Graphic Lasso: Identifiability of Multi-Layer Model

Jiaxin Hu

January 8, 2021

Suppose we have a dataset with p variables and K categories. In multi-layer model, we assume the rank of decomposition r is known, and the precision matrices are of form

$$\Omega^k = \Theta_0 + \sum_{l=1}^{r} u_{lk}\Theta_l, \quad \text{for} \quad k = 1, ..., K. \tag{1}$$

The identifiability problem for $\{\Theta_0, \Theta_1, ..., \Theta_r, \mathbf{u}_1, ..., \mathbf{u}_r\}$ is actually an identifiability problem for tensor decomposition.

Let $\mathcal{Y} \in \mathbb{R}^{p\times p\times K}$ denote the collection of K networks, where $\mathcal{Y}[,,k] = \Omega^k, k \in [K]$. Let $\mathcal{C} \in \mathbb{R}^{p\times p\times(r+1)}$ denote the collection of "core" networks, where $\mathcal{C}[,,1] = \sqrt{K}\Theta_0$, $\mathcal{C}[,,l] = \Theta_{l-1}, l = 2, ..., (r+1)$. Let $\boldsymbol{U} \in \mathbb{R}^{K\times(r+1)} = (\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_r)$ denote the factor matrix, where $\mathbf{u}_0 = \mathbf{1}_K/\sqrt{K}$. Rewrite the model (1) in tensor form.

$$\mathcal{Y} = \mathcal{C} \times_3 \boldsymbol{U}. \tag{2}$$

Therefore, the identifiability problem for $\{\Theta_l, \mathbf{u}_l\}$ becomes the identifiability problem for $\{\mathcal{C}, \boldsymbol{U}\}$. Let Unfold$(\cdot)$ denote the unfold representation of a tensor on mode 3. The model (2) is equal to the following matrix factorization model

$$\text{Unfold}(\mathcal{Y}) = \boldsymbol{U}\text{Unfold}(\mathcal{C}).$$

For simplicity, let $\mathbf{u}_k$ denote the k-th row of $\boldsymbol{U}$.

1 No sparsity constrain of U

1.1 Add constrain for distinct singular values

By SVD, we have decomposition Unfold$(\mathcal{Y}) = \tilde{\boldsymbol{U}}\Sigma\tilde{\boldsymbol{V}}$. If Unfold$(\mathcal{Y})$ has $(r+1)$ distinct singular values, the SVD is unique up to orthogonal rotation by letting the diagonal elements of Σ be in descending order. Further, since the first column of $\boldsymbol{U}$ should be $\mathbf{1}_K/\sqrt{K}$, the orthogonal rotation from $\tilde{\boldsymbol{U}}$ to $\boldsymbol{U}$ is unique, and then $\boldsymbol{U}$ and $\mathcal{C}$ are identifiable.

Next, we discuss the constrain to let Unfold$(\mathcal{Y})$ has $(r+1)$ distinct singular values. For simplicity, let $Y = \text{Unfold}(\mathcal{Y})$. Rewrite Y. We have

$$Y = \begin{pmatrix} \text{vec}^T(\Omega^1) \\ \vdots \\ \text{vec}^T(\Omega^K) \end{pmatrix}, \quad \sigma_i(Y) = \text{eigen}_i\left\{YY^T\right\} = \text{eigen}_i \begin{pmatrix} \left\|\Omega^1\right\|_F^2 & \langle\Omega^1, \Omega^2\rangle & \cdots & \langle\Omega^1, \Omega^K\rangle \\ \langle\Omega^1, \Omega^2\rangle & \left\|\Omega^2\right\|_F^2 & \cdots & \langle\Omega^2, \Omega^K\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle\Omega^1, \Omega^K\rangle & \langle\Omega^2, \Omega^K\rangle & \cdots & \left\|\Omega^K\right\|_F^2 \end{pmatrix},$$

for $i = 1$,

1.2 Add sparsity constrain of Θ_ls

2 Membership constrain of U

2.1 Hard membership constrain

We assume $\boldsymbol{U}$ is a hard membership matrix, i.e., for each row of $\boldsymbol{U}$ there is only 1 copy of 1 and massive 0.

2.1.1 Irreducible condition

Without intercept, by the Proposition 1 in Wang et al, $\{\boldsymbol{U}, \Theta_l\}$ are identifiable if $\mathcal{C}$ is irreducible on mode 3. The irreducibility is equivalent to saying that $\Theta_k \neq \Theta_l$, for all $k \neq l, k, l \in [r]$.

2.1.2 Add intercept Θ_0

Suppose we still want to cluster K categories into r groups. We need to add one of following constrains.

1. **Set to 0 constrain**

   Let $\tilde{\Theta}_1 = 0$. Consider the new parameters $\{\tilde{\boldsymbol{U}}, \tilde{\Theta}_0, \tilde{\Theta}_2, ..., \tilde{\Theta}_r\}$. The model (1) becomes

   $$\Omega^k = \tilde{\Theta}_0 + \tilde{\Theta}_{i_k}, \quad \text{for} \quad k = 1, ..., K.$$

   Then, $\tilde{\boldsymbol{U}}$ is identifiable by replacing the first column of $\boldsymbol{U}$ as $\mathbf{1}_K$. The collection of $\{\Theta_l\}$ are also identifiable by replacing $\tilde{\Theta}_0 = \Theta_1$, and $\tilde{\Theta}_l = \Theta_l - \Theta_1, l = 2, ..., r$.

2. **Sum to 0 constrain**

   To let $\{\tilde{\boldsymbol{U}}, \tilde{\Theta}_0, \tilde{\Theta}_1, ..., \tilde{\Theta}_r\}$ identifiable, we need one of the following constrains

   $$\sum_{l=1}^{r} \Theta_l = 0, \tag{3}$$

   or

   $$\sum_{l=1}^{r} m_l \Theta_l = 0, \tag{4}$$

   where $m_l = |\mathbf{u}_l|, l = 1, ..., r$. In some sense, the weighted sum to 0 constrain (4) is better because Θ_0 at this point is the average network of K categories; however, constrain (4) would be more difficult since we do not know the true value of m_l. In contrast, sum to 0 constrain (3) is more computationally convenient, and Θ_0 is the average network of $\{\Theta_l\}$ rather than the average of K categories.

2.2 Mixed membership constrain

We assume $\boldsymbol{U}$ is a mixed membership matrix, i.e., $\sum_{l=1}^{r} u_{kl} = 1, u_{kl} \geq 0$, for all $k \in [K]$.

First, we assume there is no intercept Θ_0. Note that mixed membership is a weaker constrain than hard membership. Therefore, the irreducible assumption on $\{\Theta_l\}$ is necessary condition.

Counter Example 1. Suppose we have the following decomposition

$$\Omega^k = \sum_{l=1}^{r} u_{kl}\Theta_l, \quad k \in [K]$$

Consider the new parameters

$$\tilde{\Theta}_l = c_l\Theta_l, \quad \text{and} \quad \tilde{u}_{kl} = \frac{1}{c_l}u_{kl}, \quad \text{for} \quad l \in [r], k \in [K],$$

where $(c_1, ..., c_r)$ are constants satisfy the following equation

$$\boldsymbol{U}\begin{pmatrix} \frac{1}{c_1} \\ \vdots \\ \frac{1}{c_r} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{5}$$

Then, we have

$$\Omega^k = \sum_{l=1}^{r} u_{kl}\Theta_l = \sum_{l=1}^{r} \tilde{u}_{kl}\tilde{\Theta}_l.$$

Conclusion 1: Assume $\boldsymbol{U}$ is full rank.
To avoid the case described in counter example 1, we need a condition to let $c_1 = \cdots = c_r = 1$. Note that $(c_1, ..., c_r)$ satisfy the equation (2.2). Note that $\boldsymbol{U}\mathbf{1}_r = \mathbf{1}_K$. If $\boldsymbol{U}$ is full rank, then $\mathbf{1}_r$ is the unique solution to equation (2.2). Then, the new parameter $\tilde{u}_{kl} = u_{kl}$ and $\tilde{\Theta}_l = \Theta_l$.

Counter Example 2. Let $u_{k,(r)}, u_{k,(r-1)}$ denote the largest and second largest elements in the row of $\boldsymbol{U}$, respectively. Suppose there is a dominate group for each category, i.e., $u_{k,(r)} > u_{k,(r-1)}$. However, the dominate group is not a sufficient condition for identifiability.

Consider a simple example.

$$\begin{aligned}\Omega^k &= 0.5\begin{pmatrix} 0 & 0 \\ 0 & x \end{pmatrix} + 0.2\begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix} + 0.3\begin{pmatrix} 0 & 0 \\ 0 & z \end{pmatrix} \\ &= 0.35\begin{pmatrix} 0 & 0 \\ 0 & x \end{pmatrix} + 0.4\begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix} + 0.25\begin{pmatrix} 0 & 0 \\ 0 & z \end{pmatrix}.\end{aligned}$$

Let (x, y, z) satisfy the equation $0.5x + 0.2y + 0.3z = 0.35x + 0.4y + 0.25z$. For example, $x = 4, y = 5, z = 8$. Then, $\Theta_1, \Theta_2, \Theta_3$ are irreducible but the dominate group of k category can be the first group OR the second group.

Conclusion 2: Dominate Group condition is not sufficient. (Irreducibility may not be sufficient too.) By counter example 2, the dominate group condition is not sufficient at this point. However, the reason why dominate group is not sufficient may be the $\Theta_1, \Theta_2, \Theta_3$ share the same structure but different magnitude. It seems like we use three Θs to describe one particular signal. In hard membership case, magnitude is sufficient to distinguish the signal because we only choose one group. Whereas, magnitude seems not sufficient in the mixed membership case. Therefore, irreducibility may also not be sufficient.

Conclusion 3: Non overlap condition is sufficient up to permutation

Proposition 1. *Assume $\boldsymbol{U}$ is full rank, $\{\Theta_l\}$ are nonzero matrices with non-overlapped nonzero sets. Then, $\boldsymbol{U}$ and $\{\Theta_l\}$ are identifiable.*

Following is the proof for the $r = 2$ case.

Proof. Consider the model

$$\Omega^k = u_{k1}\Theta^1 + u_{k2}\Theta^2, \quad k = 1, ..., K, \tag{6}$$

where $\Omega^k, \Theta_1, \Theta_2 \in \mathbb{R}^{p\times p}$. Let $I_1 = \left\{(i,j)|\Theta^1_{i,j} \neq 0\right\}$, $I_2 = \left\{(i,j)|\Theta^2_{i,j} \neq 0\right\}$, and $|I_1| = n_1 \geq 1, |I_2| = n_2 \geq 1$ with $n_1 + n_2 = p^2$. By the non-overlap condition, we have $I_1 \cap I_2 = \emptyset$. Let $\mathbf{u}_i = (u_{1i}, ..., u_{Ki}), i = 1, 2$, and $\Omega_{i,j} = (\Omega^1_{i,j}, ..., \Omega^K_{i,j}) \in \mathbb{R}^K$. By the full rankness of $\boldsymbol{U}$, $\mathbf{u}_1$ and $\mathbf{u}_2$ are linear independent. Rewrite the model (2.2) in the vector form.

$$\begin{bmatrix} \boldsymbol{I}_K\Theta^1_{i_1,j_1} & 0 \\ \vdots & 0 \\ \boldsymbol{I}_K\Theta^1_{i_{n_1},j_{n_1}} & 0 \\ 0 & \boldsymbol{I}_K\Theta^2_{i'_1,j'_1} \\ 0 & \vdots \\ 0 & \boldsymbol{I}_K\Theta^2_{i'_{n_2},j'_{n_2}} \end{bmatrix}_{Kp^2\times 2K} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}_{2K} = \begin{bmatrix} \Omega_{i_1,j_1} \\ \vdots \\ \Omega_{i_{n_1},j_{n_1}} \\ \Omega_{i_2,j_2} \\ \vdots \\ \Omega_{i_{n_2},j_{n_2}} \end{bmatrix}_{Kp^2}, \tag{7}$$

where $(i_k, j_k) \in I_1, k = 1, ..., n_1$, $(i'_l, j'_l) \in I_2, l = 1, ..., n_2$, and $\boldsymbol{I}_K$ is K-dimension identity matrix.

Suppose $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\Theta}^1, \tilde{\Theta}^2)$ also satisfy the model (2.2), with the corresponding nonzero sets $\tilde{I}_1$ and $\tilde{I}_2$.

First, assume $\tilde{I}_1 = I_1$ and $\tilde{I}_2 = I_2$. For all $(i,j) \in I_1$ and $(k,l) \in I_2$, we have $\Theta^1_{i,j}, \Theta^2_{k,l} \neq 0$ and $\tilde{\Theta}^1_{i,j}, \tilde{\Theta}^2_{k,l} \neq 0$. Then, we have

$$\begin{bmatrix} \Theta^1_{i,j}\mathbf{u}_1 \\ \Theta^2_{k,l}\mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\Theta}^1_{i,j}\tilde{\mathbf{u}}_1 \\ \tilde{\Theta}^2_{k,l}\tilde{\mathbf{u}}_2 \end{bmatrix} = \begin{bmatrix} \Omega_{i,j} \\ \Omega_{k,l} \end{bmatrix},$$

which implies that $\tilde{\mathbf{u}}_1 = \frac{\Theta^1_{i,j}}{\tilde{\Theta}^1_{i,j}}\mathbf{u}_1$ and $\tilde{\mathbf{u}}_2 = \frac{\Theta^2_{k,l}}{\tilde{\Theta}^2_{k,l}}\mathbf{u}_2$. By the definition of mixed membership, we have

$$\begin{bmatrix} \tilde{\mathbf{u}}_1 & \tilde{\mathbf{u}}_2 \end{bmatrix} \mathbf{1}_2 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \begin{pmatrix} \frac{\Theta^1_{i,j}}{\tilde{\Theta}^1_{i,j}} \\ \frac{\Theta^2_{k,l}}{\tilde{\Theta}^2_{k,l}} \end{pmatrix} = \mathbf{1}_K.$$

Since we also have $\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \mathbf{1}_2 = \mathbf{1}_K$ and $\mathbf{u}_1, \mathbf{u}_2$ are linear independent, we obtain that $\Theta^1_{i,j} = \tilde{\Theta}^1_{i,j}$ and $\Theta^2_{k,l} = \tilde{\Theta}^2_{k,l}$, and thus $\mathbf{u}_1 = \tilde{\mathbf{u}}_1$, $\mathbf{u}_2 = \tilde{\mathbf{u}}_2$.

Second, assume $\tilde{I}_1 \neq I_1$ and $\tilde{I}_2 \neq I_2$. Since the sparsity of $\{\Omega^k\}$ does not change, we have $I_1 \cup I_2 = \tilde{I}_1 \cup \tilde{I}_2$. Besides, to avoid the permutation case, we also assume $\tilde{I}_1 \neq I_2$ and $\tilde{I}_2 \neq I_1$. Then, there exist $(i_0, j_0) \in I_1 \cap \tilde{I}_1$ and $(i,j) \in I_1 \cap \tilde{I}_2$. We have

$$\Theta^1_{i_0,j_0}\mathbf{u}_1 = \tilde{\Theta}^1_{i_0,j_0}\tilde{\mathbf{u}}_1, \quad \text{and} \quad \Theta^1_{i,j}\mathbf{u}_1 = \tilde{\Theta}^2_{i,j}\tilde{\mathbf{u}}_2.$$

Therefore $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{u}}_2$ are linear dependent, which contradicts to the assumption that $\boldsymbol{U}$ should be full rank. Thus, such (i,j) does not exist and $\tilde{I}_1 = I_1, \tilde{I}_2 = I_2$. Then by the first case, we know that $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \tilde{\Theta}^1, \tilde{\Theta}^2) = (\mathbf{u}_1, \mathbf{u}_2, \Theta^1, \Theta^2)$.

Therefore, the parameters $(\mathbf{u}_1, \mathbf{u}_2, \Theta^1, \Theta^2)$ are identifiable up to permutation.

□

Remark 1. Note that non-overlap is a much stronger condition than irreducibility. Then we only requires $\boldsymbol{U}$ are full rank. If we can find a condition between irreducibility and non-overlap, the dominate group condition for $\boldsymbol{U}$ may be helpful.

3 Consistency

3.1 Magnitude constrain

3.2 Transfer the error from Ω^k to $\{\mathbf{u}_l, \Theta_l\}$