Do Multiple Networks improve Community Detection Performance?

Jan Greve, Kangcheng Lin

Introduction

In the project, we investigate the impact of multiple networks on the signal recovery in the problem of community detection. We model the problem as follows: Let $\mathbf{G}$ be a simple graph on n vertices and $\mathbf{X} = (X_1, X_2, ..., X_n)$ be a collection of labels associated with the vertices. Similar to [1], we have employed several observation models. The first one is the signal-plus-noise model:

$$\mathbf{Y} = \mathbf{X}S^{1/2} + N \tag{1}$$

where S is an $l \times l$ positive semidefinite matrix, known as the matrix SNR, and N is an $n \times l$ matrix with i.i.d standard Gaussian entries.
The second model is the symmetric matrix estimation model described by

$$\mathbf{Z} = \sqrt{\frac{t}{n}}\mathbf{X}\mathbf{R}\mathbf{X}^T + \xi \tag{2}$$

where ξ is a standard Gaussian Wigner matrix. We also define the signal part of the above as follows:

$$\mathbf{W} = \frac{1}{\sqrt{n}}\mathbf{X}\mathbf{R}\mathbf{X}^T \tag{3}$$

For the last model, the observations contains an n-node simple graph, which is represented by its adjacency matrix $\mathbf{G} \in \{0,1\}^{n \times n}$, where $G_{ij} = G_{ji} = 1$ if there is an edge between nodes i and j, or

$$G_{ij} \sim \text{Bernoulli}(\delta + \sqrt{\frac{\delta(1-\delta)}{n}} X_i^T R X_j), \quad i < j \tag{4}$$

Channel Universality

To approximate the potentially non-Gaussian (in our case Bernoulli) output channel with more manageable Gaussian channel, we utilize Channel Universality [2]. For a single adjacency matrix G as above, the Channel Universality is defined as:

$$I(W; G) = I(W; W + \sqrt{\Delta}\xi) + O(\sqrt{n}) \tag{5}$$

where Δ is the inverse Fisher information. Notice that this is equivalent to channel Z. Thus, we only need to establish a relationship between the first and the second channel.

Multiple Networks

For the sake of simplicity, we assume that multiple networks are realizations from the same W. Thus, we are essentially getting multiple networks with identical within-and-across group connection probability. Thus we have the following DAG:

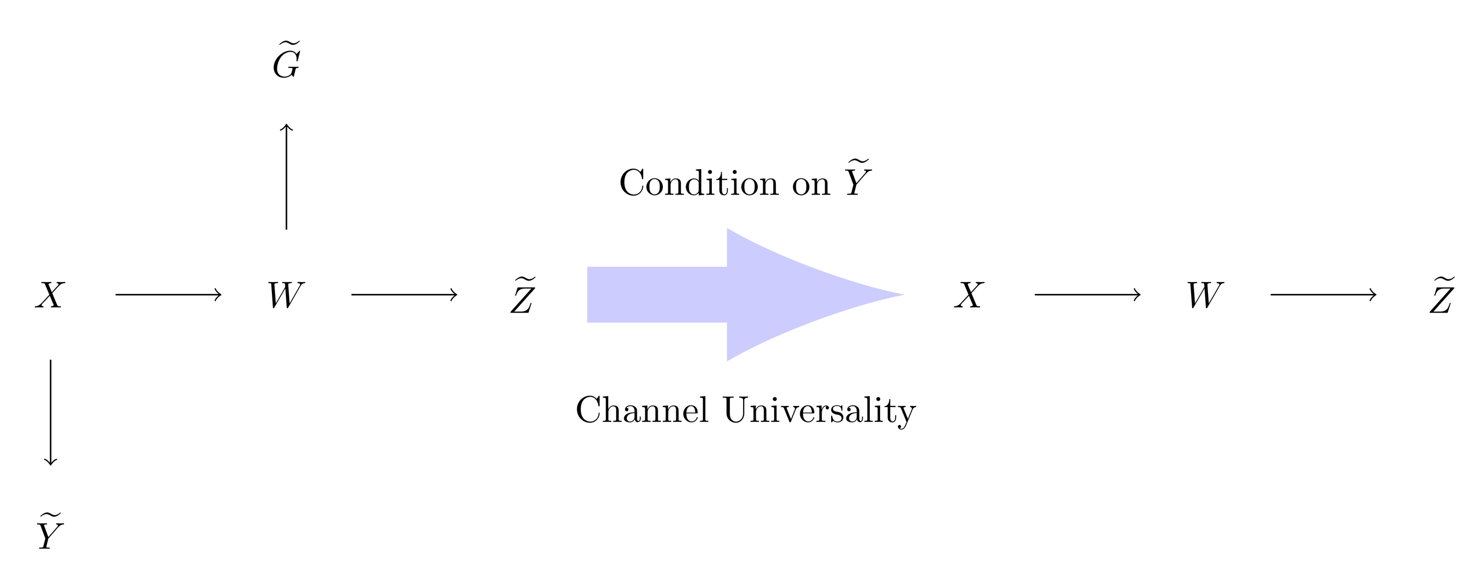

Figure 1: DAG of Our Multiple Netwkork Specification

where m stands for number of networks and $\tilde{Y} = \{Y_1, ...Y_m\}, \tilde{Z} = \{Z_1, ...Z_m\}, \tilde{G} = \{G_1, ..., G_m\}$. Following the derivation of Channel Universality in [2] we have

$$\log(P_{\text{out}}(\tilde{G}|W, \tilde{Y} = \tilde{y})) = \log(P_{\text{out}}(G_1|W, \tilde{Y} = \tilde{y}) \times \cdots \times P_{\text{out}}(G_m|W, \tilde{Y} = \tilde{y})) \approx m \log(P_{\text{out}}(G|W, Y = y))$$

Thus, following the definition of the mutual information, we have

$$I(W; \tilde{G}|\tilde{Y} = \tilde{y}) = mI(W; G|Y = y) = mI(W; W + \sqrt{\Delta}\xi|Y = y) + O(\sqrt{n})$$

Now that we obtained the relationship between mutual information of one network and multiple realizations through the term m, we proceed to examine how that affects the signal recovery performance through the minimization of the potential function.

Mathematical Section

Denote $\mathbb{S}^d$ and $\mathbb{S}^d_+$ as the spaces of $d \times d$ symmetric matrices and symmetric positive semi-definite matrices, respectively. Following [3], we define the mutual information function $I_X : \mathbb{S}^{k-1}_+ \to [0, \infty)$ such that $I_X(S) = I(\mathbf{X}; \mathbf{Y})$ and matrix-valued MMSE function $M_X : \mathbb{S}^{k-1}_+ \to \mathbb{S}^{k-1}_+$ such that $M_X(S) = \mathbb{E}[Cov(\mathbf{X}|\mathbf{Y})]$
Now, given that $\tilde{G} = (G_1, G_2, ..., G_m)$, we have that

$$G_{ij}^{(m)} \sim \text{Bernoulli}(\delta + \sqrt{\delta(1-\delta)} X_i^T R X_j / m) \tag{6}$$

from which we can derive that $\Delta = \frac{1}{m}$ in the assumption of Channel Universality, implying that $t = m$, and

$$Z = \sqrt{\frac{m}{n}}\mathbf{X}\mathbf{R}\mathbf{X}^T + \xi = \sqrt{\frac{1}{n}}\mathbf{X}\tilde{\mathbf{R}}\mathbf{X}^T + \xi \tag{7}$$

where $\tilde{\mathbf{R}} = \sqrt{m}R$. The second equation allows us to use the result in [1], which leads us to a potential function

$$\mathcal{F}(S) = I_X(S) + \frac{1}{4}tr((\tilde{\mathbf{R}} - \tilde{\mathbf{R}}^{-1}S)^2) \tag{8}$$

Taking the derivative with respect to S and using the I-MMSE relation [3], we obtain a fixed point equation

$$M_X(S) = I - \frac{1}{m}\mathbf{R}^{-1}S\mathbf{R}^{-1} \tag{9}$$

Simulation Results

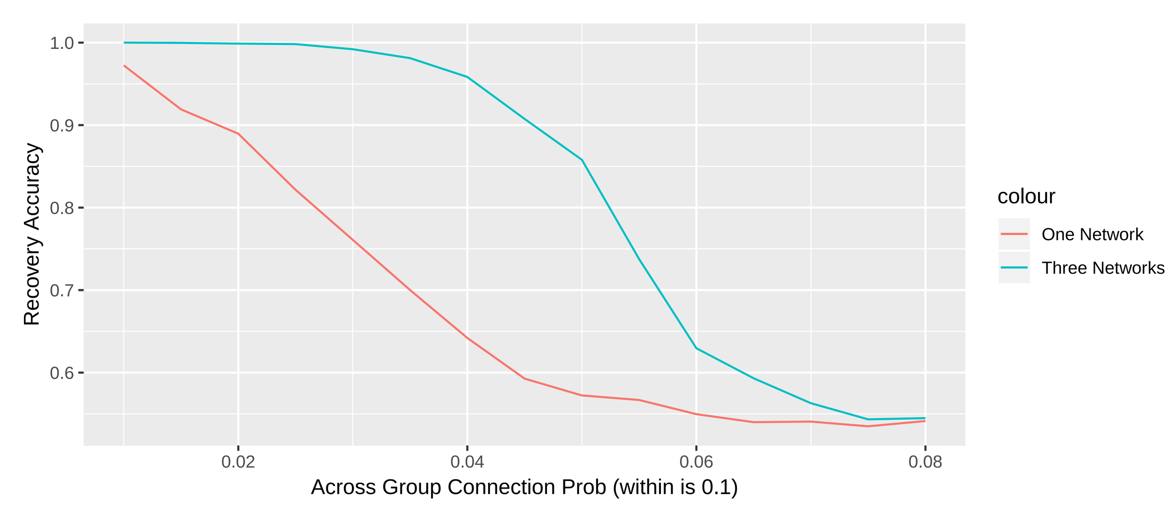

Figure 2: Recovery Performance with One and Three Networks

For comparison, we run spectral clustering on one and three networks with identical characteristics. Networks are symmetric with 2 communities with fixed within-community connection probability of 0.1 and varying across group probability ranging from 0.01 to 0.08.
The recovery algorithm we used is our implementation of [4] and when recovering from multiple networks, we run the algorithm on aggregated adjacency matrix.
As we can observe from the figure above, the recovery performance (averaged out 100 trials on every case) of the multiple network specification is always superior to the single one, albeit the difference is marginal in very high/low signal-to-noise ratio regime on both tails.

References

[1] Galen Reeves, Vaishakhi Mayya, and Alexander Volfovsky. The geometry of community detection via the mmse matrix. 2019.

[2] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016.

[3] Galen Reeves, Henry D Pfister, and Alex Dytso. Mutual information as a function of matrix snr for linear gaussian channels. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1754–1758. IEEE, 2018.

[4] Se-Young Yun and Alexandre Proutiere. Accurate community detection in the stochastic block model via spectral algorithms. *arXiv preprint arXiv:1412.7335*, 2014.