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

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

whereas the same theorem for $L_c$ reads

$$\tilde{\lambda}_k = \min_{\text{dim}(U) = k \in \text{span}(U)} \max_{x_c \in \text{span}(U)} \left\{ \frac{x_c^T L_c x_c}{x_c^T x_c} \mid x_c \neq 0 \right\} = \min_{\text{dim}(U) = k \in \text{span}(U)} \max_{C x \in \text{span}(U)} \left\{ \frac{C^T L C x}{x^T x} \mid x \neq 0 \right\},$$

where in the second equality we set $L_c = C L C^T$ and $x_c = C x$ and the third equality holds since $\Pi$ 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 $\lambda_k$, given in (1). As such, the former’s solution must be strictly larger (since it is a more constrained problem) and we have that $\tilde{\lambda}_k \geq \lambda_k$.

1

2 Proof of Theorem 3.1

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

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

$$\mathbb{P} \left( \lambda_k \leq u_k^T \tilde{L} u_k \leq \lambda_k (1 + \epsilon) \right) \geq 1 - \frac{\vartheta_k(T, \phi)}{4 \epsilon}, \tag{2}$$

where

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

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

$$u_k^T \tilde{L} u_k = u_k^T \Pi L \Pi u_k = u_k^T (I - \Pi^\perp) L (I - \Pi^\perp) u_k = u_k^T L u_k - 2 u_k^T \Pi L^\perp u_k + u_k^T \Pi^\perp \Pi^\perp u_k = \lambda_k - 2 \lambda_k u_k^T \Pi^\perp u_k + u_k^T \Pi^\perp \Pi^\perp u_k \tag{4}$$

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

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

\[
T_1 = \sum_{e_{ij} \in \mathcal{E}} b_{ij} (u_k(i) - u_k(j))^2,
\]

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

\[
T_2 = \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left( \frac{y(i)^2}{\sum_{e_{il} \in \mathcal{E}, e_{il} \neq e_{ij}} w_{il}} + \frac{y(j)^2}{\sum_{e_{jl} \in \mathcal{E}, e_{jl} \neq e_{ij}} w_{jl}} \right)
\]

\[
= \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left( \frac{u_k(i) - u_k(j)}{2} \right)^2 (d_i - w_{ij}) + \left( \frac{u_k(i) + u_k(j)}{2} \right)^2 (d_j - w_{ij})
\]

\[
= \sum_{e_{ij} \in \mathcal{E}} b_{ij} w_{ij} (u_k(i) - u_k(j))^2 \frac{d_i + d_j - 2w_{ij}}{4w_{ij}}.
\]

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

\[
\|\Pi^\perp u_k\|_2^2 = \sum_{e_{ij} \in \mathcal{E}} b_{ij} \left( \left( \frac{u_k(i) - u_k(j)}{2} \right)^2 + \left( \frac{u_k(i) + u_k(j)}{2} \right)^2 \right)
\]

\[
= \sum_{e_{ij} \in \mathcal{E}} b_{ij} w_{ij} (u_k(i) - u_k(j))^2 \frac{1}{2w_{ij}}.
\]

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

\[
u_k^\perp \tilde{L} u_k - \lambda_k = \sum_{e_{ij} \in \mathcal{E}} b_{ij} w_{ij} (u_k(i) - u_k(j))^2 \left( 1 + \frac{d_i + d_j - 2w_{ij} - \lambda_k}{4w_{ij}} \right)
\]

\[
= \frac{1}{4} \sum_{e_{ij} \in \mathcal{E}} b_{ij} w_{ij} (u_k(i) - u_k(j))^2 \left( \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right)
\]

and furthermore

\[
\mathbb{E} \left[ u_k^\perp \tilde{L} u_k \right] - \lambda_k = \frac{1}{4} \sum_{e_{ij} \in \mathcal{E}} P(e_{ij} \in \mathcal{E}_F) \left( \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right) w_{ij} (u_k(i) - u_k(j))^2.
\]

The expression above is always smaller than

\[
\mathbb{E} \left[ u_k^\perp \tilde{L} u_k \right] - \lambda_k \leq \frac{\lambda_k}{4} \max_{e_{ij} \in \mathcal{E}_F} \left\{ P(e_{ij} \in \mathcal{E}_F) \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right\} = \frac{\lambda_k}{4} \vartheta_k(T, \phi),
\]

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

\[
P \left( u_k^\perp \tilde{L} u_k \geq \lambda_k (1 + \epsilon) \right) = P \left( \frac{u_k^\perp \tilde{L} u_k - \lambda_k}{\lambda_k} \geq \epsilon \right) \leq \frac{\mathbb{E} \left[ u_k^\perp \tilde{L} u_k \right] - \lambda_k}{\epsilon \lambda_k} \leq \frac{\vartheta_k(T, \phi)}{4 \epsilon},
\]

which gives the desired probability bound.
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

\[
P_{ij} \frac{1 - e^{-TP_{ij}}}{P_{ij}} \leq P(e_{ij} \in \mathcal{E}_F) = P(b_{ij} = 1) \leq P_{ij} \frac{1 - e^{-TP_{ij}}}{1 - e^{-P_{ij}}}
\]

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

\[
P(X_{ij}(t)) = P(X_{ij}(t-1) \cap \{ e_{ij} \text{ is not selected at } t \})
\]

\[
= P(X_{ij}(t-1)) \prod_{pq \in \mathcal{N}_{ij}} (1 - p_{pq}) = \prod_{\tau=1}^{t} \left( \prod_{pq \in \mathcal{N}_{ij}} (1 - p_{pq}) \right) = a_{ij}^t.
\]

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

\[
P(e_{ij} \in \mathcal{E}_F) = \sum_{t=1}^{T} P(e_{ij} \text{ is selected at } t)
\]

\[
= \sum_{t=1}^{T} p_{ij} P(X_{ij}(t-1))
\]

\[
= p_{ij} \sum_{t=0}^{T-1} a_{ij}^t = p_{ij} \frac{1 - a_{ij}^T}{1 - a_{ij}}.
\]

(15)

According to the Weierstrass product inequality

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

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

\[
1 - a_{ij}^T \geq 1 - \left(1 - P_{ij}\right)^T = \frac{1 - e^{\log(1 - P_{ij})T}}{P_{ij}} \geq \frac{1 - e^{-TP_{ij}}}{P_{ij}},
\]

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

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

(17)

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

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

\[
\vartheta_k(T, \phi) \leq \max_{e_{ij} \in \mathcal{E}} \left\{ p_{ij} \frac{1 - e^{-TP_{ij}}}{1 - e^{-P_{ij}}} \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right\}
\]

\[
\leq \max_{e_{ij} \in \mathcal{E}} \left\{ P_{ij} \frac{1 - e^{-TP_{ij}}}{1 - e^{-P_{ij}}} \right\} \max_{e_{ij} \in \mathcal{E}} \left\{ p_{ij} \frac{d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right\}.
\]

(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 \( P_{ij} \) max when \( P_{ij} \) is also maximized. Setting \( c_1 = NP_{\text{max}} \) and after straightforward algebraic manipulation, we find:

\[
\partial_k(T, \phi) \leq P_{\text{max}} \frac{1 - e^{-c_1T/N}}{1 - e^{-P_{\text{max}}}} \max_{e_{ij} \in E} \left\{ \frac{p_{ij} d_i + d_j + 2(w_{ij} - 2\lambda_k)}{w_{ij}} \right\}
\]

\[
= P_{\text{max}} \frac{1 - e^{-c_1T/N}}{1 - e^{-P_{\text{max}}}} \max_{e_{ij} \in E} \left\{ \frac{\phi_{ij}}{\sum_{e_{pq} \in N_{ij}} \phi_{pq} \left( \frac{\sum_{e_{pq} \in N_{ij}} u_{pq}}{w_{ij}} + 3 - \frac{4\lambda_k}{w_{ij}} \right)} \right\}.
\]

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

\[
\lim_{N \rightarrow \infty} \partial_k(T, \phi) \leq (1 - e^{-c_1T/N}) \max_{e_{ij} \in E} \left\{ \sum_{e_{pq} \in N_{ij}} \phi_{pq} \left( \frac{\sum_{e_{pq} \in N_{ij}} u_{pq}}{w_{ij}} + 3 - \frac{4\lambda_k}{w_{ij}} \right) \right\}.
\]

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

### 3 Proof of Theorem 4.2

We adopt a variational approach and reason that, since

\[
\bar{\lambda}_k = \min_U \max_x \left\{ \frac{x^T L x}{x^T x}, \; x \in U \text{ and } x \neq 0 \right\} \text{dim}(U) = k \mid x = \Pi x \right\},
\]

(21)

for any matrix \( Z \) the following inequality holds

\[
\bar{\lambda}_k \leq \max_x \left\{ \frac{x^T L x}{x^T x} \mid x \in \text{span}(\bar{Z}) \text{ and } x \neq 0 \right\}
\]

(22)

as long as the columnspace of \( Z \) is of dimension \( k \) and does not intersect with the nullspace of \( \Pi \).

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

\[
Z(:, i) = \begin{cases} C^T \bar{u}_i & \text{if } i < k, \\ z & \text{if } i = k, \end{cases} \quad \text{where } z = \Pi(I - Y_{k-1} \bar{Y}_{k-1}^T)u_k.
\]

(23)

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

\[
x^T L x = (a^T Y_{k-1}^T + b \bar{z}^T) \Pi L \Pi(Y_{k-1} \bar{a} + b \bar{z})
\]

\[
= a^T Y_{k-1}^T \bar{L} (Y_{k-1} \bar{a} + b \bar{z})
\]

\[
= a^T Y_{k-1}^T \bar{L} \bar{Y}_{k-1} \bar{a} + b^2 \bar{z}^T \bar{L} \bar{z} + 2b \bar{z}^T \bar{L} Y_{k-1} \bar{a}
\]

\[
= a^T Y_{k-1}^T \bar{L} \bar{Y}_{k-1} \bar{a} + b^2 \bar{z}^T \bar{L} \bar{z},
\]

(24)

where in the last step we exploited the fact that, by construction, \( \bar{z} \) does not lie in the span of \( \bar{U}_{k-1} \) (matrix \( \bar{L} \) does not rotate its own eigenvectors). Since \( Y_{k-1} \bar{a} \in \text{span}(\bar{U}_{k-1}) \), the first term in the equation above is bounded by \( \bar{\lambda}_{k-1} \) 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 \( \bar{\lambda}_{k-1} \), implying that

\[
\bar{\lambda}_k \leq \max \left\{ \bar{\lambda}_{k-1}, \frac{z^T \bar{L} \bar{z}}{\bar{z}^T \bar{z}} \right\}
\]

(25)
with the two cases corresponding to the choices \(a(k - 1) = 1\) and \(b = 1\), respectively. In addition, we have that
\[
zk \, Lz = u_k^T (I - Y_{k-1} Y_{k-1}^T) \Pi \Pi (I - Y_{k-1} Y_{k-1}^T) u_k = \sum_{i \geq k} \bar{\lambda}_i (\tilde{u}_i^T C u_k)^2
\]  
(26)
and \(\|z\|^2 = \|\Pi (I - Y_{k-1} Y_{k-1}^T) u_k\|^2 = \sum_{i \geq k} (\tilde{u}_i^T C u_k)^2\), meaning that
\[
\frac{z^T L z}{z^T z} = \frac{\sum_{i \geq k} \bar{\lambda}_i (\tilde{u}_i^T C u_k)^2}{\sum_{i \geq k} (\tilde{u}_i^T C u_k)^2} \leq \frac{u_k^T L u_k}{\|u_k\|^2}
\]  
(27)
and therefore the relation \(\bar{\lambda}_k \leq \max \left\{ \bar{\lambda}_{k-1}, (1 + \epsilon_k) \frac{\lambda_k}{\sum_{i \geq k} \theta_{ki}} \right\}\) holds whenever \(k \leq K\).

4 Proof of Theorem 4.3

Proof. Li’s Lemma [2] allows to express \(\theta_k\) based on the squared inner products \((\tilde{u}_j^T C u_i)^2\) of the eigenvectors \(u_i\) of the Laplacian \(L\) and the lifted eigenvectors \(C^T \tilde{u}_j\) of the coarsened Laplacian \(L_c\).
\[
\theta_k = \left\| \sin \left( U_k, C^T \tilde{U}_k \right) \right\|_F^2 = \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2, \quad (32)
\]
Moreover, the summed RSS inequalities for each \(i \leq k\) give:
\[
\sum_{i \leq k} (1 + \epsilon_i \lambda_i) \geq \sum_{i \leq k} u_i^T L u_i = \sum_{i \leq k, j = 1}^n \bar{\lambda}_j (\tilde{u}_j^T C u_i)^2 = \sum_{j \leq k} \bar{\lambda}_j \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2 + \sum_{j > k} \bar{\lambda}_j \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2. \quad (29)
\]
To continue, we use the equality
\[
\sum_{2 \leq j \leq k} \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2 = \sum_{2 \leq i \leq k} \left( \|\Pi u_i\|^2 - \sum_{j > k} (\tilde{u}_j^T C u_i)^2 \right) \quad (30)
\]
based on which
\[
\bar{\lambda}_k (1 + \epsilon_i) \lambda_i \geq \bar{\lambda}_2 \sum_{2 \leq i \leq k} \left( \|\Pi u_i\|^2 - \sum_{j > k} (\tilde{u}_j^T C u_i)^2 \right) \quad (31)
\]
Our first \(\sin \Theta\) bound is obtained by using the inequality \(\bar{\lambda}_2 \leq \bar{\lambda}_2\) and re-arranging the terms:
\[
\left\| \sin \left( U_k, C^T \tilde{U}_k \right) \right\|_F^2 \leq \sum_{2 \leq i \leq k} \frac{(1 + \epsilon_i) \lambda_i - \lambda_2 \|\Pi u_i\|^2}{\bar{\lambda}_{k+1} - \bar{\lambda}_2} \quad (32)
\]
For the second bound, we instead perform the following manipulation
\[
\sum_{j \leq k} \bar{\lambda}_j \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2 \geq \sum_{j \leq k} \bar{\lambda}_j \sum_{i \leq k} (\tilde{u}_j^T C u_i)^2 = \sum_{j \leq k} \bar{\lambda}_j \left( 1 - \sum_{i > k} (\tilde{u}_j^T C u_i)^2 \right) 
\geq \sum_{j \leq k} \bar{\lambda}_j - \lambda_k \sum_{i \leq k} \left( \|\Pi^+ u_i\|^2 + \sum_{j > k} (\tilde{u}_j^T C u_i)^2 \right), \quad (33)
\]
which together with (28) and (29) results to
\[
\left\| \sin \left( U_k, C^T \tilde{U}_k \right) \right\|_F^2 \leq \sum_{i \leq k} \frac{(1 + \epsilon_i) \lambda_i - \lambda_i + \lambda_k \|\Pi^+ u_i\|^2}{\bar{\lambda}_{k+1} - \lambda_k} \leq \sum_{2 \leq i \leq k} \frac{\epsilon_i \lambda_i + \lambda_k \|\Pi^+ u_i\|^2}{\bar{\lambda}_{k+1} - \lambda_k}. \quad (34)
\]
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 and later adapted by Martin et al. In particular, these works proved that:

\[ F_K(\Psi, S^*)^{1/2} \leq F_K(\Psi, S^*)^{1/2} + 2\gamma_K, \]

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

\[ \gamma_K^2 = ||U_K - C^\top \tilde{U}_K Q||_F^2 \leq 2 \sin \Theta \left( U_K, C^\top \tilde{U}_K \right) \leq 2 \sum_{k=2}^K c_k \lambda_k + \lambda_K ||\Pi^1 u_k||^2 \]

where the last inequality follows from Theorem 4.3 and \( \lambda_{K+1} \geq \lambda_K + 1 \). At this point, we could opt to take a union bound with respect to the events \( \{ c_k \geq \epsilon \} \) and \( \{ ||\Pi^1 u_k||^2 \geq \epsilon \} \) 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:

\[
\sum_{k=2}^K E[c_k \lambda_k + \lambda_K E[||\Pi^1 u_k||_2^2]] = \sum_{k=2}^K \sum_{e_{ij} \in \mathcal{E}} P(e_{ij} \in \mathcal{E}_F) w_{ij}(u_k(i) - u_k(j))^2 \left[ \frac{d_i + d_j + 2w_{ij} + 2\lambda_K - 4\lambda_k}{4w_{ij}} \right] \\
\leq \sum_{k=2}^K \max_{e_{ij} \in \mathcal{E}} \left\{ P(e_{ij} \in \mathcal{E}_F) \left[ \frac{d_i + d_j + 2w_{ij} + 2\lambda_K - 4\lambda_k}{4w_{ij}} \right] \right\} \\
\leq \sum_{k=2}^K \lambda_k P_{\text{max}} \frac{1 - e^{-\lambda K}}{1 - e^{-\lambda_{\text{max}}}} \max_{e_{ij} \in \mathcal{E}} \left\{ \phi_{ij} \left( \frac{\sum_{e_{pq} \in N_i j} w_{pq}}{w_{ij}} + 3 + \frac{2\lambda_K - 4\lambda_k}{w_{ij}} \right) \right\} \\
= \frac{1 - e^{-c_1 T/N}}{4} \sum_{k=2}^K \lambda_k \max_{e_{ij} \in \mathcal{E}} \left\{ \phi_{ij} \left( \sum_{e_{pq} \in N_i j} w_{pq} / w_{ij} + 3 + \frac{2\lambda_K - 4\lambda_k}{w_{ij}} \right) \right\},
\]

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

\[
E[\gamma_K^2] \leq \frac{1 - e^{-4\theta_{\text{max}} T/N}}{2 \delta K} \sum_{k=2}^K \lambda_k \left( 1 + \max_{e_{ij} \in \mathcal{E}} \left\{ \frac{3w_{ij} + 2\lambda_K - 4\lambda_k}{\sum_{e_{pq} \in N_{i j}} w_{pq}} \right\} \right) \\
\leq \frac{1 - e^{-4\theta_{\text{max}} T/N}}{2 \delta K} \sum_{k=2}^K \lambda_k \left( 1 + \max_{e_{ij} \in \mathcal{E}} \left\{ \frac{6 + 4\lambda_K - 8\lambda_k}{d_{\text{avg}} \theta_{\text{min}}} \right\} \right).
\]

The last inequality used the relation \( \min_{e_{ij}} \sum_{e_{pq} \in N_{i j}} w_{pq} = \theta_{\text{min}} d_{\text{avg}} / 2 \) and the fact that \( w_{ij} \leq 1 \). Setting \( c_3 = \sum_{k=2}^K \lambda_k / \sum_{k=2}^K \lambda_k \), gives

\[
E[\gamma_K^2] \leq \frac{1 - e^{-4\theta_{\text{max}} T/N}}{2 \delta K} \left( \sum_{k=2}^K \lambda_k \right) \left( 1 + \frac{6 + 4\lambda_K - 8\lambda_k}{d_{\text{avg}} \theta_{\text{min}}} \right).
\]

From Markov’s inequality, then

\[
P \left( \left[ F_K(\Psi, S^*)^{1/2} - F_K(\Psi, S^*)^{1/2} \right]^2 \geq \epsilon \sum_{k=2}^K 2\lambda_k (1 - e^{-4\theta_{\text{max}} T/N}) \right) \leq \frac{1}{\epsilon} \left( 1 + \frac{6 + 4\lambda_K - 8\lambda_k}{d_{\text{avg}} \theta_{\text{min}}} \right).
\]

The final result follows by the inequality \( 1 - e^{-4\theta_{\text{max}} T/N} \leq 4\theta_{\text{max}} \) (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.

![MNIST digit graph](image)

References


