Reading Group Notes

Benjamin Draves

July 21, 2020

1 Eigen-Scaling Random Dot Product Graph

In Draves and Sussman (2020) we analyze the joint embedding of multiple networks over a common vertex set, $\mathcal{V}$. Suppose $n := |\mathcal{V}|$ and we observe m networks over this vertex set. We adopt a Random Dot Product Graph (RDPG) framework for the analysis of these embedding techniques.

To *jointly* model these networks, we propose a joint latent position model we entitled the *Eigen-Scaling Random Dot Product Graph* (ESRDPG). Under the ESRDPG, we assume that each vertex $v \in \mathcal{V}$ is associated with a latent vector $\mathbf{X}_v \in \mathbb{R}^d$ drawn i.i.d. from an inner product distribution F over an appropriate subset of $\mathbb{R}^d$. To capture network level heterogeneity, we associate each network with a *diagonal* matrix $\mathbf{C}^{(g)}$ for each $g \in [m]$. From here, we assume the probability that vertex i and j share an edge in graph $g \in [m]$ is given by $\mathbf{P}_{ij}^{(g)} = \mathbf{x}_i^T\mathbf{C}^{(g)}\mathbf{x}_i$. In essence, the weighting matrices alter the kernel for each graph. Finally, conditional on the latent position $\mathbf{X}$, the entries of random adjacency matrices are Bernoulli random variables specified by $\mathbf{A}_{ij}^{(g)}|\mathbf{X}_i, \mathbf{X}_j \sim \text{Bern}(\mathbf{X}_i^T\mathbf{C}^{(g)}\mathbf{X}_j)$.

Assumptions on this joint graph distribution are made to simplify analysis. Some assumptions are more restrictive than others, which I'll comment on below.

1. $\min_{i\in[d]} \max_{g\in[m]} \mathbf{C}_{ii}^{(g)} > 0$: This ensures that the weighting matrices aren't fully 'removing' a dimension of the latent space. If for some $i \in [d]$, $\mathbf{C}_{ii}^{(g)} = 0$ for all $g \in [m]$, then the following analysis would hold just with results written in $\mathbb{R}^{d-1}$.
2. $\mathbf{C}^{(g)} \geq 0$: This allows us to focus on the embedding of the *positive-definite* part of the adjacency matrices. We've recently extended these into embeddings that utilize both the high and low end of the spectrum, but the analysis becomes more complex: see Rubin-Delanchy et al. (2017).
3. $\Delta = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$ for $\mathbf{y} \sim F$ is diagonal: This enables an analytic computation of the bias.

2 Joint Embedding Techniques

In this work, we focus on the estimation of the *scaled latent positions*, $\mathbf{L} \in \mathbb{R}^{nm\times d}$ given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{X}\sqrt{\mathbf{C}^{(1)}} \\ \mathbf{X}\sqrt{\mathbf{C}^{(2)}} \\ \vdots \\ \mathbf{X}\sqrt{\mathbf{C}^{(m)}} \end{bmatrix}.$$

For ease of notation, let $\mathbf{L}^{(g)} = \mathbf{X}\sqrt{\mathbf{C}^{(g)}}$ We analyze estimators by their ability to recover row-wise estimates of $\mathbf{L}$ of the form $\sqrt{\mathbf{C}^{(g)}}\mathbf{X}_i$. We consider the following four estimators.

1. Adjacency Spectral Embedding (ASE): Ignore the shared structure between graphs. Estimate $\mathbf{L}^{(g)}$ with $\text{ASE}(\mathbf{A}^{(g)}, d) = \mathbf{U}_{\mathbf{A}^{(g)}}\mathbf{S}_{\mathbf{A}^{(g)}}^{1/2}$. Analyzed in Sussman et al. (2012) and Athreya et al. (2016).

2. Mean Embedding (Abar): Ignore the differences between graphs and estimate all $\mathbf{L}^{(g)}$ with $\text{ASE}(\bar{\mathbf{A}}, d) = \mathbf{U}_{\bar{\mathbf{A}}}\mathbf{S}_{\bar{\mathbf{A}}}^{1/2}$ where $\bar{\mathbf{A}} = m^{-1}\sum_{g=1}^{m}\mathbf{A}^{(g)}$. Analyzed in Tang et al. (2019).

3. Omnibus Embedding (Omni): Leverage joint structure by *simultaneously* embedding the adjacency matrices to arrive at a joint estimate matrix $\hat{\mathbf{L}} = \text{ASE}(\tilde{\mathbf{A}}, d) \in \mathbb{R}^{nm\times d}$ where $\tilde{\mathbf{A}} \in \mathbb{R}^{nm\times nm}$ is the omnibus matrix of the adjacency matrix. The i.i.d. case (i.e. $\mathbf{C}^{(g)} = \mathbf{I}$ for all $g \in [m]$) was analyzed in Levin et al. (2017). We provide the analysis of these estimates under the ESRDPG below.

4. Mean Omnibus Embedding (Omnibar): Since the omnibus embedding provides m different estimates, we can combine these estimates to produce a global estimate of the latent position. Let $\hat{\mathbf{X}}^{(g)}$ be g-th $n \times d$ block of $\hat{\mathbf{L}}$ and define $\bar{\mathbf{X}} = m^{-1}\sum_{g=1}^{m}\hat{\mathbf{X}}^{(g)}$. Then $\bar{\mathbf{X}}$ is the Omnibar estimator for each $\mathbf{L}^{(g)}$.

A stated goal of this work is to understand the (dis)advantages of each embedding technique stated above. We choose to first compare these estimators with respect to mean square error when estimating the rows of $\mathbf{L}$. Several numerical studies showed that Omni was often the most robust estimator in finite samples (see Figures 2, 4). In order to complete the comparison rigorously, we needed to establish the asymptotic properties of the Omni and Omnibar estimators under the ESRDPG. This is the content of our Main Results.

3 Main Results

3.1 Takeaways

There are two theorems that establish the fist and second moment properties of the omnibus embedding estimate under the ESRDPG. Theorem 1 reveals that the omnibus embedding is a *biased* estimator of $\mathbf{L}$ and the corresponding residual term concentrates at a rate of $O(m^{3/2}n^{-1/2}\log nm)$. Theorem 2 reveals that the rows of $\hat{\mathbf{L}}$ behave normally around this biased term when scaled by $\sqrt{n}$. This bias and variance term is also a function of each latent position. Figure 3 further highlights this point.

Together these results suggests there is a bias-variance tradeoff inherent in the omnibus embedding. The numerical examples throughout further support that the bias is offset by a sizable variance reduction in finite samples. This tradeoff typically makes the omnibus embedding the most robust estimator with respect to MSE in experimental settings.

These results, as well as identifying the asymptotic covariance between rows of $\hat{\mathbf{L}}$ (Corollary 1), allow for a formal comparison of the MSE of the four embedding estimators (Table 1). The ASE estimator is the only unbiased estimator while the other three estimators exhibit coordinate-scaling type biases. The variance terms are much more difficult to interpret but can be loosely interpreted as weighted - linear combinations of individual graph variances. For simple models, these explicit expressions can enable determination of the best MSE estimator of the estimators considered.

3.2 Asymptotic Expansion & Comments on Proof

Let $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{P}}$ be the omnibus matrix of $\{\mathbf{A}^{(g)}\}_{g=1}^{m}$ and $\{\mathbf{P}^{(g)}\}_{g=1}^{m}$, respectively. Levin et al. (2017) show that $\hat{\mathbf{Z}} = \text{ASE}(\tilde{\mathbf{A}}, d)$ concentrates around $\mathbf{Z} = \text{ASE}(\tilde{\mathbf{P}}, d)$ and the rows of $\sqrt{n}(\hat{\mathbf{Z}} - \mathbf{Z})$ behave normally (after appropriate rotation). In the i.i.d. case of Levin et al. (2017), $\mathbf{Z}$ can be written directly in terms of the latent positions as there exists $\mathbf{W} \in \mathcal{O}_d$ such that $\mathbf{Z}\mathbf{W} = \mathbf{1}_m \otimes \mathbf{X}$.

Under the ESRDPG, we needed to do more work to relate $\mathbf{Z}$ back to the latent positions $\mathbf{X}$. To that end, consider the following expansion of $\tilde{\mathbf{P}}$

$$\tilde{\mathbf{P}} = (\mathbf{I}_{m\times m} \otimes \mathbf{X})\tilde{\mathbf{C}}(\mathbf{I}_{m\times m} \otimes \mathbf{X})^T$$

where $\tilde{\mathbf{C}}$ is the omnibus matrix of $\{\mathbf{C}^{(g)}\}_{g=1}^m$. We show that $\tilde{\mathbf{C}}$, and by extension $\tilde{\mathbf{P}}$, is rank $2d$ with d positive and d negative eigenvalues. Let $\mathbf{S} = \text{ASE}(\tilde{\mathbf{C}}, d)$ and let it's g-th $d \times d$ block be written as $\mathbf{S}^{(g)}$ (See Definition 3.1). Then, by letting $\mathbf{L}_S = (\mathbf{I}_{m\times m} \otimes \mathbf{X})\mathbf{S}$ we have

$$\tilde{\mathbf{P}} = [(\mathbf{I}_{m\times m} \otimes \mathbf{X})\mathbf{S}][(\mathbf{I}_{m\times m} \otimes \mathbf{X})\mathbf{S}]^T + \tilde{\mathbf{P}}^- = \mathbf{L}_S\mathbf{L}_S^T + \tilde{\mathbf{P}}^-$$

where $\tilde{\mathbf{P}}^-$ is the negative definite part of $\tilde{\mathbf{P}}$. From here we can directly relate $\mathbf{Z}$ to $\mathbf{L}_S$ (see Lemma 2). These results suggest that we can extend the results of Levin et al. (2017) by considering the asymopitic expansion (See Appendix A)

$$\hat{\mathbf{L}} - \mathbf{L} = \underbrace{(\hat{\mathbf{L}} - \mathbf{L}_S)}_{2\to\infty,\text{ CLT}} + \underbrace{(\mathbf{L}_S - \mathbf{L})}_{\text{Bias}}.$$

The development of the residual bounds and the central limit theorem follow closely with that of Levin et al. (2017). The only slight difference is the eigenvalue rate of growth of $\tilde{\mathbf{P}}$ (see Lemma 4). In most work on RDPG CLTs, $\lambda_d(\mathbf{P}) = \Theta(n)$. Under i.i.d. latent positions, $\lambda_d(\tilde{\mathbf{P}}) = \Theta(nm)$. Under the ESRDPG, however, it is only guaranteed that $\lambda_d(\tilde{\mathbf{P}}) = \omega(\sqrt{m}n)$. This slower rate occurs in very few models (which are discussed in the Section 5), but results in the additional factor of m in the residual bound given in Theorem 1.

4 Statistical Consequences

Having established this Bias-Variance Tradeoff, we then turned to analyzing its ramifications in subsequent inference procedures. We chose to analyze the following inference problems.

1. Comm. Detection: Recover community labels within each graph using GMM applied to the latent position estimates provided by methods (1-4).
2. Multiplex Comm. Detection: Regarding the networks as describing a multiplex network, recover community labels shared across layers using the omnibus embedding estimates.
3. Two Graph Hypothesis Testing: Test if two graphs share the same set of latent positions.

Through several numerical studies, we show that GMM applied to the omnibus and omnibar estimates are competitive with the GMM applied to the ASE and Abar estimates with respect to missclassification rate under the Comm. Detection. We then show that the GMM applied to the Omnibar estimates is competitive with other Multiplex Comm. Detection methods that utilize different joint embedding techniques (Wang et al. 2017; Arroyo et al. 2019; Nielsen and Witten 2018). These results suggest that the bias does not directly harm, and may even aid, in consequent inference tasks. See Section 4.2 for a full discussion of these problems.

In Section 4.3 we develop a pivotal test statistic W that tests the hypothesis $\mathbf{H}_0 : \mathbf{C}^{(1)} = \mathbf{C}^{(2)}$. The statistic is reminiscent of a Hotelling's T^2 statistic and utilizes the distributional results established in the Central Limit Theorem. Under H_0, this statistic is approximately distributed $W \dot{\sim} \chi^2_{nd}$. We compare this new proposed statistic to the semi-parameteric statistic, T, proposed in Levin et al. (2017). As W incorporates covariance corrections (i.e. Mahalanobis distances) and T does not (i.e. Euclidean distance) we anticipate W will achieve higher power. In a simulation study, we find W does achieve higher power. However, its estimate in practice $\hat{W}$ is over-powered under H_0. Its level-corrected version $\tilde{W}$ achieves lower power than T for small networks sizes but achieves higher power for moderate network sizes ($n \geq 100$).

References

Arroyo, Jesús et al. (2019). "Inference for multiple heterogeneous networks with a common invariant subspace". In: arXiv preprint: 1906.10026.

Athreya, A. et al. (2016). "A Limit Theorem for Scaled Eigenvectors of Random Dot Product Graphs". In: *Sankhya A* 78.1, pp. 1–18.

Draves, Benjamin and Daniel L. Sussman (2020). "Bias-Variance Tradeoffs in Joint Spectral Embeddings". In: arXiv: 2005.02511 [math.ST].

Levin, Keith et al. (2017). "A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference". In: arXiv preprint: 1705.09355.

Nielsen, Agnes Martine and Daniela Witten (2018). "The Multiple Random Dot Product Graph Model". In: arXiv preprint: 1811.12172.

Rubin-Delanchy, Patrick et al. (2017). "A statistical interpretation of spectral embedding: the generalised random dot product graph". In: arXiv preprint: arXiv:1709.05506.

Sussman, Daniel L. et al. (2012). "A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs". In: *Journal of the American Statistical Association* 107.499, pp. 1119–1128.

Tang, Runze et al. (2019). "Connectome Smoothing via Low-Rank Approximations". In: *IEEE Transactions on Medical Imaging* 38, pp. 1446–1456.

Wang, Shangsi et al. (2017). "Joint Embedding of Graphs". In: arXiv preprint: 1703.03862.