Reading Group Notes

Joshua Agterberg

July 7, 2020

1 Notation

I will not bold anything, but I will use a bit different notation to Fan et al. (2019b) to keep things consistent with our previous discussion. In other words, P is the probability matrix, $P = UDU^\top$ and $A = \hat{U}\hat{D}\hat{U}^\top + \hat{U}_\perp \hat{D}_\perp \hat{U}_\perp^\top$, where $\hat{U}$ is the matrix of top d eigenvectors. Note that Fan et al. (2019b) uses H and W as the mean matrix and noise matrices respectively and use U and $\hat{U}$ for the eigenvectors. Also note that I will use B as the SBM connectivity matrix as is typically used; they use P for the same purpose, so hopefully there is no confusion.

They assume that they have K communities; I will also refer to the number of communities as K, but they assume throughout that P is rank K; that is $d = K$. I will make clear where something that is written as K is really only referring to the dimension of U ($= d$) or the number of communities ($= K$), as they are similar, but different things. Also, Carey likes to study $d < K$, and though the results of Fan et al. (2019b) hold for $d = K$, the results in Fan et al. (2019a) hold for a generic sequence of matrices.

Finally, I will use θ instead of the ρ_n I've used previously; this is because Fan et al. (2019b) consider the DCMMSBM in which the parameter θ or Θ corresponds to either a sparsity or degree-correction matrix respectively.

2 Notes on Fan et al. (2019b)

This paper is the only paper (that I know of) to study the test

$$H_0 : \pi_i = \pi_j$$
$$H_A : \pi_i \neq \pi_j$$

under the degree corrected mixed membership model (DCMMSBM), where π_i and π_j are the membership vectors of vertex i and j respectively. They assume that either $P = \theta \Pi B \Pi^\top$ where Π is the $n \times K$ matrix whose rows sum to one or $P = \Theta \Pi B \Pi^\top \Theta$, where Π is as above and Θ is a diagonal matrix of degree correction parameters. Again they assume B is full rank and that P has distinct eigenvalues (though I believe this can be relaxed following Cape et al. (2019b), but it might come at a high cost).

Their asymptotic results are motivated in part by those in Fan et al. (2019a) which considers asymptotics for bilinear forms and empirical eigenvalues. Note that the results in Fan et al. (2019a) should match those in Tang (2018) and Cape et al. (2019a) for eigenvalues and eigenvectors respectively (more on that in a bit).

Also note that if you don't believe that the rows of $\hat{U}$ or $\hat{X}$ are dependent in the limit, then this should convince you (as they explicitly use the limiting covariance).

Recall from last week that the rows of $\hat{U}$ look a lot like the rows of U and that the rows of U reveal the communities in the mixed-membership setting. The main idea in Fan et al. (2019b) is that the matrix $\hat{U}$ can be approximately viewed as a "data matrix;" that is, each row of $\hat{U}$ looks like a d-dimensional observation. As such, one can consider a Hotelling two-sample test for two different observations (e.g. testing equality of the mean of two normals). From Fan et al. (2019a) (and also Cape et al. (2019a)), we know that the rows of $\hat{U}$ are approximately Gaussian about the rows of U (up to orthogonal transformation). Therefore, under an appropriate asymptotic regime, one should be able to show that a Hotelling test statistic is asymptotically χ^2 under the null and noncentral χ^2 under the alternative.

Their test statistic is then of the form

$$T_{MM} := (\hat{U}_i - \hat{U}_j)^\top (\hat{\Sigma}_{ij}^{(1)})^{-1} (\hat{U}_i - \hat{U}_j),$$

in the mixed-membership setting and

$$T_{DC} := (Y_i - Y_j)^\top (\hat{\Sigma}_{ij}^{(2)})^{-1} (Y_i - Y_j)$$

where Y_i is $\hat{U}_i$ divided by its first component. The reason for the division here is because under the degree-corrected model, the ratios of the eigenvectors reveal their memberships. The Hotelling test statistic requires a covariance above; a quick examination and notation-match shows that they are using the covariance of the leading term in the CLT (EUD^{-1}). As I've said time and time again, this is the dominant term that comes up in all the limiting results; e.g. Cape et al. (2019a,b); Rubin-Delanchy et al. (2020); Athreya et al. (2016); Tang and Priebe (2018); Tang et al. (2017a). Note that when considering the scaled eigenvectors, the term is instead $EU|D|^{-1/2}$, but it is in the same spirit. Since they are considering the eigenvectors $\hat{U}_i$ and $\hat{U}_j$, they consider the covariance matrix $\mathrm{cov}((e_i - e_j)^\top EUD^{-1})$. They later show that if you have a decent enough estimator for this covariance, you can get the same results.

2.1 Discussion of their conditions

Their notation and conditions are a bit nonstandard in the literature, but they also are the only group (besides us maybe) to be consistently pushing out asymptotic normality results for spiked eigenvectors/values.

2.1.1 Condition 1

Their first condition is that P has distinct eigenvalues and that $\alpha_n \to \infty$ where $\alpha_n^2 := \max_i \sum_{j=1}^n \mathrm{Var}(E_{ij})$. Note that $\mathrm{Var}(E_{ij}) = \mathbb{E}(E_{ij}^2) = P_{ij}(1 - P_{ij})$. If you go to L. Lu and X. Peng (2013) (which shows that $\|E\| \leq O(\sqrt{\Delta})$) and read the proof carefully, you realize their actually bound this by $\max_i \sum_j P_{ij} =: \Delta$. In other words, the term α_n^2, while a bit unconventional in the graph literature does have a natural place in the analysis, and the max expected degree typically replaces it. The inclusion of α_n as opposed to Δ seems to me to really be stemming from the careful analysis they do that requires a finer handling of terms than one might normally need.

2.1.2 Condition 2

Their next condition is $\lambda_K(\Pi^\top \Pi) \geq c_0 n$, $\lambda_K(B) \geq c_0$, and $\theta \geq n^{-c_1}$. The first condition $\lambda_K(\Pi^\top \Pi)$ is saying that the memberships are approximately equal – there is no huge volume of vertices all in one community or others. Recall the bounds in Mao et al. (2019) had a precise dependence on this term in the denominator of their result and only required that it be larger than θ^{-1}; again, I attribute this to the precise asymptotic regime they are studying.

Their assumption $\lambda_K(B) \geq c_0$ is mostly nonrestrictive if B is full-rank; because they are considering a sequence of matrices $P = P(n)$, they need to assume that the eigenvalues of B do not degenerate as $n \to \infty$.

Finally, their assumption $\theta \geq n^{-c_1}$ for some $0 < c_1 < 1$ is somewhat restrictive compared to other results. Thinking of θ as a sparsity parameter, one might typicaly assume that $\theta \geq \omega(\log(n)/n)$ or even $\theta \geq \omega(\log^4(n)/n)$ which is not satisfied if θ is of order n^{-c_1}. The only other time I have seen a sparsity parameter required to have this order is in Tang et al. (2017b) which requires $\omega(\sqrt{n})$. As far as I can tell, the reason for θ having this order is that the sparsity ends up affecting the scaling and the variance in the asymptotic results (see Cape et al. (2019a); on the order of $n\sqrt{\theta}$), and hence for this setting the $\log(n)/n$ information-theoretical threshold might be too sparse.

2.1.3 Condition 3

Their condition 3 is a bit weird, but just seems to be mostly technical. Checking that the eigenvalues of $n^2\theta\Sigma_{ij}^{(1)}$ are bounded away from 0 and ∞ is I guess just something you have to do.

Quick remark– from Cape et al. (2019a), we know that the scaling for the eigenvectors to get asymptotic normality is $n\sqrt{\theta}$, so their scaling of $n^2\theta$ makes sense as it is the scaling needed since they have a quadratic form of their eigenvectors.

2.2 Their Results

Their results are quite nice without going too deep into the math.

Theorem 1 (Theorems 1 and 2 of Fan et al. (2019b))**.** *Under the null $\pi_i = \pi_j$ and the eigenvalues of $n^2\theta\Sigma_{ij}^{(1)}$ are bounded away from 0 and infinity, then $T_{MM} \to \chi^2_K$; a χ^2 distribution with K degrees of freedom assuming knowledge of the true $\Sigma_{ij}^{(1)}$.*

If instead $\sqrt{n\theta}||\pi_i - \pi_j|| \to \infty$ then for any large $C > 0$, we have $P(T_{MM} > C) \to 1$ as $n \to \infty$. Moreover, if the eigenvalues of $n^2\theta\Sigma_{ij}^{(1)}$ are bounded away from 0 and infinity, $\|\pi_i - \pi_j\| \sim (n\theta)^{-1/2}$ and $(U_i - U_j)^\top(\Sigma_{ij}^{(1)})^{-1}(U_i - U_j) \to \mu$ where μ is some constant, then $T_{mm} \to \chi^2_K(\mu)$, a noncentral χ^2 with noncentrality parameter μ.

Finally, if $\hat{K}$ and $\hat{\Sigma}$ are estimators of K and $\Sigma_{ij}^{(1)}$ satisfying $\mathbb{P}(\hat{K} = K) = 1 - o(1)$ and $n^2\theta\|\hat{\Sigma} - \Sigma_{ij}^{(1)}\| = o_{\mathbb{P}}(1)$, then the asymptotic results still hold.

The beauty of the result is that using the test statistic applied to the eigenvectors gets exactly the same asymptotic results as if one were performing a two sample test of equality of means for two multivariate normal distributions!

Theorem 2 (Theorems 3 and 4 of Fan et al. (2019b))**.** *Under the null $\pi_i = \pi_j$ and the eigenvalues of $n^2\theta\Sigma_{ij}^{(1)}$ are bounded away from 0 and infinity, then $T_{DC} \to \chi^2_{K-1}$; a χ^2 distribution with K degrees of freedom assuming knowledge of the true $\Sigma_{ij}^{(2)}$.*

If instead $\lambda_2(\pi_i\pi_i^\top + \pi_j\pi_j^\top) \gg (n\theta_{\min}^2)^{-1}$ *then for any large* $C > 0$*, we have* $P(T_{DC} > C) \to 1$ *as* $n \to \infty$*.*

Finally, if $\hat{K}$ *and* $\hat{\Sigma}$ *are estimators of* K *and* $\Sigma_{ij}^{(2)}$ *satisfying* $\mathbb{P}(\hat{K} = K) = 1 - o(1)$ *and* $n^2\theta\|\hat{\Sigma} - \Sigma_{ij}^{(1)}\| = o_{\mathbb{P}}(1)$*, then the asymptotic results still hold.*

The conditions for the previous theorem are mostly similar to that of the MMSBM, though Condition 5 is a bit stringent. They require that B is positive definite and irreducible with $\max_k B_{kk} = 1$. I do not think this is required, but rather is needed for their analysis, though I didn't dig deep enough to figure out where they used it.

Finally, a remark on the estimators for $\Sigma_{ij}^{(1)}$ and $\Sigma_{ij}^{(2)}$: they note from Fan et al. (2019a) (and also Tang (2018)) that the eigenvalues are biased estimators of the true eigenvalues. Since $\Sigma_{ij}^{(1)} = \mathrm{cov}((e_i - e_j)^\top EUD^{-1})$ depends on the eigenvalues, they include an explicit debiasing procedure using the limiting eigenvalue result which results in an estimator that is $o_{\mathbb{P}}((n^2\theta)^{-1})$.

3 Discussion Points

- Their result depends on the limiting eigenvector forms in Fan et al. (2019a) in which the eigenvalues are determined as the solution to a fixed-point equation. However, Tang (2018) has an explicit calculation for these eigenvalues (and a joint distribution) that depends on a number of things – but one can write it down without solving a fixed-point equation. Could one use this in practice?
- What would the one-sample version of the test be?

References

A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman. A Limit Theorem for Scaled Eigenvectors of Random Dot Product Graphs. *Sankhya A*, 78(1):1–18, Feb. 2016. ISSN 0976-8378. doi: 10.1007/s13171-015-0071-x. URL https://doi.org/10.1007/s13171-015-0071-x.

J. Cape, M. Tang, and C. E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250, Mar. 2019a. ISSN 0006-3444. doi: 10.1093/biomet/asy070. URL https://academic.oup.com/biomet/article/106/1/243/5280315.

J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, Oct. 2019b. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1752. URL https://projecteuclid.org/euclid.aos/1564797852.

J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic Theory of Eigenvectors for Large Random Matrices. *arXiv:1902.06846 [math, stat]*, Feb. 2019a. URL http://arxiv.org/abs/1902.06846. arXiv: 1902.06846.

J. Fan, Y. Fan, X. Han, and J. Lv. SIMPLE: Statistical Inference on Membership Profiles in Large Networks. *arXiv:1910.01734 [math, stat]*, Oct. 2019b. URL http://arxiv.org/abs/1910.01734. arXiv: 1910.01734.

L. Lu and X. Peng. Spectra of edge-independent random graphs. *Electronic Journal of Combinatorics*, 20, 2013.

X. Mao, P. Sarkar, and D. Chakrabarti. Estimating Mixed Memberships with Sharp Eigenvector Deviations. *arXiv:1709.00407 [physics, stat]*, Nov. 2019. URL http://arxiv.org/abs/1709.00407. arXiv: 1709.00407.

P. Rubin-Delanchy, J. Cape, M. Tang, and C. E. Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *arXiv:1709.05506 [cs, stat]*, Jan. 2020. URL http://arxiv.org/abs/1709.05506. arXiv: 1709.05506.

M. Tang. The eigenvalues of stochastic blockmodel graphs. *arXiv:1803.11551 [cs, stat]*, Mar. 2018. URL http://arxiv.org/abs/1803.11551. arXiv: 1803.11551.

M. Tang and C. E. Priebe. Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, Oct. 2018. ISSN 0090-5364, 2168-8966. doi: 10.1214/17-AOS1623. URL https://projecteuclid.org/euclid.aos/1534492839.

M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe. A Semiparametric Two-Sample Hypothesis Testing Problem for Random Graphs. *Journal of Computational and Graphical Statistics*, 26 (2):344–354, Apr. 2017a. ISSN 1061-8600. doi: 10.1080/10618600.2016.1193505. URL https://doi.org/10.1080/10618600.2016.1193505.

M. Tang, J. Cape, and C. E. Priebe. Asymptotically efficient estimators for stochastic blockmodels: the naive MLE, the rank-constrained MLE, and the spectral. *arXiv:1710.10936 [stat]*, Oct. 2017b. URL http://arxiv.org/abs/1710.10936. arXiv: 1710.10936.