Spectrally approximating large graphs with smaller graphs: supplementary material

Andreas Loukas and Pierre Vandergheynst

1 Proof of Theorem 4.1

Proof. The Courant-Fischer min-max theorem for L reads

$$\lambda_k = \min_{\dim(U)=k} \max_{x \in \text{span}(U)} \left\{ \frac{x^\top L x}{x^\top x} \mid x \neq 0 \right\}, \tag{1}$$

whereas the same theorem for L_c reads

$$\begin{aligned} \widetilde{\lambda}_k = \min_{\dim(U_c)=k} \max_{x_c \in \text{span}(U_c)} \left\{ \frac{x_c^\top L_c x_c}{x_c^\top x_c} \mid x_c \neq 0 \right\} &= \min_{\dim(U_c)=k} \max_{Cx \in \text{span}(U_c)} \left\{ \frac{x^\top \Pi L \Pi x}{x^\top \Pi x} \mid x \neq 0 \right\} \\ &= \min_{\dim(U)=k, U \subseteq \text{im}(\Pi)} \max_{x \in \text{span}(U)} \left\{ \frac{x^\top L x}{x^\top x} \mid x \neq 0 \right\}, \end{aligned}$$

where in the second equality we set $L_c = CLC^\top$ and $x_c = Cx$ and the third equality holds since Π is a projection matrix (see Property 1). Notice how, with the exception of the constraint that $x = \Pi x$, the final optimization problem is identical to the one for λ_k, given in (1). As such, the former's solution must be strictly larger (since it is a more constrained problem) and we have that $\widetilde{\lambda}_k \geq \lambda_k$. □

2 Proof of Theorem 3.1

We now proceed to derive the main statement of Theorem 3.1. Our approach will be to control $u_k^\top \widetilde{L} u_k$ through its expectation.

Lemma 2.1. *For any k such that $\lambda_k \leq 0.5 \min_{e_{ij} \in \mathcal{E}} \left\{ \frac{d_i + d_j}{2} + w_{ij} \right\}$ the matrix L_c produced by REC abides to*

$$P\Big(\lambda_k \leq u_k^\top \widetilde{L} u_k \leq \lambda_k (1+\epsilon)\Big) \geq 1 - \frac{\vartheta_k(T,\phi)}{4\epsilon}, \tag{2}$$

where

$$\vartheta_k(T,\phi) = \max_{e_{ij} \in \mathcal{E}} \left\{ P\left(e_{ij} \in \mathcal{E}_F\right) \frac{d_i + d_j + 2(w_{ij} - \lambda_k)}{w_{ij}} \right\}. \tag{3}$$

Proof. Denote by $\Pi^\perp$ the projection matrix defined such that $\Pi + \Pi^\perp = I$. We can then write

$$\begin{aligned} u_k^\top \widetilde{L} u_k = u_k^\top \Pi L \Pi u_k = u_k^\top (I - \Pi^\perp) L (I - \Pi^\perp) u_k &= u_k^\top L u_k - 2 u_k^\top L \Pi^\perp u_k + u_k^\top \Pi^\perp L \Pi^\perp u_k \\ &= \lambda_k - 2\lambda_k u_k^\top \Pi^\perp u_k + u_k^\top \Pi^\perp L \Pi^\perp u_k \end{aligned} \tag{4}$$

Let us now consider term $u_k^\top \Pi^\perp L \Pi^\perp u_k$, where for compactness we set $y = \Pi^\perp u_k$.

$$y^\top L y = \sum_{e_{ij} \in \mathcal{E}} w_{ij} (y(i) - y(j))^2 = \underbrace{\sum_{e_{ij} \in \mathcal{E}_F} w_{ij} (y(i) - y(j))^2}_{T_1} + \underbrace{\sum_{v_i \in \mathcal{V}_F} \sum_{v_j \notin \mathcal{V}_F} w_{ij} y(i)^2}_{T_2}. \tag{5}$$

In the last step above, we exploited the fact that $y(i) = 0$ whenever $v_i \notin \mathcal{V}_F$.

Since $\mathcal{E}_F$ is a matching of $\mathcal{E}$, any coarsening that occurs involves a merging of two adjacent vertices v_i, v_j with $(\Pi x)(i) = (\Pi x)(j)$, implying that for every $e_{ij} \in \mathcal{E}_F$:

$$(y(i) - y(j))^2 = ((\Pi^\perp u_k)(i) + (\Pi u_k)(i) - (\Pi^\perp u_k)(j) - (\Pi u_k)(j))^2 = (x(i) - x(j))^2$$

and therefore

$$T_1 = \sum_{e_{ij} \in \mathcal{E}} b_{ij}\, w_{ij}\, (u_k(i) - u_k(j))^2, \tag{6}$$

with b_{ij} a Bernoulli random variable indicating whether $e_{ij} \in \mathcal{E}_F$. For T_2, notice that the terms in the sum correspond to boundary edges and, moreover, whenever $e_{ij} \in \mathcal{E}_F$ all vertices adjacent to v_i and v_j do not belong in $\mathcal{V}_F$. Another way to express T_2 therefore is

$$\begin{aligned} T_2 &= \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left(y(i)^2 \sum_{e_{i\ell} \in \mathcal{E}, e_{i\ell} \neq e_{ij}} w_{i\ell} + y(j)^2 \sum_{e_{j\ell} \in \mathcal{E}, e_{j\ell} \neq e_{ij}} w_{j\ell} \right) \\ &= \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left(\left(u_k(i) - \frac{u_k(i) + u_k(j)}{2} \right)^2 (d_i - w_{ij}) + \left(u_k(j) - \frac{u_k(i) + u_k(j)}{2} \right)^2 (d_j - w_{ij}) \right) \\ &= \sum_{e_{ij} \in \mathcal{E}} b_{ij}\, w_{ij} (u_k(i) - u_k(j))^2\, \frac{d_i + d_j - 2w_{ij}}{4w_{ij}}. \end{aligned} \tag{7}$$

A similar result also holds for the remaining term $u_k^\top \Pi^\perp u_k = \|\Pi^\perp u_k\|_2^2$ of (4):

$$\begin{aligned} \|\Pi^\perp u_k\|_2^2 &= \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left(\left(u_k(i) - \frac{u_k(i) + u_k(j)}{2} \right)^2 + \left(u_k(i) - \frac{u_k(i) + u_k(j)}{2} \right)^2 \right) \\ &= \sum_{e_{ij} \in \mathcal{E}} b_{ij}\, w_{ij}\, (u_k(i) - u_k(j))^2\, \frac{1}{2w_{ij}}. \end{aligned} \tag{8}$$

If we substitute (6), (7), and (8) into (4) we find that

$$\begin{aligned} u_k^\top \widetilde{L} u_k - \lambda_k &= \sum_{e_{ij} \in \mathcal{E}} b_{ij}\, w_{ij} (u_k(i) - u_k(j))^2 \left(1 + \frac{d_i + d_j - 2w_{ij}}{4w_{ij}} - \frac{\lambda_k}{w_{ij}} \right) \\ &= \frac{1}{4} \sum_{e_{ij} \in \mathcal{E}} b_{ij}\, w_{ij} (u_k(i) - u_k(j))^2 \left(\frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right) \end{aligned} \tag{9}$$

and furthermore

$$\mathbf{E}\left[u_k^\top \widetilde{L} u_k \right] - \lambda_k = \frac{1}{4} \sum_{e_{ij} \in \mathcal{E}} P(e_{ij} \in \mathcal{E}_F) \left(\frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right) w_{ij} (u_k(i) - u_k(j))^2. \tag{10}$$

The expression above is always smaller than

$$\mathbf{E}\left[u_k^\top \widetilde{L} u_k \right] - \lambda_k \leq \frac{\lambda_k}{4} \max_{e_{ij} \in \mathcal{E}} \left\{ P(e_{ij} \in \mathcal{E}_F)\, \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right\} = \frac{\lambda_k}{4}\, \vartheta_k(T, \phi), \tag{11}$$

where $\vartheta_k(T, \phi)$ is a function of the sampling probabilities, the eigenvalue λ_k, and the degree distribution of G. Noticing that (9) is a non-negative random variable whenever $\lambda_k \leq 0.5 \min_{e_{ij} \in \mathcal{E}} \frac{d_i + d_j}{2} + w_{ij}/2$ (the condition is equivalent to $d_i + d_j + 2(w_{ij} - 2\lambda_k) > 0$ implying that $u_k^\top \widetilde{L} u_k - \lambda_k$ is a sum of non-negative terms) and using Markov's inequality, we find that

$$P\left(u_k^\top \widetilde{L} u_k \geq \lambda_k (1 + \epsilon) \right) = P\left(\frac{u_k^\top \widetilde{L} u_k - \lambda_k}{\lambda_k} \geq \epsilon \right) \leq \frac{\mathbf{E}\left[u_k^\top \widetilde{L} u_k \right] - \lambda_k}{\epsilon \lambda_k} \leq \frac{\vartheta_k(T, \phi)}{4\epsilon}, \tag{12}$$

which gives the desired probability bound. $\square$

The RSS constant therefore depends on the probability that each edge e_{ij} is contracted. This is given by:

Lemma 2.2. *At the termination of REC, each edge e_{ij} of $\mathcal{E}$ can be found in $\mathcal{E}_F$ with probability*

$$p_{ij}\,\frac{1-e^{-TP_{ij}}}{P_{ij}} \leq P(e_{ij}\in\mathcal{E}_F) = P(b_{ij}=1) \leq\, p_{ij}\frac{1-e^{-TP_{ij}}}{1-e^{-P_{ij}}} \tag{13}$$

where $p_{ij}=\phi_{ij}/\Phi$ and $P_{ij}=\sum_{e_{pq}\in\mathcal{N}_{ij}} p_{pq}$.

Proof. The event $X_{ij}(t)$ that edge e_{ij} is still in the candidate set $\mathcal{C}$ at the end of the t-th iteration is

$$\begin{aligned} P(X_{ij}(t)) &= P(X_{ij}(t-1)\cap\{e_{ij}\text{ is not selected at }t\}) \\ &= P(X_{ij}(t-1))\prod_{pq\in\mathcal{N}_{ij}}(1-p_{pq}) = \prod_{\tau=1}^{t}\left(\prod_{pq\in\mathcal{N}_{ij}}(1-p_{pq})\right) = a_{ij}^t. \end{aligned} \tag{14}$$

Therefore, the probability that e_{ij} is selected after T iterations can be written as

$$\begin{aligned} P(e_{ij}\in\mathcal{E}_F) &= \sum_{t=1}^{T} P(e_{ij}\text{ is selected at }t) \\ &= \sum_{t=1}^{T} p_{ij}P(X_{ij}(t-1)) \\ &= p_{ij}\sum_{t=0}^{T-1} a_{ij}^t = p_{ij}\,\frac{1-a_{ij}^T}{1-a_{ij}}. \end{aligned} \tag{15}$$

According to the Weierstrass product inequality

$$a_{ij} = \prod_{e_{pq}\in\mathcal{N}_{ij}}(1-p_{pq}) \geq 1-\sum_{e_{pq}\in\mathcal{N}_{ij}} p_{pq} \tag{16}$$

and since the function $f(x)=(1-x^T)/(1-x)$ is monotonically increasing in $[0,1]$ and setting $P_{ij}=\sum_{e_{pq}\in\mathcal{N}_{ij}} p_{pq}$ we have that

$$\frac{1-a_{ij}^T}{1-a_{ij}} \geq \frac{1-(1-P_{ij})^T}{P_{ij}} = \frac{1-e^{\log(1-P_{ij})T}}{P_{ij}} \geq \frac{1-e^{-TP_{ij}}}{P_{ij}},$$

where the last step takes advantage of the series expansion $\log(1-p) = -\sum_{i=1}^{\infty} p^i/i \leq -p$. Similarly, for the upper bound

$$a_{ij} = \prod_{e_{pq}\in\mathcal{N}_{ij}}(1-p_{pq}) = e^{\log\left(\prod_{e_{pq}\in\mathcal{N}_{ij}}(1-p_{pq})\right)} = e^{\sum_{e_{pq}\in\mathcal{N}_{ij}}\log(1-p_{pq})} \leq e^{-\sum_{e_{pq}\in\mathcal{N}_{ij}} p_{pq}} = e^{-P_{ij}} \tag{17}$$

and therefore $\frac{1-a_{ij}^T}{1-a_{ij}} \leq \frac{1-e^{-T\,P_{ij}}}{1-e^{-P_{ij}}}$, as claimed. $\square$

Based on Lemma 2.2, the expression of $\vartheta_k(T,\phi)$ is

$$\begin{aligned} \vartheta_k(T,\phi) &\leq \max_{e_{ij}\in\mathcal{E}}\left\{p_{ij}\frac{1-e^{-TP_{ij}}}{1-e^{-P_{ij}}}\,\frac{d_i+d_j+2(w_{ij}-2\lambda_k)}{w_{ij}}\right\} \\ &\leq \max_{e_{ij}\in\mathcal{E}}\left\{P_{ij}\,\frac{1-e^{-TP_{ij}}}{1-e^{-P_{ij}}}\right\}\max_{e_{ij}\in\mathcal{E}}\left\{\frac{p_{ij}}{P_{ij}}\,\frac{d_i+d_j+2(w_{ij}-2\lambda_k)}{w_{ij}}\right\}. \end{aligned} \tag{18}$$

The function $f(P_{ij}) = P_{ij}\,\frac{1-e^{-TP_{ij}}}{1-e^{-P_{ij}}}$ has a positive derivative in the domain of interest and thus it attains its maximum at $P_{\max}$ when P_{ij} is also maximized. Setting $c_1 = NP_{\max}$ and after straightforward algebraic manipulation, we find:

$$
\begin{aligned}
\vartheta_k(T,\phi) &\le P_{\max}\,\frac{1-e^{-c_1T/N}}{1-e^{-P_{\max}}}\,\max_{e_{ij}\in\mathcal{E}}\left\{\frac{p_{ij}}{P_{ij}}\,\frac{d_i+d_j+2(w_{ij}-2\lambda_k)}{w_{ij}}\right\} \\
&= P_{\max}\,\frac{1-e^{-c_1T/N}}{1-e^{-P_{\max}}}\,\max_{e_{ij}\in\mathcal{E}}\left\{\frac{\phi_{ij}}{\sum_{e_{pq}\in\mathcal{N}_{ij}}\phi_{pq}}\,\left(\frac{\sum_{e_{pq}\in\mathcal{N}_{ij}} w_{pq}}{w_{ij}}+3-\frac{4\lambda_k}{w_{ij}}\right)\right\}. \qquad (19)
\end{aligned}
$$

For any potential function and graph such that $P_{\max} = O(1/N)$, at the limit $c_2 = \frac{P_{\max}}{1-e^{-P_{\max}}}\to 1$ and the above expression reaches

$$
\lim_{N\to\infty}\vartheta_k(T,\phi) \le (1-e^{-c_1T/N})\,\max_{e_{ij}\in\mathcal{E}}\left\{\frac{\phi_{ij}}{\sum_{e_{pq}\in\mathcal{N}_{ij}}\phi_{pq}}\,\left(\frac{\sum_{e_{pq}\in\mathcal{N}_{ij}} w_{pq}}{w_{ij}}+3-\frac{4\lambda_k}{w_{ij}}\right)\right\}. \qquad (20)
$$

The final probability estimate is achieved by using Lemma 2.1 along with the derived bound on $\vartheta_k(T,\phi)$.

3 Proof of Theorem 4.2

We adopt a variational approach and reason that, since

$$
\widetilde{\lambda}_k = \min_U \max_x \left\{\frac{x^\top Lx}{x^\top x},\, x\in U \text{ and } x\neq 0 \,|\, \dim(U)=k \,|\, x=\Pi x\right\}, \qquad (21)
$$

for any matrix Z the following inequality holds

$$
\widetilde{\lambda}_k \le \max_x \left\{\frac{x^\top Lx}{x^\top x} \,|\, x\in \operatorname{span}(Z) \text{ and } x\neq 0\right\} \qquad (22)
$$

as long as the columnspace of Z is of dimension k and does not intersect with the nullspace of Π.

Write $\widetilde{U}_{k-1}$ to denote the $n\times(k-1)$ matrix with the $k-1$ first eigenvectors of L_c and further set $Y_{k-1} = C^\top \widetilde{U}_{k-1}$. We will consider the $N\times k$ matrix Z with

$$
Z(:,i) = \begin{cases} C^\top \widetilde{u}_i & \text{if } i<k \\ z & \text{if } i=k, \end{cases} \quad \text{where} \quad z = \Pi(I - Y_{k-1}Y_{k-1}^\top)u_k. \qquad (23)
$$

It can be confirmed that Z's columnspace meets the necessary requirements. Now, we can express any $x\in\operatorname{span}(Z)$ as $x = Y_{k-1}a + bz = \Pi(Y_{k-1}a+bz)$ with $\|a\|^2 + b^2\|z\|^2 = 1$ and therefore

$$
\begin{aligned}
x^\top Lx &= (a^\top Y_{k-1}^\top + bz^\top)\Pi L\Pi(Y_{k-1}a+bz) \\
&= (a^\top Y_{k-1}^\top + bz^\top)\widetilde{L}(Y_{k-1}a+bz) \\
&= a^\top Y_{k-1}^\top \widetilde{L} Y_{k-1}a + b^2\, z^\top \widetilde{L} z + 2b\, z^\top \widetilde{L} Y_{k-1}a \\
&= a^\top Y_{k-1}^\top \widetilde{L} Y_{k-1}a + b^2\, z^\top \widetilde{L} z, \qquad (24)
\end{aligned}
$$

where in the last step we exploited the fact that, by construction, z does not lie in the span of $\widetilde{U}_{k-1}$ (matrix $\widetilde{L}$ does not rotate its own eigenvectors). Since $Y_{k-1}a \in \operatorname{span}(\widetilde{U}_{k-1})$, the first term in the equation above in bounded by $\widetilde{\lambda}_{k-1}$ and the equality is attained only when $a(k-1)=1$ (in which case b must be zero). By the variational argument however, we are certain that the upper bound in (22) has to be at least as large as $\widetilde{\lambda}_{k-1}$, implying that

$$
\widetilde{\lambda}_k \le \max\left\{\widetilde{\lambda}_{k-1}, \frac{z^\top Lz}{z^\top z}\right\} \qquad (25)
$$

with the two cases corresponding to the choices $a(k-1)=1$ and $b=1$, respectively. In addition, we have that

$$z^\top Lz = u_k^\top (I - Y_{k-1}Y_{k-1}^\top)\Pi L \Pi (I - Y_{k-1}Y_{k-1}^\top) u_k = \sum_{i\geq k} \widetilde{\lambda}_i \, (\widetilde{u}_i^\top C u_k)^2 \tag{26}$$

and $\|z\|^2 = \|\Pi(I - Y_{k-1}Y_{k-1}^\top)u_k\|^2 = \sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2$, meaning that

$$\frac{z^\top \widetilde{L} z}{z^\top z} = \frac{\sum_{i\geq k} \widetilde{\lambda}_i \, (\widetilde{u}_i^\top C u_k)^2}{\sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2} \leq \frac{u_k^\top \widetilde{L} u_k}{\sum_{i\geq k}(\widetilde{u}_i^\top C u_k)^2} \tag{27}$$

and therefore the relation $\widetilde{\lambda}_k \leq \max\left\{\widetilde{\lambda}_{k-1}, (1+\epsilon_k)\frac{\lambda_k}{\sum_{i\geq k}\theta_{ki}}\right\}$ holds whenever $k \leq K$.

4 Proof of Theorem 4.3

Proof. Li's Lemma [2] allows to express ϑ_k based on the squared inner products $(\widetilde{u}_j^\top C u_i)^2$ of the eigenvectors u_i of the Laplacian L and the lifted eigenvectors $C^\top \widetilde{u}_j$ of the coarsened Laplacian L_c.

$$\vartheta_k = \left\|\sin\Theta\left(U_k, C^\top \widetilde{U}_k\right)\right\|_F^2 = \left\|\widetilde{U}_{k^\perp}^\top C U_k\right\|_F^2 = \sum_{i\leq k}\sum_{j>k}(\widetilde{u}_j^\top C u_i)^2 \tag{28}$$

Moreover, the summed RSS inequalities for each $i \leq k$ give:

$$\sum_{i\leq k}(1+\epsilon_i)\lambda_i \geq \sum_{i\leq k} u_i^\top \widetilde{L} u_i = \sum_{i\leq k}\sum_{j=1}^{n} \widetilde{\lambda}_j (\widetilde{u}_j^\top C u_i)^2 = \sum_{j\leq k}\widetilde{\lambda}_j \sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2 + \sum_{j>k}\widetilde{\lambda}_j \sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2. \tag{29}$$

To continue, we use the equality

$$\sum_{2\leq j\leq k}\sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2 = \sum_{2\leq i\leq k}\left(\|\Pi u_i\|_2^2 - \sum_{j>k}(\widetilde{u}_j^\top C u_i)^2\right) \tag{30}$$

based on which

$$\widetilde{\lambda}_{k+1}\sum_{j>k}\sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2 + \widetilde{\lambda}_2 \sum_{2\leq i\leq k}\left(\|\Pi u_i\|_2^2 - \sum_{j>k}(\widetilde{u}_j^\top C u_i)^2\right) \leq \sum_{i\leq k}(1+\epsilon_i)\lambda_i = \sum_{2\leq i\leq k}(1+\epsilon_i)\lambda_i. \tag{31}$$

Our first sinΘ bound is obtained by using the inequality $\lambda_2 \leq \widetilde{\lambda}_2$ and re-arranging the terms:

$$\left\|\sin\Theta\left(U_k, C^\top \widetilde{U}_k\right)\right\|_F^2 \leq \sum_{2\leq i\leq k}\frac{(1+\epsilon_i)\lambda_i - \lambda_2\|\Pi u_i\|_2^2}{\widetilde{\lambda}_{k+1} - \lambda_2} \tag{32}$$

For the second bound, we instead perform the following manipulation

$$\begin{aligned}\sum_{j\leq k}\widetilde{\lambda}_j \sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2 &\geq \sum_{j\leq k}\lambda_j \sum_{i\leq k}(\widetilde{u}_j^\top C u_i)^2 = \sum_{j\leq k}\lambda_j\left(1 - \sum_{i>k}(\widetilde{u}_j^\top C u_i)^2\right) \\ &\geq \sum_{j\leq k}\lambda_j - \lambda_k \sum_{i\leq k}\left(\|\Pi^\perp u_i\|_2^2 + \sum_{j\geq k}(\widetilde{u}_j^\top C u_i)^2\right),\end{aligned} \tag{33}$$

which together with (28) and (29) results to

$$\left\|\sin\Theta\left(U_k, C^\top \widetilde{U}_k\right)\right\|_F^2 \leq \sum_{i\leq k}\frac{(1+\epsilon_i)\lambda_i - \lambda_i + \lambda_k\|\Pi^\perp u_i\|_2^2}{\widetilde{\lambda}_{k+1} - \lambda_k} = \sum_{2\leq i\leq k}\frac{\epsilon_i\lambda_i + \lambda_k\|\Pi^\perp u_i\|_2^2}{\widetilde{\lambda}_{k+1} - \lambda_k}. \tag{34}$$

The final bound is obtained as the minimum of (32) and (34). □

5 Proof of Corollary 5.1

Proof. The proof follows a known argument in the analysis of spectral clustering first proposed by Boutsidis [1] and later adapted by Martin et al. [3]. In particular, these works proved that:

$$\mathcal{F}_K(\Psi, \widetilde{S}^*)^{1/2} \leq \mathcal{F}_K(\Psi, S^*)^{1/2} + 2\,\gamma_K, \tag{35}$$

with $\gamma_K = \|\Psi - \widetilde{\Psi} Q\|_F = \|U_K - C^\top \widetilde{U}_K Q\|_F$ and Q being some unitary matrix of appropriate dimensions. However, as demonstrated by Yu and coauthors [4], it is always possible to find a unitary matrix Q such that

$$\gamma_K^2 = \left\|U_K - C^\top \widetilde{U}_K Q\right\|_F^2 \leq 2 \left\|\sin \Theta\left(U_K, C^\top \widetilde{U}_K\right)\right\|_F^2 \leq 2 \sum_{k=2}^{K} \frac{\epsilon_k \lambda_k + \lambda_K \|\Pi^\perp u_k\|_2^2}{\delta_K} \tag{36}$$

where the last inequality follows from Theorem 4.3 and $\widetilde{\lambda}_{K+1} \geq \lambda_{K+1}$. At this point, we could opt to take a union bound with respect to the events $\{\epsilon_k \geq \epsilon\}$ and $\{\|\Pi^\perp u_k\|_2^2 \geq \epsilon\}$ using the results of Section 3. A more careful analysis however follows the steps of the proof of Theorem 3.1 simultaneously for all terms:

$$\begin{aligned}
\sum_{k=2}^{K} \mathbf{E}[\epsilon_k]\, \lambda_k + \lambda_K \mathbf{E}\left[\|\Pi^\perp u_k\|_2^2\right] &= \sum_{k=2}^{K} \sum_{e_{ij} \in \mathcal{E}} P(e_{ij} \in \mathcal{E}_F) w_{ij} (u_k(i) - u_k(j))^2 \left[\frac{d_i + d_j + 2w_{ij} + 2\lambda_K - 4\lambda_k}{4w_{ij}}\right] \\
&\leq \sum_{k=2}^{K} \lambda_k \max_{e_{ij} \in \mathcal{E}} \left\{P(e_{ij} \in \mathcal{E}_F) \left[\frac{d_i + d_j + 2w_{ij} + 2\lambda_K - 4\lambda_k}{4w_{ij}}\right]\right\} \\
&\leq \sum_{k=2}^{K} \lambda_k \, P_{\max} \frac{1 - e^{-TP_{\max}}}{1 - e^{-P_{\max}}} \max_{e_{ij} \in \mathcal{E}} \left\{\frac{\phi_{ij}}{\sum_{e_{pq} \in \mathcal{N}_{ij}} \phi_{pq}} \left(\frac{\sum_{e_{pq} \in \mathcal{N}_{ij}} w_{pq}}{w_{ij}} + 3 + \frac{2\lambda_K - 4\lambda_k}{w_{ij}}\right)\right\} \\
&= c_2 \frac{1 - e^{-c_1 T/N}}{4} \sum_{k=2}^{K} \lambda_k \max_{e_{ij} \in \mathcal{E}} \left\{\frac{\phi_{ij}}{\sum_{e_{pq} \in \mathcal{N}_{ij}} \phi_{pq}} \left(\frac{\sum_{e_{pq} \in \mathcal{N}_{ij}} w_{pq}}{w_{ij}} + 3 + \frac{2\lambda_K - 4\lambda_k}{w_{ij}}\right)\right\},
\end{aligned} \tag{37}$$

where as before $c_1 = NP_{\max}$ and $c_2 = P_{\max}/(1 - e^{-P_{\max}})$. Assuming further that a heavy-edge potential is used, N is sufficiently large, and G has bounded degree such that $c_1 = 4\varrho_{\max} = O(1)$, the above simplifies to

$$\begin{aligned}
\mathbf{E}\left[\gamma_K^2\right] &\leq \frac{1 - e^{-4\varrho_{\max} T/N}}{2\,\delta_K} \sum_{k=2}^{K} \lambda_k \left(1 + \max_{e_{ij} \in \mathcal{E}} \left\{\frac{3w_{ij} + 2\lambda_K - 4\lambda_k}{\sum_{e_{pq} \in \mathcal{N}_{ij}} w_{pq}}\right\}\right) \\
&\leq \frac{1 - e^{-4\varrho_{\max} T/N}}{2\,\delta_K} \sum_{k=2}^{K} \lambda_k \left(1 + \max_{e_{ij} \in \mathcal{E}} \left\{\frac{6 + 4\lambda_K - 8\lambda_k}{d_{\text{avg}} \varrho_{\min}}\right\}\right).
\end{aligned} \tag{38}$$

The last inequality used the relation $\min_{e_{ij}} \sum_{e_{pq} \in \mathcal{N}_{ij}} w_{pq} = \varrho_{\min} d_{\text{avg}}/2$ and the fact that $w_{ij} \leq 1$. Setting $c_3 = \frac{\sum_{k=2}^K \lambda_k^2}{\sum_{k=2}^K \lambda_k}$, gives

$$\mathbf{E}\left[\gamma_K^2\right] \leq \frac{1 - e^{-4\varrho_{\max} T/N}}{2\,\delta_K} \left(\sum_{k=2}^{K} \lambda_k\right) \left(1 + \frac{6 + 4\lambda_K - 8\,c_3}{d_{\text{avg}} \varrho_{\min}}\right). \tag{39}$$

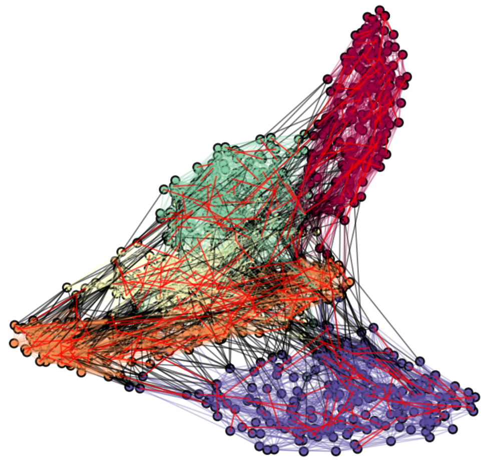
From Markov's inequality, then

$$P\left(\left[\mathcal{F}_K(\Psi, \widetilde{S}^*)^{1/2} - \mathcal{F}_K(\Psi, S^*)^{1/2}\right]^2 \geq \epsilon \sum_{k=2}^{K} \frac{2\lambda_k (1 - e^{-4\varrho_{\max} T/N})}{\delta_K}\right) \leq \frac{1}{\epsilon} \left(1 + \frac{6 + 4\lambda_K - 8\,c_3}{d_{\text{avg}} \varrho_{\min}}\right). \tag{40}$$

The final result follows by the inequality $1 - e^{-4\varrho_{\max} T/N} \leq 4r\,\varrho_{\max}$ (see (9) in the main document). □

6 The MNIST digit graph

The following figure illustrates an instance of the clustering problem we considered. The graph is constructed from $N = 1000$ images, each depicting a digit between 0 and 4 from the MNIST database. Contracted edges are shown in red.



References

[1] C. Boutsidis, P. Kambadur, and A. Gittens. Spectral clustering via the power method-provably. In *International Conference on Machine Learning*, pages 40–48, 2015.

[2] R.-C. Li. Relative perturbation theory:(ii) eigenspace variations. Technical report, 1994.

[3] L. Martin, A. Loukas, and P. Vandergheynst. Fast approximate spectral clustering for dynamic networks. *arXiv preprint arXiv:1706.03591*, 2017.

[4] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2014.