Submitted to Bernoulli
arXiv: arXiv:1305.7388

Addendum – Addenda – Da Dum Da Dum

EDMUND BLACKADDER,

Department of Applied Mathematics and Statistics, Johns Hopkins University
E-mail: ebladder@jhu.edu

We extend Perfect results through a cunning argument; an argument that, in fact, is even more cunning than a fox that was once appointed Distinguished Professor of Cunning at Oxford University.

Keywords: wibble wibble, nibble nibble, quibble quibble.

***Definition* 1** (Random Dot Product Graph (d-dimensional))**.** Let F be a distribution on a set $\mathcal{X} \subset \mathbb{R}^d$ satisfying $x^\top y \in [0,1]$ for all $x, y \in \mathcal{X}$. We say $(\mathbf{X}, \mathbf{A}) \sim \mathrm{RDPG}(F)$ with sparsity factor $\rho_n \leq 1$ if the following hold. Let $X_1, \ldots, X_n \sim F$ be independent random variables and define

$$\mathbf{X} = [X_1 \mid \cdots \mid X_n]^\top \in \mathbb{R}^{n\times d} \text{ and } \mathbf{P} = \rho_n \mathbf{X}\mathbf{X}^\top \in [0,1]^{n\times n}. \tag{1}$$

The X_i are the latent positions for the random graph. The matrix $\mathbf{A} \in \{0,1\}^{n\times n}$ is defined to be a symmetric matrix with all zeroes on the diagonal such that for all $i < j$, conditioned on X_i, X_j the A_{ij} are independent and

$$A_{ij} \sim \mathrm{Bernoulli}(\rho_n X_i^\top X_j), \tag{2}$$

namely,

$$\mathbb{P}[\mathbf{A} \mid \mathbf{X}] = \prod_{i<j} (\rho_n X_i^\top X_j)^{A_{ij}} (1 - \rho_n X_i^\top X_j)^{(1-A_{ij})} \tag{3}$$

Remark. We denote the second moment matrix for the X_i by $\Delta = \mathrm{E}(X_1 X_1^\top)$. For the remainder of this work we shall assume that Δ is of rank d. Finally, let δ_d denote $\lambda_d(\mathbb{E}[X_1 X_1^\top])$, the smallest eigenvalue of Δ

***Definition* 2** (Embedding of $\mathbf{A}$ and $\mathbf{P}$)**.** Suppose that $\mathbf{A}$ is as in Definition 1. Let $\mathbf{U}\mathbf{S}\mathbf{U}^\top$ be the spectral decomposition of $|\mathbf{A}| = (\mathbf{A}^\top \mathbf{A})^{1/2}$. Then our estimate for the $\rho_n^{1/2}\mathbf{X}$ (up to rotation) is $\hat{\mathbf{X}} = \mathbf{U}_{\mathbf{A}} \mathbf{S}_{\mathbf{A}}^{1/2}$, where $\mathbf{S}_{\mathbf{A}} \in \mathbb{R}^{d\times d}$ is the diagonal submatrix of $\mathbf{S}$ with the d largest eigenvalues (in magnitude) of $|\mathbf{A}|$ and $\mathbf{U}_{\mathbf{A}} \in \mathbb{R}^{n\times d}$ is the submatrix of $\mathbf{U}$ whose orthonormal columns are the corresponding eigenvectors. Similarly, we let $\mathbf{U}_{\mathbf{P}} \mathbf{S}_{\mathbf{P}} \mathbf{U}_{\mathbf{P}}^\top$ denote the spectral decomposition of $\mathbf{P}$. Note that $\mathbf{P}$ is of rank d.

We start with a simple result[1]

***Proposition* 1.** Let $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$. Let $\mathbf{W}_1 \mathbf{\Sigma} \mathbf{W}_2^\top$ be the singular value decomposition of $\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A}$. Then for sufficiently large n,

$$\|\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A} - \mathbf{W}_1 \mathbf{W}_2^\top\|_F = O((n\rho_n)^{-1})$$

with high probability.

Proof. Let $\sigma_1, \sigma_2, \dots, \sigma_d$ denote the singular values of $\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A}$ (the diagonal entries of $\mathbf{\Sigma}$). Then $\sigma_i = \cos(\theta_i)$ where the θ_i are the principal angles between the subspaces spanned by $\mathbf{U}_\mathbf{A}$ and $\mathbf{U}_\mathbf{P}$. Furthermore, by the Davis-Kahan $\sin(\Theta)$ theorem (see e.g., Theorem 3.6 in Stewart and Sun [1990]),

$$\|\mathbf{U}_\mathbf{A} \mathbf{U}_\mathbf{A}^\top - \mathbf{U}_\mathbf{P} \mathbf{U}_\mathbf{P}^\top\| = \max_i |\sin(\theta_i)| \leq \frac{\|\mathbf{A} - \mathbf{P}\|}{\lambda_d(\mathbf{P})} \leq \frac{C\sqrt{n\rho_n}}{n\rho_n} = O((n\rho_n)^{-1/2})$$

for sufficiently large n. Here $\lambda_d(\mathbf{P})$ denotes the d-th largest eigenvalue of $\mathbf{P}$.

We thus have

$$\begin{aligned}
\|\mathbf{U}_\mathbf{P}^\top \mathbf{U}_\mathbf{A} - \mathbf{W}_1 \mathbf{W}_2^\top\|_F = \|\mathbf{\Sigma} - \mathbf{I}\|_F &= \sqrt{\sum_{i=1}^{d} (1 - \sigma_i)^2} \\
&\leq \sum_{i=1}^{d} (1 - \sigma_i) \leq \sum_{i=1}^{d} (1 - \sigma_i^2) \\
&= \sum_{i=1}^{d} \sin^2(\theta_i) \leq d \|\mathbf{U}_\mathbf{A} \mathbf{U}_\mathbf{A}^T - \mathbf{U}_\mathbf{P} \mathbf{U}_\mathbf{P}^\top\|^2 = O((n\rho_n)^{-1})
\end{aligned}$$

as desired. □

From now on, we shall denote by $\mathbf{W}^*$ the orthogonal matrix $\mathbf{W}_1 \mathbf{W}_2^\top$ as defined in the above proposition. Next, we have

***Lemma* 2.** Let $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$. Then for sufficiently large n,

$$\|\mathbf{W}^* \mathbf{S}_\mathbf{A} - \mathbf{S}_\mathbf{P} \mathbf{W}^*\|_F = O(1); \quad \text{and} \quad \|\mathbf{W}^* \mathbf{S}_\mathbf{A}^{1/2} - \mathbf{S}_\mathbf{P}^{1/2} \mathbf{W}^*\|_F = O((n\rho_n)^{-1/2})$$

with high probability.

[1] Many of the bounds in here have a missing $\sqrt{\log n}$ factor compared with the corresponding bounds in CLT, Perfect and Semipar. This is mainly due to a tighter concentration inequality for $\|\mathbf{A} - \mathbf{P}\|$ from [Lu and Peng, 2013]

Proof. Let $\mathbf{R} = \mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}$. We note that $\mathbf{R}$ is the residual after projecting $\mathbf{U_A}$ orthogonally onto the column space of $\mathbf{U_P}$. In particular, this implies that

$$\|\mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}\|_F \leq \|\mathbf{U_A} - \mathbf{U_P}\mathbf{T}\|_F$$

for all $d \times d$ matrices $\mathbf{T}$. Therefore, $\|\mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}\|_F = O((n\rho_n)^{-1/2})$. We derive that

$$\begin{aligned}
\mathbf{W}^*\mathbf{S_A} &= (\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A})\mathbf{S_A} + \mathbf{U_P}^\top\mathbf{U_A}\mathbf{S_A} = (\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A})\mathbf{S_A} + \mathbf{U_P}^\top\mathbf{A}\mathbf{U_A} \\
&= (\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A})\mathbf{S_A} + \mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_A} + \mathbf{U_P}^\top\mathbf{P}\mathbf{U_A} \\
&= (\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A})\mathbf{S_A} + \mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{R} + \mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A} + \mathbf{U_P}^\top\mathbf{P}\mathbf{U_A} \\
&= (\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A})\mathbf{S_A} + \mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{R} + \mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A} + \mathbf{S_P}\mathbf{U_P}^\top\mathbf{U_A}
\end{aligned}$$

Writing $\mathbf{S_P}\mathbf{U_P}^\top\mathbf{U_A} = \mathbf{S_P}(\mathbf{U_P}^\top\mathbf{U_A} - \mathbf{W}^*) + \mathbf{S_P}\mathbf{W}^*$ and rearranging terms, we obtain

$$\begin{aligned}
\|\mathbf{W}^*\mathbf{S_A} - \mathbf{S_P}\mathbf{W}^*\|_F &\leq \|\mathbf{W}^* - \mathbf{U_P}^\top\mathbf{U_A}\|_F(\|\mathbf{S_A}\| + \|\mathbf{S_P}\|) + \|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{R}\|_F + \|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\|_F \\
&\leq O(1) + O(1) + \|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\|_F
\end{aligned}$$

with high probability. Now, $\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}$ is a $d \times d$ matrix whose ij-th entry is of the form

$$\boldsymbol{u}_i^\top(\mathbf{A} - \mathbf{P})\boldsymbol{u}_j = \sum_{k=1}^{n}\sum_{l=1}^{n}(\mathbf{A}_{kl} - \mathbf{P}_{kl})\boldsymbol{u}_{ik}\boldsymbol{u}_{jl} = 2\sum_{k<l}(\mathbf{A}_{kl} - \mathbf{P}_{kl}) + \sum_{k=1}\mathbf{P}_{kk}\boldsymbol{u}_{ik}\boldsymbol{u}_{jk}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are the i-th and j-th columns of $\mathbf{U_P}$. Thus, conditioned on $\mathbf{P}$, $\boldsymbol{u}_i^\top(\mathbf{A} - \mathbf{P})\boldsymbol{u}_j$ is a sum of independent mean 0 random variables and a term of order $O(1)$. [2] Now, by Hoeffding's inequality [3].

$$\mathbb{P}[|\sum_{k<l} 2(\mathbf{A}_{kl} - \mathbf{P}_{kl})\boldsymbol{u}_{ik}\boldsymbol{u}_{jl}| \geq t] \leq 2\exp\Big(\frac{-2t^2}{\sum_{k<l}(2\boldsymbol{u}_{ik}\boldsymbol{u}_{jl})^2}\Big) \leq 2\exp(-t^2).$$

Therefore, each entry of $\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}$ is of order $O(1)$ with high probability, and as a consequence, $\|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\|_F$ is of order $O(1)$ with high probability. Hence, $\|\mathbf{W}^*\mathbf{S_A} - \mathbf{S_P}\mathbf{W}^*\| = O(1)$ with high probability. We establish $\|\mathbf{W}^*\mathbf{S_A}^{1/2} - \mathbf{S_P}^{1/2}\mathbf{W}^*\|_F = O(n^{-1/2})$ by noting that the ij-th entry of $\mathbf{W}^*\mathbf{S_A}^{1/2} - \mathbf{S_P}^{1/2}\mathbf{W}^*$ can be written as

$$\mathbf{W}^*_{ij}(\lambda_i^{1/2}(\mathbf{A}) - \lambda_j^{1/2}(\mathbf{P})) = \mathbf{W}^*_{ij}\frac{\lambda_i(\mathbf{A}) - \lambda_j(\mathbf{P})}{\lambda_i^{1/2}(\mathbf{A}) + \lambda_j^{1/2}(\mathbf{P})}$$

and that the eigenvalues $\lambda_i^{1/2}(\mathbf{A})$ and $\lambda_j^{1/2}(\mathbf{P})$ are all of order $O(\sqrt{n\rho_n})$. □

[2] Note that conditioned on $\mathbf{P}$, $\mathbf{U_P}$ is fixed and non-random.

[3] One might be able to obtain a tighter bound using e.g., Bernstein inequality, however, as the other terms above are already of order $O(1)$, a tighter bound is not necessary

We then have

Theorem 3. Let $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$. Then there exists a rotation matrix $\mathbf{W}$ such that for sufficiently large n

$$\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_F = \|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S_P}^{-1/2}\|_F + O((n\rho_n)^{-1/2})$$

with high probability.

Proof. Let $\mathbf{R}_1 = \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A} - \mathbf{U_P}\mathbf{W}^*$ and $\mathbf{R}_2 = (\mathbf{W}^*\mathbf{S}_\mathbf{A}^{1/2} - \mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^*)$. We then have

$$\begin{aligned}\hat{\mathbf{X}} - \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^* &= \mathbf{U_A}\mathbf{S}_\mathbf{A}^{1/2} - \mathbf{U_P}\mathbf{W}^*\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{U_P}(\mathbf{W}^*\mathbf{S}_\mathbf{A}^{1/2} - \mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^*) \\ &= (\mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A})\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{R}_1\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{U_P}\mathbf{R}_2 \\ &= \mathbf{U_A}\mathbf{S}_\mathbf{A}^{1/2} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{R}_1\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{U_A}\mathbf{R}_2\end{aligned}$$

Now, $\mathbf{U_P}\mathbf{U_P}^\top\mathbf{P} = \mathbf{P}$ and $\mathbf{U_A}\mathbf{S}_\mathbf{A}^{1/2} = \mathbf{A}\mathbf{U_A}\mathbf{S}_\mathbf{A}^{-1/2}$ and hence

$$\hat{\mathbf{X}} - \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^* = (\mathbf{A} - \mathbf{P})\mathbf{U_A}\mathbf{S}_\mathbf{A}^{-1/2} - \mathbf{U_P}\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_A}\mathbf{S}_\mathbf{A}^{-1/2} + \mathbf{R}_1\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{U_A}\mathbf{R}_2$$

Writing $\mathbf{R}_3 = \mathbf{U_A} - \mathbf{U_P}\mathbf{W}^* = \mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A} + \mathbf{R}_1$ we then have

$$\begin{aligned}\hat{\mathbf{X}} - \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^* &= (\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2} + \mathbf{U_P}\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2} \\ &\quad + (\mathbf{I} - \mathbf{U_P}\mathbf{U_P}^\top)(\mathbf{A} - \mathbf{P})\mathbf{R}_3\mathbf{S}_\mathbf{A}^{-1/2} + \mathbf{R}_1\mathbf{S}_\mathbf{A}^{1/2} + \mathbf{U_A}\mathbf{R}_2\end{aligned}$$

Now $\|\mathbf{R}_1\|_F = O((n\rho_n)^{-1})$, $\|\mathbf{R}_2\|_F = O((n\rho_n)^{-1/2})$ and $\|\mathbf{R}_3\|_F = O((n\rho_n)^{-1/2})$; indeed, $\mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}$ is the residual after we projected $\mathbf{U_A}$ orthogonally onto the column space of $\mathbf{U_P}$ and hence $\|\mathbf{U_A} - \mathbf{U_P}\mathbf{U_P}^\top\mathbf{U_A}\|_F = \min_\mathbf{T}\|\mathbf{U_A} - \mathbf{U_P}\mathbf{T}\|_F = O((n\rho_n)^{-1/2})$. Furthermore, we have

$$\|\mathbf{U_P}\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2}\|_F \leq \|\mathbf{U_P}^\top(\mathbf{A} - \mathbf{P})\mathbf{U_P}\|_F\|\mathbf{S}_\mathbf{A}^{-1/2}\|_F = O((n\rho_n)^{-1/2})$$

by Hoeffding's inequality. Therefore

$$\begin{aligned}\|\hat{\mathbf{X}} - \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^*\|_F &= \|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2}\|_F + O(n^{-1/2}) \\ &= \|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S}_\mathbf{P}^{-1/2}\mathbf{W}^* + (\mathbf{A} - \mathbf{P})\mathbf{U_P}(\mathbf{S}_\mathbf{P}^{-1/2}\mathbf{W}^* - \mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2})\|_F + O(n^{-1/2})\end{aligned}$$

Similar to our derivation of Lemma 2, we can show that

$$\|\mathbf{S}_\mathbf{P}^{-1/2}\mathbf{W}^* - \mathbf{W}^*\mathbf{S}_\mathbf{A}^{-1/2}\|_F = O((n\rho_n)^{-3/2})$$

and hence

$$\begin{aligned}\|\hat{\mathbf{X}} - \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^*\|_F &= \|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S}_\mathbf{P}^{-1/2}\mathbf{W}^*\|_F + O((n\rho_n)^{-1/2}) \\ &= \|(\mathbf{A} - \mathbf{P})\mathbf{U_P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F + O((n\rho_n)^{-1/2}).\end{aligned} \tag{4}$$

Finally, to complete the proof, we note that $\mathbf{X} = \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}$ for some orthogonal matrix $\mathbf{W}$. As $\mathbf{W}^*$ is also orthogonal, therefore $\mathbf{X}\tilde{\mathbf{W}} = \mathbf{U_P}\mathbf{S}_\mathbf{P}^{1/2}\mathbf{W}^*$ for some orthogonal $\tilde{\mathbf{W}}$. □

We now arrive at the extension of the perfect clustering result.

***Corollary* 4.** Let $(\mathbf{A}, \mathbf{X}) \sim \mathrm{RDPG}(F)$. Then there exists a rotation matrix $\mathbf{W}$ such that for sufficiently large n [4]

$$\max_i \|\hat{\mathbf{X}}_i - \mathbf{W}\mathbf{X}_i\| \leq \frac{Cd^{1/2}\log n}{\sqrt{n\rho_n}}$$

with high probability.

Proof. We have that

$$\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_F = \|(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F + O((n\rho_n)^{-1/2})$$

and hence

$$\begin{aligned}\max_i \|\hat{\mathbf{X}}_i - \mathbf{W}\mathbf{X}_i\| &\leq \frac{1}{\lambda_d^{1/2}(\mathbf{P})} \max_i \|((\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}})_i\| + O((n\rho_n)^{-1/2}) \\ &\leq \frac{d^{1/2}}{\lambda_d^{1/2}(\mathbf{P})} \max_j \|(\mathbf{A} - \mathbf{P})\boldsymbol{u}_j\|_\infty + O((n\rho_n)^{-1/2})\end{aligned}$$

where $\boldsymbol{u}_j$ denotes the j-th column of $\mathbf{U}_{\mathbf{P}}$. Now, for a given j and a given index i, the i-th element of the vector $(\mathbf{A} - \mathbf{P})\boldsymbol{u}_j$ is of the form

$$\sum_k (\mathbf{A}_{ik} - \mathbf{P}_{ik})\boldsymbol{u}_{jk}$$

and once again, by Hoeffding's inequality, the above term can be bounded with high probability by a constant. Taking the union bound over all index i and all columns j of $\mathbf{U}_{\mathbf{P}}$ then yields

$$\max_i \|\hat{\mathbf{X}}_i - \mathbf{W}\mathbf{X}_i\| \leq \frac{Cd^{1/2}}{\lambda_d^{1/2}(\mathbf{P})}\log n + O((n\rho_n)^{-1/2}) \leq \frac{Cd^{1/2}\log n}{(n\rho_n)^{1/2}}.$$

as desired. □

References

A. Athreya, V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang. A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 2015.

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

G. W. Stewart and J.-G. Sun. *Matrix pertubation theory.* Academic Press, 1990.

[4] If $\boldsymbol{x}$ is a vector then $\|\boldsymbol{x}\|$ denotes the l_2 norm of $\boldsymbol{x}$.