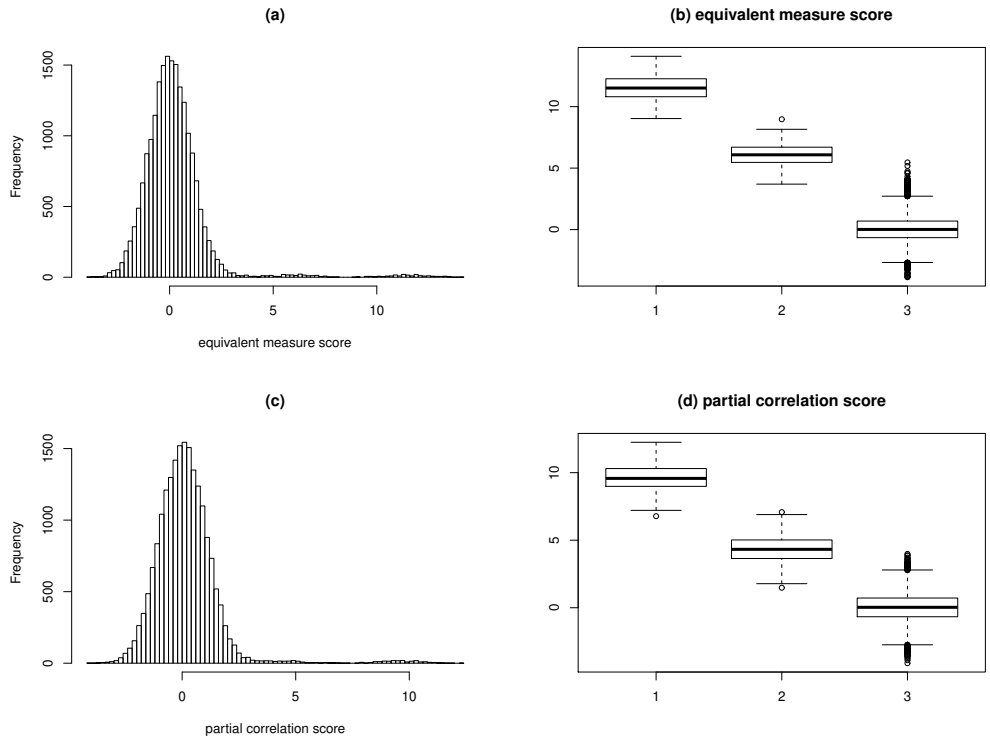

Figure: Comparison of ψ-scores and sample partial correlation scores. (a) Histogram of ψ-scores. (b) Boxplots of ψ-scores for the first-order dependent nodes (labeled by 1), the second-order dependent nodes (labeled by 2) and other nodes (labeled by 3). (c) Histogram of sample partial correlation scores. (d) Boxplots of sample partial correlation scores for the first-order dependent nodes (labeled by 1), the second-order dependent nodes (labeled by 2) and other nodes (labeled by 3).

An Illustrative Example

Table: Average areas under the Precision-Recall curves resulting from different methods: ψ-learning, true partial correlation, gLasso, nodewise regression, *qp*-average, PC, and correlation. The numbers in parentheses represent the standard deviation of the average area.

n	ψ	partial	gLasso	nodewise	*qp*-ave	PC	corr
500	0.9940	0.9831	0.8259	0.9466	0.7268	0.5285	0.7696
	(0.0002)	(0.0011)	(0.0010)	(0.0019)	(0.0025)	(0.0040)	(0.0017)
100	0.7925	—	0.5336	0.6207	0.5819	0.4945	0.5215
	(0.0086)	—	(0.0030)	(0.0024)	(0.0030)	(0.0026)	(0.0029)

T-cell Data

This dataset results from one experiment investigating the expression response of human T-cells to phorbol myristate acetate (PMA). It contains the temporal expression levels of 58 genes for 10 unequally spaced time points. At each time point there are 34 separate measurements. The data have been log-transformed and quantile normalized and are available at the R package *longitudinal* (Opgen-Rhein and Strimmer, 2008). As in other work, we ignore its longitudinal structure and treat the observations as independent in studying the GGM.

T-cell Data

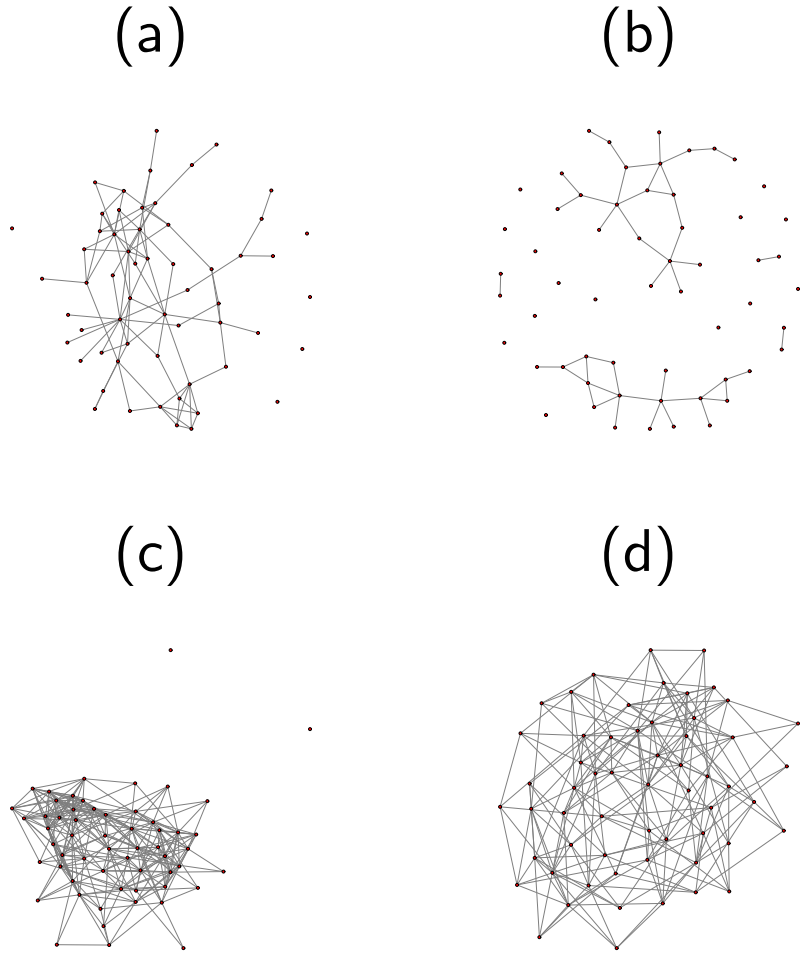

Figure: Networks identified by (a) the ψ-learning method with q-value=0.01, (b) the ψ-learning method with q-value=0.0001, (c) gLasso, and (d) nodewise regression for the T-cell data.

Phase Transition

The correlation and partial correlation screening suffers from a phase transition phenomenon: As the threshold decreases, the number of discoveries increases abruptly.
This explaines the ψ-learning method outperforms the other methods in the high-precision region, while this is reversed in the low-precision region.

T-cell Data: Power Law

To assess the quality of the networks resulted from different methods, we fit the power law to them. A nonnegative random variable X is said to have a power law distribution if

$$P(X = x) \propto x^{-\alpha},$$

for some positive constant α. The power law states that the majority of vertices are of very low degree, although some are of much higher degree–two or three orders of magnitude higher in some cases.

T-cell Data

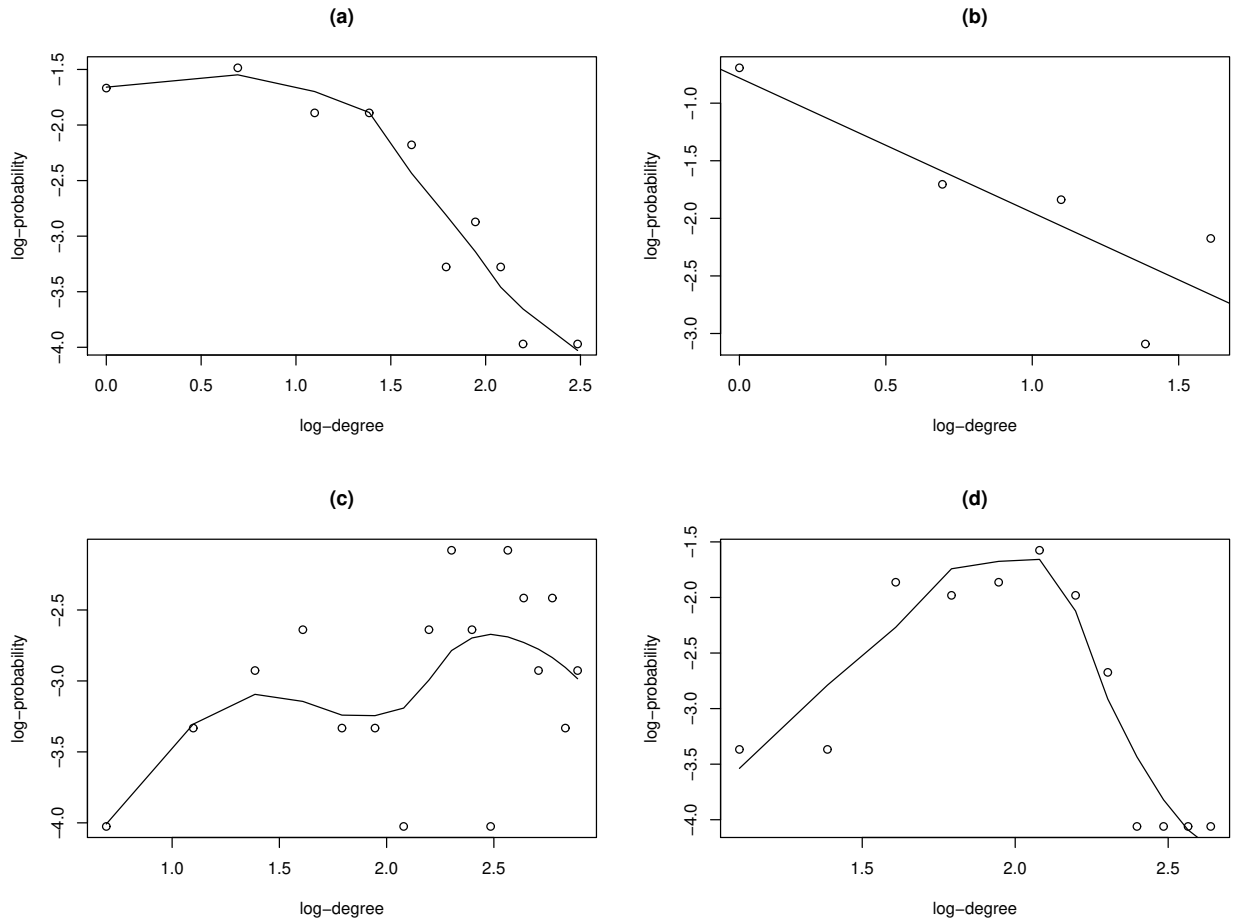

Figure: Log-log plots of the degree distributions of the four networks shown in Figure 5: (a) ψ-learning with q-value=0.01; (b) ψ-learning with q-value=0.0001; (c) gLasso; and (d) nodewise regression.

Breast Cancer Data

This dataset contains 49 breast cancer samples and 3883 genes arising in a study of molecular phenotypes for clinical prediction (West *et al.*, 2001). The cleaned data is available at `http://strimmerlab.org/ data.html`. The full graph consists of 7,536,903 edges.

Breast Cancer Data

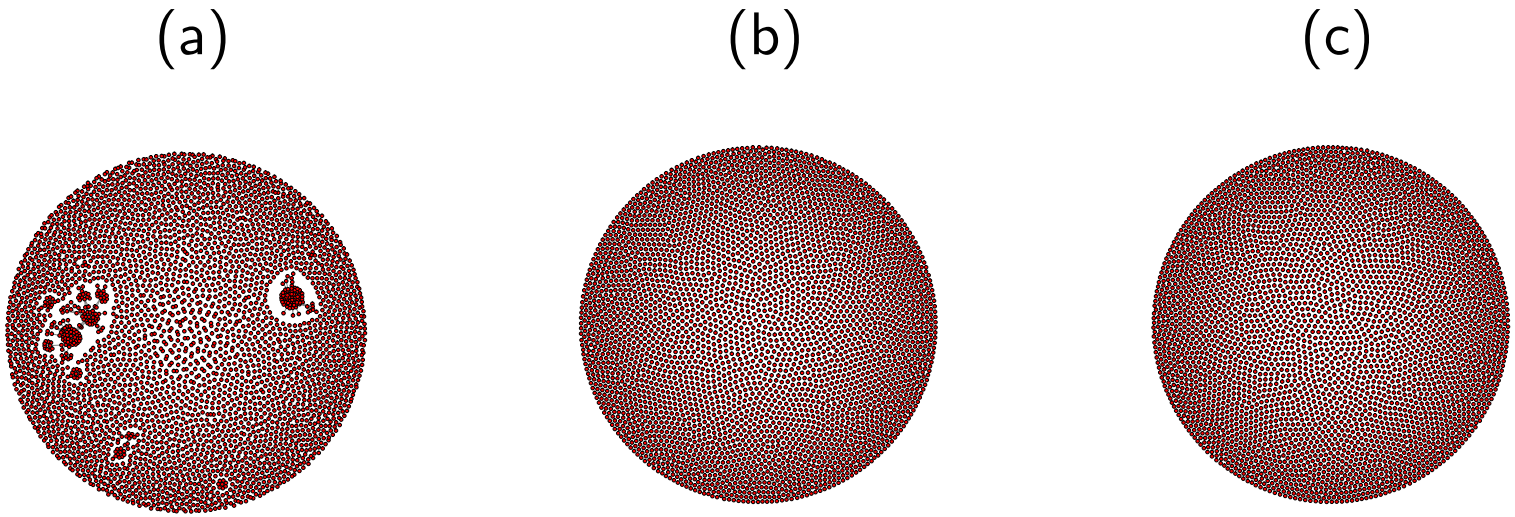

Figure: Networks identified by (a) the ψ-learning method with q-value=0.01, (b) gLasso, and (c) nodewise regression for the breast cancer data.

Breast Cancer Data

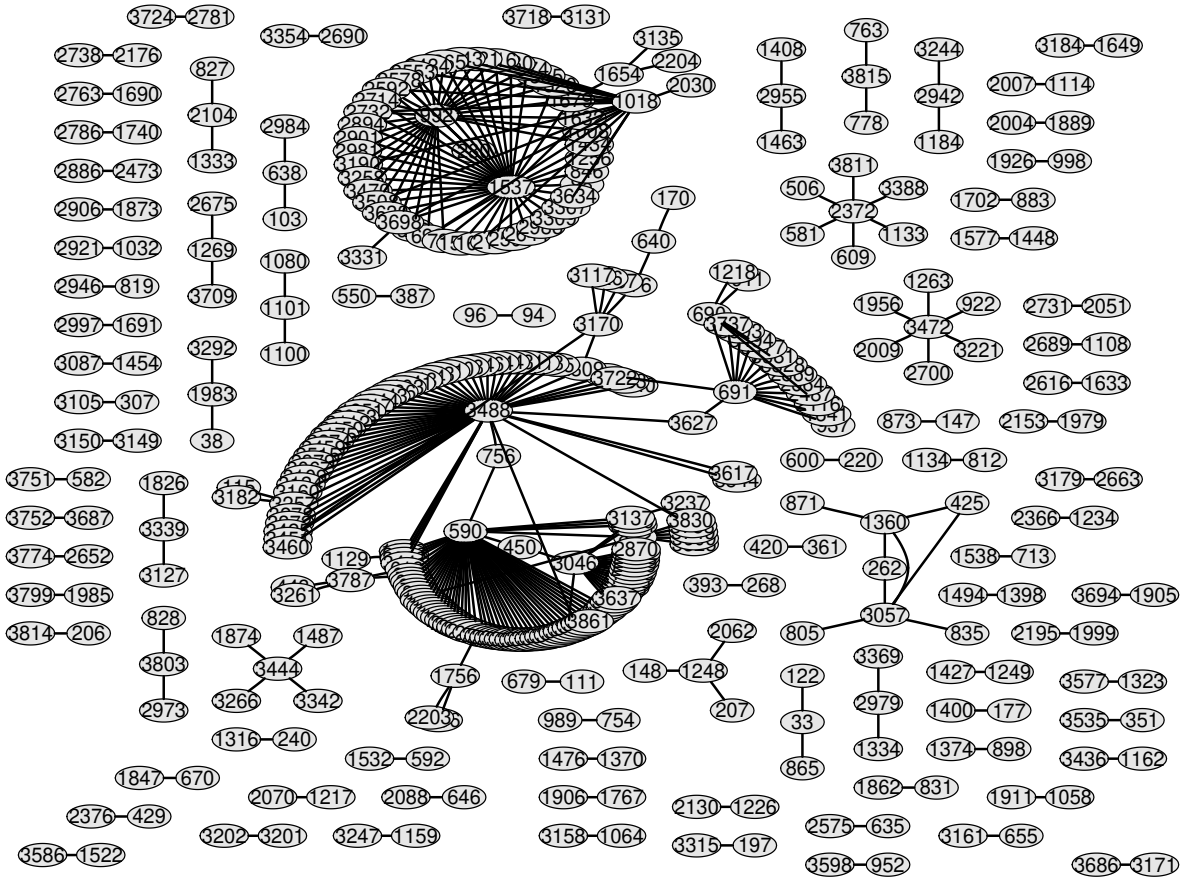

Figure: Edges of the network identified by the ψ-learning method for the breast cancer data, where the genes numbered by 590, 3488, 1537, and 992 are CD44, IGFBP-5, HLA, and STARD3, respectively.

Biological Results

Based on the result of ψ-learning, 13 hub genes are identified by setting the cutoff value of connectivity at 5, and the top 4 genes are STARD3, IGFBP-5, CD44 and HLA. Our examination shows very exciting results: All of the 13 genes are related to breast cancer.

- STARD3 (also named as MLN64 and CAB1) is in the 17q11-q21 region in which amplification is found in about 25% of primary breast carcinomas, and its gene expression is associated with poor clinical outcome, such as increase risk of relapse and poor prognosis (http://atlasgeneticsoncology.org/).
- IGFBP-5 has been shown to play an important role in the molecular mechanism of breast cancer, especially in metastasis (Akkiprik *et al.*, 2008), and it is considered to be a potential therapeutic target for breast cancer.
- CD44 is involved in many essential biological functions associated with the pathology activities of cancer cells, and CD44 expression is commonly used as a marker for breast cancer stem cells (Louderbough *et al.*, 2011).
- HLA expression is associated with the genetic susceptibility (Chaudhuri *et al.*, 2000), tumor progression and recurrent risk of breast cancer (Kaneko *et al.*, 2011).

Data Integration

To integrate multiple sources of data, the ψ-partial correlation coefficient can be transformed to a Z-score via Fisher's transformation:

$$\psi_{z_{ij}} = \frac{\sqrt{n - |S_{ij}| - 3}}{2} \log \left[\frac{1 + \hat{\psi}_{ij}}{1 - \hat{\psi}_{ij}} \right], \quad i, j = 1, 2, \ldots, p. \tag{13}$$

Then ψ_z-scores from different sources of data can be combined using the meta-analysis method:

$$\psi_{c_{ij}} = \frac{\sum_{k=1}^{K} w_k \psi_{z_{ij}}^{(k)}}{\sqrt{\sum_{k=1}^{K} w_k^2}}, \quad i, j = 1, 2, \ldots, p, \tag{14}$$

where K denotes the total number of data sources, $\psi_{z_{ij}}^{(k)}$ denotes the ψ_z-score from source k, and w_k denotes the weight assigned on source k.

Network Comparison

$$\psi_{d_{ij}} = [\psi^{(1)}_{z_{ij}} - \psi^{(2)}_{z_{ij}}]/\sqrt{2}, \tag{15}$$

(i) (ψ-correlation calculation) Perform steps (a) and (b) of the ψ-learning algorithm independently for each source of data.

(ii) (ψ_d-score calculation) Calculate the difference of ψ-scores in in (15).

(iii) (ψ_d-*score screening*) Conduct a multiple hypothesis test to identify the pairs of vertices for which $\psi_{d_{ij}}$ is differentially distributed from the standard normal $N(0, 1)$.

An Illustrative Example (continuation)

Generate two datasets from $N(0, \frac{1}{2}C^{-1})$ and $N(0, C^{-1})$, respectively, where C is as early specified with $p = 200$. Both datasets have a sample size of $n = 100$.

An Illustrative Example

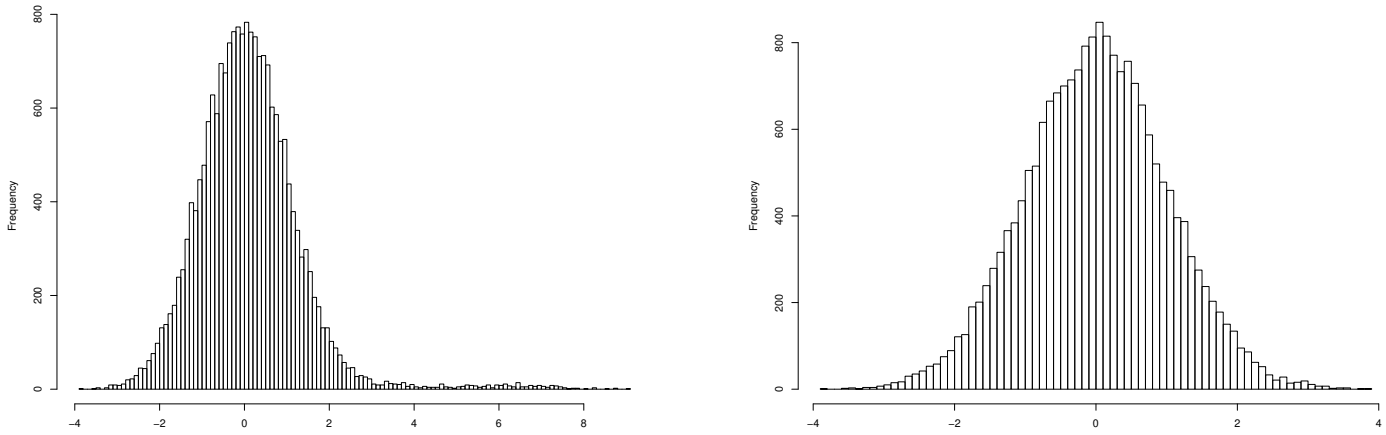

Figure: Histograms of $\psi'_{c_{ij}}$ (left) and $\psi'_{d_{ij}}$ (right) for the two simulated datasets.

An Illustrative Example

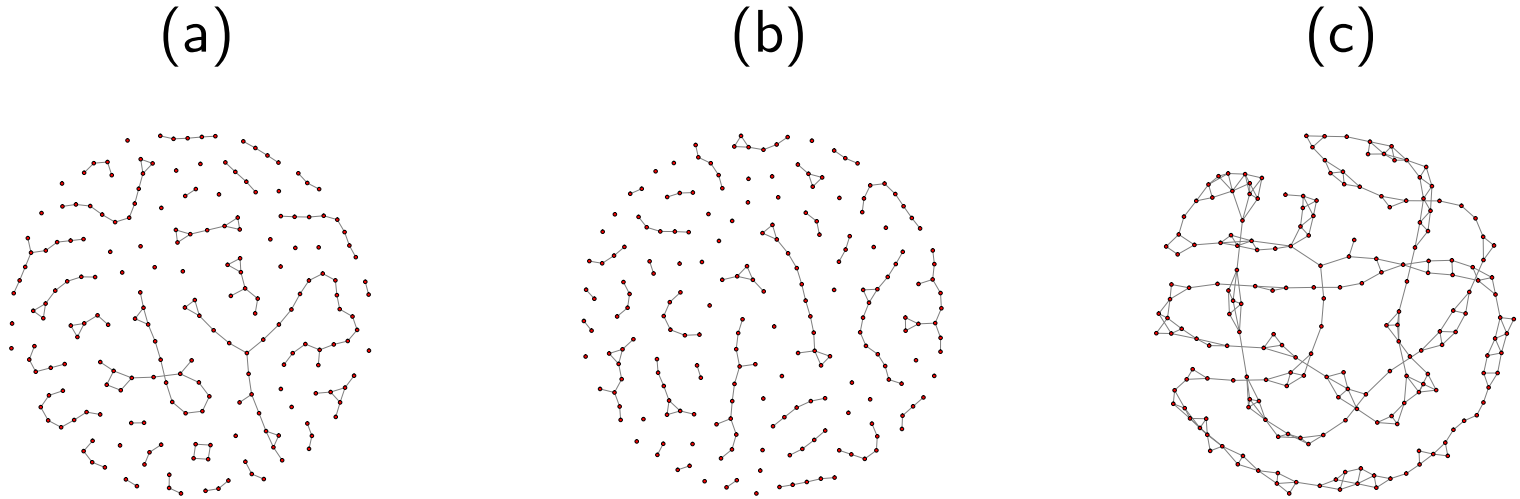

Figure: Networks identified by the ψ-learning algorithm for the simulation experiment: (a) network for dataset 1 (precision,recall)=(0.963,0.388), (b) network for dataset 2 (precision,recall)=(0.958,0.343), and (c) integrated network of dataset 1 and dataset 2 (precision,recall)=(0.934,0.602).

Covariate Adjustment

- Let $W_1, \ldots, W_q$ denote the external variables. To adjust their effects, we can replace the the empirical correlation coefficient used in step (a) of the ψ-learning algorithm by the p-value obtained in testing the hypotheses $H_0 : \beta_{q+1} = 0 \Leftrightarrow H_1 : \beta_{q+1} \neq 0$ for the regression

$$X^{(i)} = \beta_0 + \beta_1 W_1 + \cdots + \beta_q W_q + \beta_{q+1} X^{(j)} + \epsilon, \qquad (16)$$

where ϵ denotes a vector of Gaussian random errors.

- Similarly, we replace the ψ-correlation coefficient used in step (c) of the ψ-learning algorithm by p-value obtained in testing the hypotheses $H_0 : \beta_{q+1} = 0 \Leftrightarrow H_1 : \beta_{q+1} \neq 0$ for the regression

$$X^{(i)} = \beta_0 + \beta_1 W_1 + \cdots + \beta_q W_q + \beta_{q+1} X^{(j)} + \sum_{k \in S_{ij}} \gamma_k X^{(k)} + \epsilon, \qquad (17)$$

where S_{ij}, as defined previously, denotes the selected separator of $X^{(i)}$ and $X^{(j)}$.

Time Complexity

In terms of p, the computational complexity of the ψ-learning algorithm is bounded by $O(p^2(\log p)^b)$, where $b = 3(2\kappa' + \tau)/\delta$, $O(p^2)$ is for the total number of ψ-scores that need to calculate, and $O((\log p)^b)$ is for the computational complexity of inverting a neighboring matrix of size $O(n^{2\kappa'+\tau})$.

When $\delta > 0$, the computational complexity of the algorithm is nearly $O(p^2)$.

Time Complexity

- The **gLasso** algorithm is known to have a computational complexity of $O(p^3)$. With its fast implementation (Witten *et al.*, 2011; Mazumder and Hastie, 2012), which makes use of the block diagonal structure in graphical lasso solutions, the computational complexity can be reduced to $O(p^{2+\nu})$, where $0 < \nu \leq 1$ may depend on the number of blocks and the size of each block.
- Since the ordinary Lasso has a computational complexity of $O(np\min(n,p))$ (Meinshausen, 2007), the **nodewise regression** algorithm has a complexity of $O(p^3(\log p)^{2/\delta})$ if $p > n$ and $O(p^4(\log p)^{1/\delta})$ otherwise.
- Other algorithms usually have higher computational complexities. For example, the PC algorithm (Spirtes *et al.*, 2000) has a computational complexity bounded by $O(p^{2+m})$, where m is the maximum size of the neighborhoods.

Time Complexity

We have compared the CPU times of ψ-learning and gLasso (the fastest regularization method) for the breast cancer example. On a single core of Intel® Xeon® CPU E5-2690(2.90Ghz), ψ-learning took 61 minutes, and gLasso took 6,993 minutes.

Here gLasso was implemented in the package *huge* and run under its default setting with the regularization parameter being determined using the stability approach.

Discussion

- We have provided a general framework for inference of GGMs based on the equivalent measure of partial correlation coefficient.
- The key idea is *correlation screening* and the key strategy is *split-and-merge*.
- Extension to Multiple Datasets Integration, covariate adjustment, network comparison, and non-normality data are all simple under the framework.