Differential Network

High-dimensional differential networks with sparsity and reduced-rank

Mose Park

Department of Statistical Data Science
University of Seoul

Selective. Lab

July 9, 2024

Index

1 Remind

2 Introduction

3 Method

4 Simulation

Overview

Graphical Model

Optimization

Sparsity
Reduced-rank

1

Remind

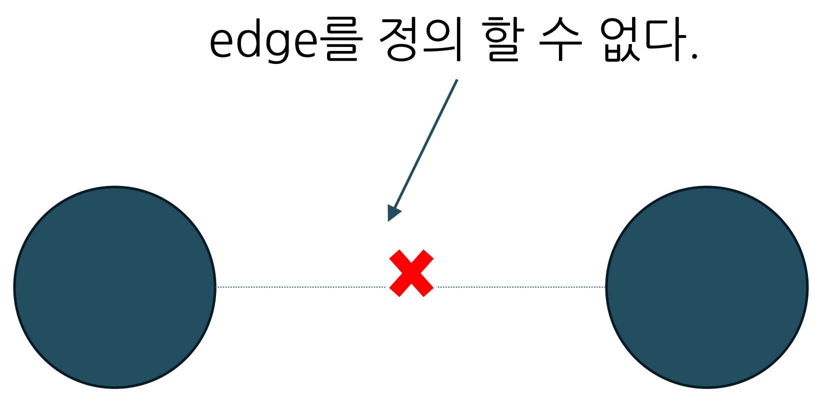
조건부 독립의 의미

$$P(A \cap B|C) = P(A|C)P(B|C)$$

C가 주어졌을 때 A,B 결합확률

$$P(A|\mathrm{B},C) = P(A|C)$$

정보 B와는 상관없음



정밀 행렬

➢ Graphical Lasso 발표에서,

$$\boldsymbol{\Theta} = \begin{pmatrix} \ddots & \cdots & \color{red}{0} \\ \vdots & \ddots & \vdots \\ \color{red}{0} & \cdots & \ddots \end{pmatrix}$$

✓ X_i 와 X_j에 해당하는 원소가 0

✓ 두 노드 사이의 $edge$는 없음

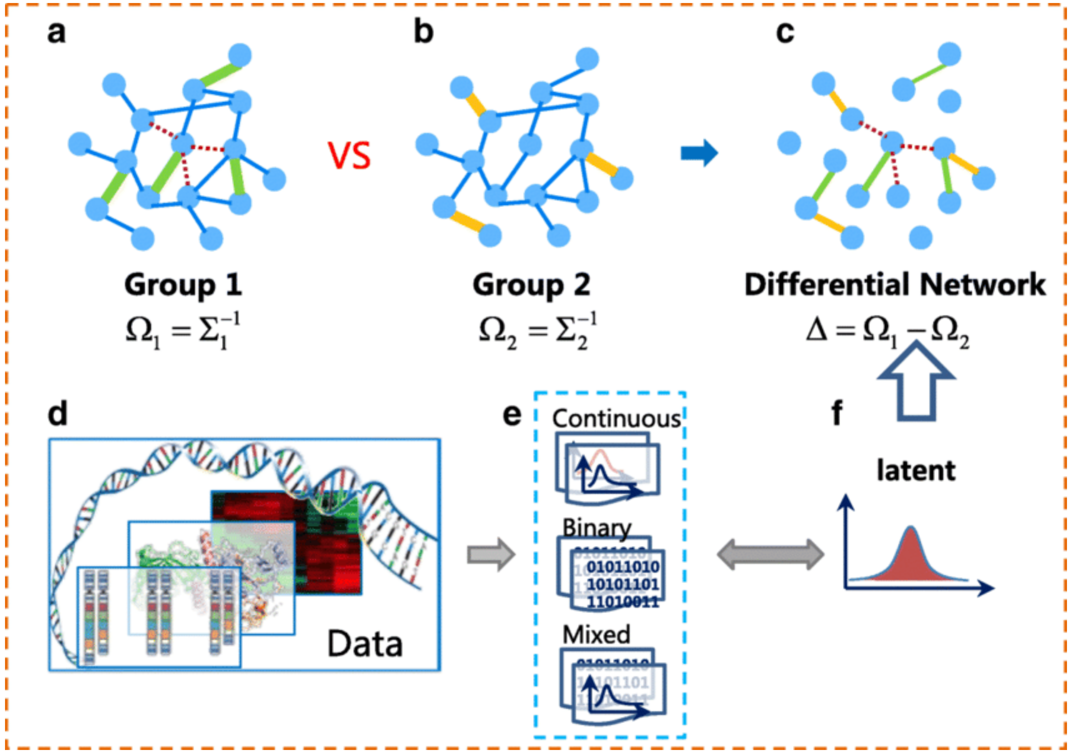
✓ X가 주어졌을 때, X_i 와 X_j 가 조건부 독립

- $\boldsymbol{\Theta} = \Sigma^{-\mathbf{1}}$: 공분산행렬의 역행렬 (정밀 행렬)
- 기본 모델 가정 : X ~ $\mathcal{N}(\mu, \Sigma)$: 다변량 정규분포
- 데이터셋 : n << p , 고차원 데이터

2

Introduction

모델소개



➢ Differential Network 이란?

- 각 그룹의 정밀 행렬에 기반하여 그룹 간 상호작용 차이 식별
- 복잡한 변수 관계 모델링 - 생물학, 금융 및 사회과학 등

➢ 선행 연구의 흐름

1. 정밀 행렬을 개별적으로 추정 → graphical lasso
2. Δ 를 직접 추정 → 추정 효율성과 수렴속도 고려하는 연구

- Δ 를 직접 추정하되, sparsity와 reduce-rank를 가정으로 최적화 문제 정의 및 해결

노테이션

➢ Vector Norms

실수 벡터 $\mathbf{a} = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$에 대해 다음과 같이 노름을 정의합니다:

- ℓ_0 노름:

$$\|\mathbf{a}\|_0 = \sum_{i=1}^{p} I(a_i \neq 0)$$

- ℓ_1 노름:

$$\|\mathbf{a}\|_1 = \sum_{i=1}^{p} |a_i|$$

- ℓ_2 노름:

$$\|\mathbf{a}\|_2 = \sqrt{\sum_{i=1}^{p} a_i^2}$$

- ℓ_∞ 노름:

$$\|\mathbf{a}\|_\infty = \max_{1 \le i \le p} |a_i|$$

➢ Matrix Norm

행렬 A의 특이값을 $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_{\min(p,q)}(A) \geq 0$로 나타내면, 다음과 같은 노름을 정의할 수 있습니다:

- 스펙트럴 노름:

$$\|A\| = \sigma_1(A) = \sup_{\|\mathbf{x}\|_2 \le 1} \|A\mathbf{x}\|_2$$

- 누클리어 노름:

$$\|A\|_* = \sum_{i=1}^{\min(p,q)} \sigma_i(A)$$

- 프로베니우스 노름:

$$\|A\|_F = \sqrt{\sum_{i=1}^{\min(p,q)} \sigma_i^2(A)} = \sqrt{\sum_{i,j} a_{ij}^2}$$

- 벡터 노름은 익숙한 형태, 원소 기반 행렬 노름은 벡터 노름으로 설명 대체
- 추가로 벡터화에 대한 설명이 있는데, 단순히 행렬의 원소를 하나의 열로 나열하는 것을 의미

- Kronecker product

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$$

$$A \otimes B = \begin{pmatrix} 1\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} & 4\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \\ 2\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} & 5\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \\ 3\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} & 6\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 & 0 \\ 2 & 3 & 8 & 12 \\ 2 & 0 & 5 & 0 \\ 4 & 6 & 10 & 15 \\ 3 & 0 & 6 & 0 \\ 6 & 9 & 12 & 18 \end{pmatrix}$$

✓ 두 행렬 A와 B의 모든 조합에 대해 곱

✓ 외적의 특수한 형태

- 기억해야할 부분은 벡터화는 크로네커 곱과 호환이 된다. (**호환성**) ⟶ $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$

아이디어의 관점

➢ 이론적 관점

$$\|\Delta^*\|_0 = o\left(\sqrt{\frac{n_1+n_2}{\log p}}\right)$$

* 일반적으로 sparisty의 조건이 위와 같음

$$o\left(\sqrt{\frac{n_1+n_2}{\log p}}\right)$$

$\text{rank}(\Delta^*) \leq \|\Delta^*\|_0$ 이므로 추정량의 rank가 sparity 조건보다 작아야 함

➢ 방법론적 관점

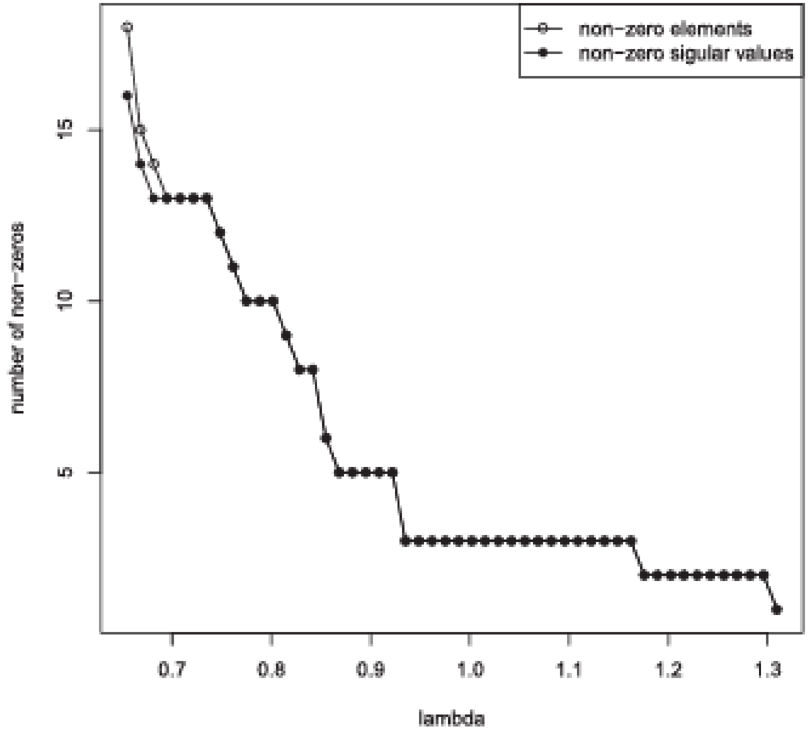
$$X^T\Delta^*X = \sum_{i=1}^{d} \lambda_i(u_i^TX)^2$$

* 행렬 분해를 통해 추정

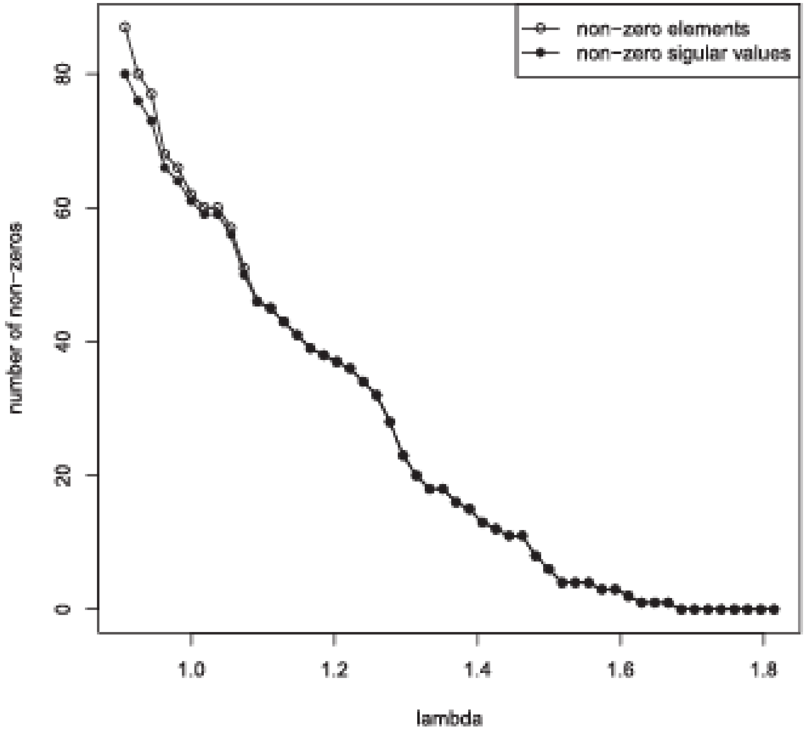
$\longrightarrow$ latent factor 식별

- n이 각각 100이고 p = 1000인 경우 0이 아닌 요소는 약 5.39보다 빠르게 증가해서는 안됨 $\longrightarrow$ 매우 sparse 함

$$\operatorname{rank}(\Delta^*) \leq \|\Delta^*\|_0$$



Spam data



Single-cell data

- 0이 아닌 요소와 특이값의 수 모두 sparse함에 따라 점차 감소하는 경향을 보임
- 단일 세포 데이터에서 더 많은 0이 아닌 요소가 있는 것으로 보아 데이터가 더 복잡하고 고차원적 특성을 가짐

3

Method

방법 1

➢ Based on Lasso $\longrightarrow$ $b = \mathrm{vec}(\Delta)$

$$\mathrm{argmin}_{b\in\mathbb{R}^{p^2}}\left(\frac{1}{2}b^T(\hat{\Sigma}_1\otimes\hat{\Sigma}_2)b - b^T\mathrm{vec}(\hat{\Sigma}_1-\hat{\Sigma}_2)+\lambda\|b\|_1\right)$$

우리가 직접 추정해야할 식 (1) 에서

$$\Delta^* = \Sigma_2^{-1} - \Sigma_1^{-1} = \Sigma_1^{-1}(\Sigma_1-\Sigma_2)\Sigma_2^{-1}$$

$$\longrightarrow \quad \mathrm{vec}(\Delta^*) = (\Sigma_2\otimes\Sigma_1)^{-1}\mathrm{vec}(\Sigma_1-\Sigma_2)$$

* 우리가 모르는 것

- 식 (1)에서 출발해 2차식 형태의 목적식을 다음과 같이 정의할 수 있다.
- 우리가 추정하는 것을 sparse하게 함 - Graph 구조 자체를 sparse하게 해서 유의미한 관계 식별

방법 2

참고 : "A direct approach for sparse quadratic discriminant analysis", Jiangetal. (2018)

- QDA

$$\hat{\Delta}_{\text{QDA}} = \underset{\Delta \in \mathbb{R}^{p \times p}}{\operatorname{argmin}} \left(\frac{1}{2}\operatorname{tr}(\Delta^T \hat{\Sigma}_1 \Delta \hat{\Sigma}_2) - \operatorname{tr}(\Delta \hat{\Sigma}_1 \hat{\Sigma}_2) + \lambda_1 \|\Delta\|_1 \right) \qquad (2)$$

우리가 직접 추정해야할 식 (1) 에서

* 교환법칙이 자유로움

$$\Sigma_2 \Delta \Sigma_1 = \Sigma_1 \Delta \Sigma_2 = \Sigma_1 - \Sigma_2 \longrightarrow \frac{1}{2}\operatorname{tr}(\Delta^T \Sigma_1 \Delta \Sigma_2) - \operatorname{tr}(\Delta(\Sigma_1 - \Sigma_2))$$

- 식 (1)에서 출발해 2차식 형태의 목적식을 trace 성질과 함께 정의할 수 있다.

방법 3

참고 : "Differential network analysis via lasso penalized d-trace loss", Yuanetal. (2017)

➢ Dtrace

$$\hat{\Delta}_{\text{Dtrace}} = \underset{\Delta \in \mathbb{R}^{p \times p}}{\operatorname{argmin}} \left(\frac{1}{4}\operatorname{tr}(\Delta^T \hat{\Sigma}_1 \Delta \hat{\Sigma}_2) + \boxed{\frac{1}{4}\operatorname{tr}(\Delta^T \hat{\Sigma}_2 \Delta \hat{\Sigma}_1)} - \operatorname{tr}(\Delta \hat{\Sigma}_1 \hat{\Sigma}_2) + \lambda \|\Delta\|_1 \right)$$

우리가 직접 추정해야할 식 (1) 에서 대칭성도 고려

$$\Sigma_2 \Delta \Sigma_1 = \Sigma_1 \Delta \Sigma_2 = \Sigma_1 - \Sigma_2$$

- 대칭성도 고려하여 문제를 좀 더 쪼개서 접근

제안

$$\hat{\Delta} = \underset{\Delta \in \mathbb{R}^{p \times p}}{\operatorname{argmin}} \left(\frac{1}{2} \operatorname{tr}(\Delta^T \hat{\Sigma}_1 \Delta \hat{\Sigma}_2) - \operatorname{tr}(\Delta \hat{\Sigma}_1 \hat{\Sigma}_2) + \lambda_1 \|\Delta\|_1 + \lambda_2 \|\Delta\|_* \right) \qquad (3)$$

$$\|\Delta\|_* = \sum_i \sigma_i(\Delta) \longrightarrow \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

0 이 된다면 rank는 감소

- nuclear norm의 경우 특이값의 합을 최소화 하려고 함
- 즉, 행렬의 rank를 감소시키는 효과 발생
- reduced 된 rank를 통해 저차원 구조로 해석력과 근본적인 구조를 볼 수 있다.

알고리즘

$$\begin{aligned} \min_{\Delta_1,\Delta_2,\Delta_3} \quad & \frac{1}{2}\mathrm{tr}(\Delta_1^T\hat{\Sigma}_1\Delta_1\hat{\Sigma}_2) - \mathrm{tr}(\Delta_1(\hat{\Sigma}_1\hat{\Sigma}_2)) + \lambda_1\|\Delta_2\|_1 + \lambda_2\|\Delta_3\|_* \\ \text{s.t.} \quad & \Delta_1 = \Delta_2 = \Delta_3 = \Delta \end{aligned} \tag{4}$$

위 선형 제약식을

$(\Delta_1, \Delta_2, \Delta_3)^T - (I_p, I_p, I_p)^T B = 0$ 로 표현 → 왜? 어떻게?

- 목적식과 선형 제약조건이 총 3가지로 복잡함 ⟶ 부분 문제로 나눠서 해결하고 갱신
- ADMM 알고리즘

* Scaled augmented Lagrangian

- 제약조건을 위반하기 어렵게 만드는 역할, ρ 는 step size

$$L(\cdot) = \frac{1}{2}\mathrm{tr}(\Delta_1^\top \hat{\Sigma}_1 \Delta_1 \hat{\Sigma}_2) + \lambda_1 \|\Delta_2\|_1 + \lambda_2 \|\Delta_3\|_* + \frac{\rho}{2}\sum_{j=1}^{3} \|\Delta_j - \Delta + A_j\|_2^2$$

➢ Step 1. $$(\Delta_1^{k+1}, \Delta_2^{k+1}, \Delta_3^{k+1}) = \arg\min_{\Delta_1,\Delta_2,\Delta_3} L(\Delta_1, \Delta_2, \Delta_3, \Delta^k, A_1^k, A_2^k, A_3^k)$$

➢ Step 2. $$\Delta^{k+1} = \arg\min_{\Delta} L(\Delta_1^k, \Delta_2^k, \Delta_3^k, \Delta, A_1^k, A_2^k, A_3^k) = \frac{1}{3}\sum_{j=1}^{3}(\Delta_j^{k+1} + A_j^{k+1})$$

➢ Step 3. $$(A_1^{k+1}, A_2^{k+1}, A_3^{k+1})^T = (A_1^k, A_2^k, A_3^k)^T + (\Delta_1^{k+1}, \Delta_2^{k+1}, \Delta_3^{k+1})^T - (I_p, I_p, I_p)^T \Delta^{k+1}$$

- 1단계는 주어진 라그랑지안 loss로 Δ 1,2,3 을 업데이트
- 2단계는 Δ 자체를 위의 값과 보조 변수(Auxiliary Variable) A를 활용해 업데이트 (평균으로 계산)
- 3단계는 보조변수 값들을 업데이트

참고 : "A direct approach for sparse quadratic discriminant analysis", Jiangetal. (2018)
Proposition 1

Sub problem

$$\Delta_1^{k+1} = \arg\min_{\Delta_1} \left\{ \frac{1}{2}\mathrm{tr}(\Delta_1^\top \hat{\Sigma}_1 \Delta_1 \hat{\Sigma}_2) - \mathrm{tr}(\Delta_1^\top \hat{\Sigma}_1 \hat{\Sigma}_2) + \frac{\rho}{2}\|\Delta_1 - \Delta^k + A_1^k\|_2^2 \right\}$$

$$\Delta_2^{k+1} = \arg\min_{\Delta_2} \{\lambda_1\|\Delta_2\|_1 + \tfrac{\rho}{2}\|\Delta_2 - \Delta^k + A_2^k\|_2^2\}$$

$$\Delta_3^{k+1} = \arg\min_{\Delta_3} \{\lambda_2\|\Delta_3\|_* + \tfrac{\rho}{2}\|\Delta_3 - \Delta^k + A_3^k\|_2^2\}$$

* Λ_i : 고유값 분해 대각행렬
* Λ = 두 그룹의 고유값의 역수

Solution

$$\Delta_1^{k+1} = B_1\left(\Lambda \circ \left(B_1^\top(\hat{\Sigma}_1 - \hat{\Sigma}_2 + \rho\Delta^k - \rho A_1^k)B_2\right)\right)B_2^\top$$

$$\Delta_2^{k+1} = \mathrm{soft}(\Delta^k - A_2^k, \frac{\lambda_1}{\rho})$$

$$\Delta_3^{k+1} = \sum_{i=1}^{p} \mathrm{soft}(d_i, \frac{\lambda_2}{\rho})u_i v_i^\top$$

→ soft threshold 연산을 통해 해를 구함 (norm 계산 때문)

- 참고 논문의 명제 1에 의해 목적식의 해가 증명이 되어있습니다.
- 풀이를 간단하게 설명드리면, 목적식의 도함수를 계산하고 벡터화 호환성과 SVD 사용해 계산할 수 있는 폼으로 유도

이론적 분석

* 정리 1 ⟶ 제안하는 추정의 정확도와 일관성을 보장

가정 (A1)부터 (A4)가 성립하고, λ_1과 λ_2가 다음을 만족한다고 가정합니다:

$$2c_1\sqrt{\frac{\log p}{n}} \le \lambda_1 \le 3c_1\sqrt{\frac{\log p}{n}}, \quad \lambda_2 \le c_2\lambda_1$$

여기서 c_1는 양의 상수이고, c_2는 충분히 작은 상수입니다. 그러면, 충분히 큰 상수 $C > 0$가 존재하여 다음을 만족합니다:

$$\|\hat{\Delta} - \Delta^*\|_2 \le C\sqrt{\frac{s\log p}{n}}, \quad \|\hat{\Delta} - \Delta^*\|_1 \le Cs\sqrt{\frac{\log p}{n}}$$

확률이 $1 - p^{-1}$보다 클 때.

- ✓ X_1과 Y_1 은 sub gausian random vector
- ✓ 두 가우시안 분포의 모멘트에 대한 정규화 조건
- ✓ Δ^*는 sparse하다.
- ✓ Δ^*는 low-rank

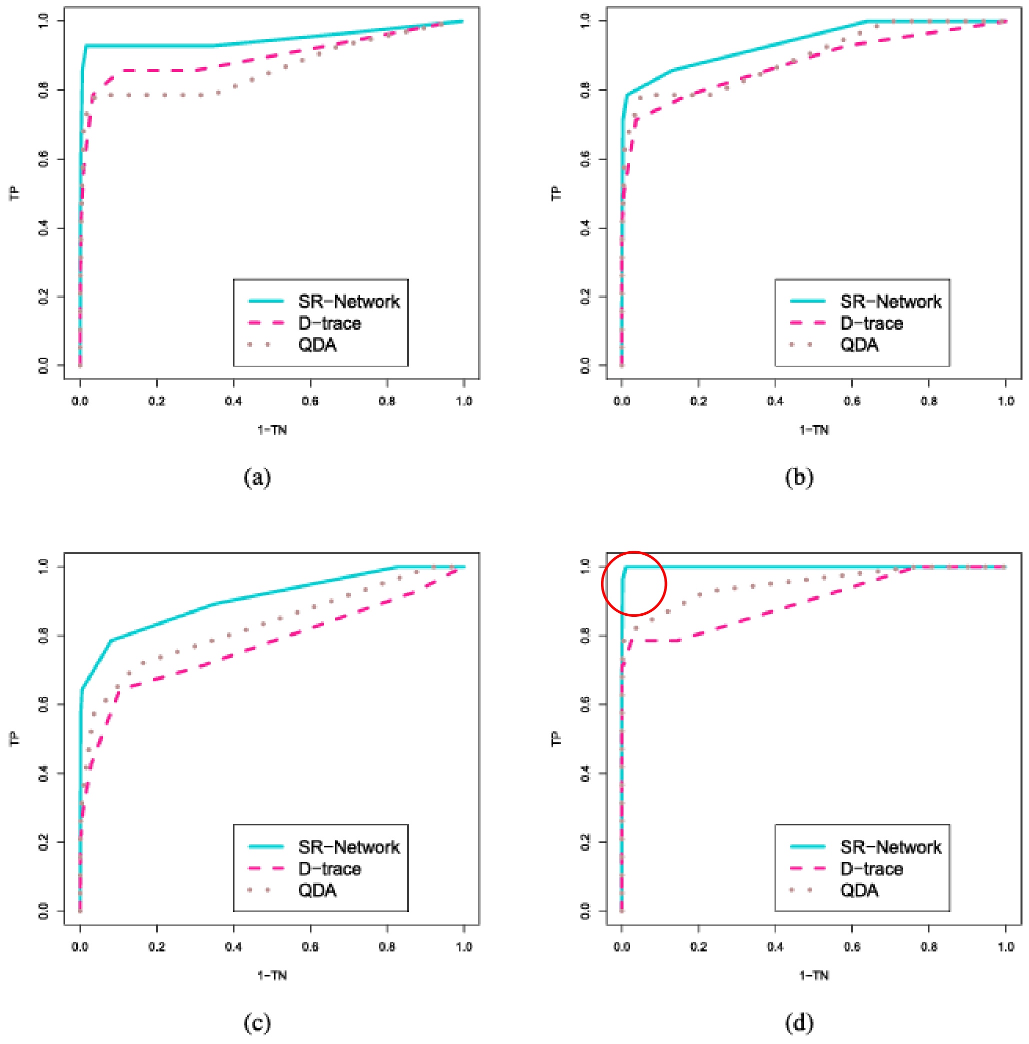
- 가정 2의 경우 표본의 비율을 통해 그룹간 적절한 비율을 유지함.
- 그리고 공분산 행렬의 조건수를 특정 상수로 제한해 조건수가 지나치게 커지는 것을 방지
- "가정이 많은데, 추정이 어느정도로 잘 한건지 비교가 어려웠음

4

Simulation

시뮬레이션 데이터

* ROC 커브

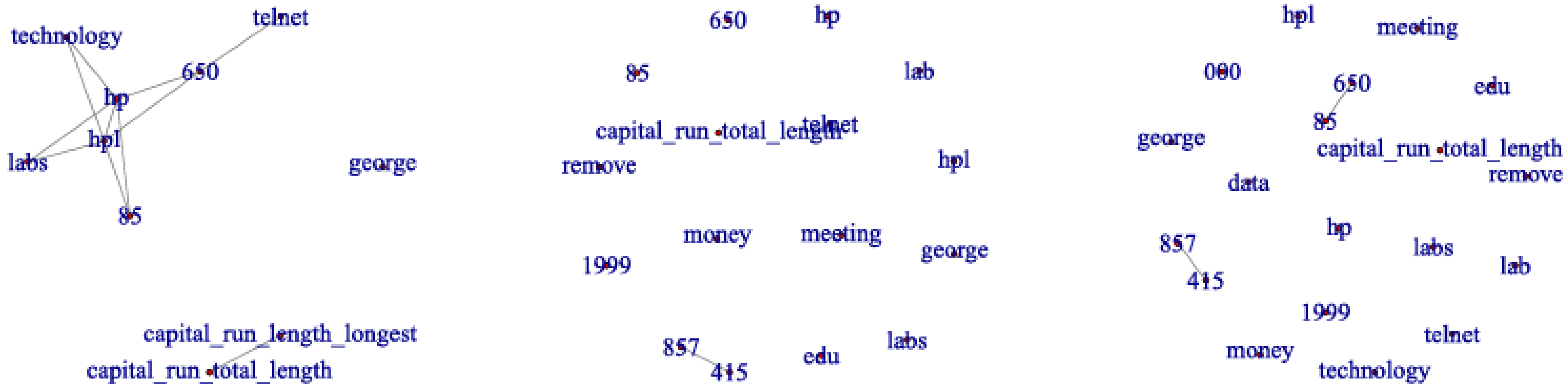


- $p = 50, 100, 120, 150$
- $X_1, \ldots, X_{n_1} \sim N(0, \Sigma_1)$
- $Y_1, \ldots, Y_{n_2} \sim N(0, \Sigma_2)$
- $n = 100$

▪ 제안하는 방법이 다른 방법에 비해 우수

▪ 하지만, p >> n인 세팅인 (d)에서 과적합으로 보이는데, 확실히 고차원 구조에서 추정이 어렵다는 것을 느꼈음

실제 데이터 - 스팸메일



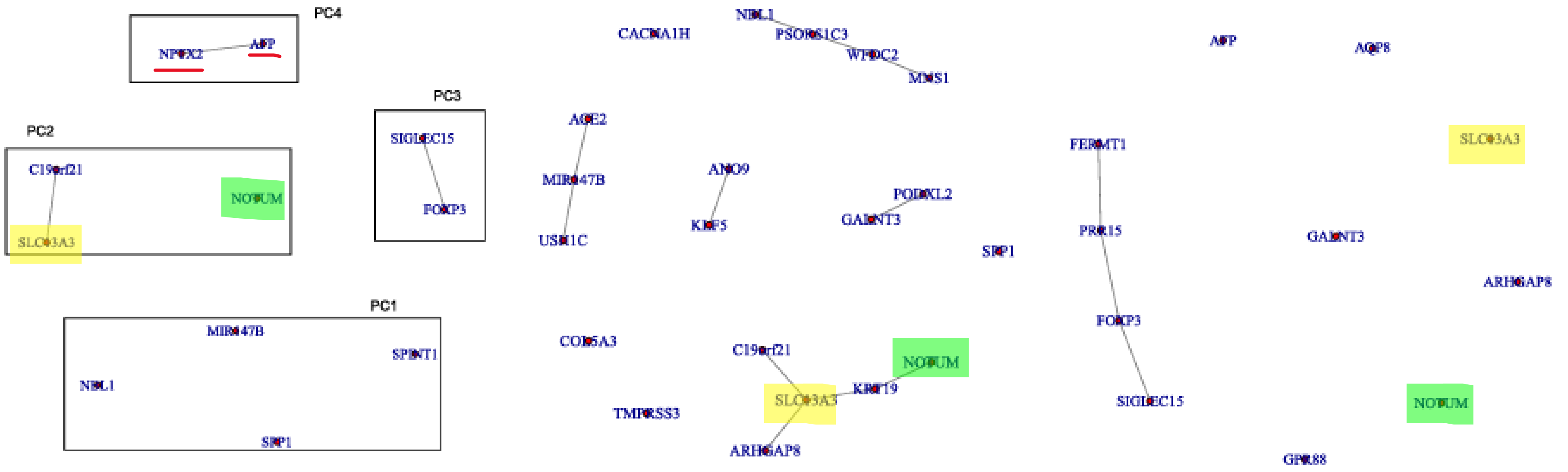
(a) SR-Network (b) D-trace (c) QDA

- 데이터셋의 경우 1813개의 스팸, 2788개의 스팸이 아닌 이메일을 56개의 변수로 구성

- 위 그림에서 재미있는 스팸인지 구분해낼 수 있는 잠재 요인을 식별
  - 특정 키워드
  - 메일 텍스트의 길이 등
  - HP 연구소에서 보낸 이메일

실제 데이터 - 간암 유전자 데이터

* 다른 방법은 잘 못찾지만 새로운 발견



(a) SR-Network (b) D-trace (c) QDA

- 첫 4개의 주성분(PCs)이 특정 유전자들에 대해 유의미한 계수를 가짐

- 형광펜은 기존 간암과 관련된 의미있는 유전자로 밝혀진 부분들 (모든 모델이 캡쳐)
- 하지만, 제안된 방법으로 latent factor를 캡쳐해 추가 연구에 대한 도움을 줄 수 있다.