Supplement for
"New Insights into Laplacian Similarity Search"

Xiao-Ming Wu[1] Zhenguo Li[2] Shih-Fu Chang[1]
[1]Department of Electrical Engineering, Columbia University
[2]Huawei Noah's Ark Lab, Hong Kong
{xmwu,sfchang}@ee.columbia.edu li.zhenguo@huawei.com

Abstract

This is the supplement for our main paper "New Insights into Laplacian Similarity Search" [3]. Here, we show the proofs of all the theoretical arguments in the main paper.

Proof of Statements in Sec. 2.1: M is positive and symmetric, i.e., $\forall i, j$, $m_{ij} > 0$, and $m_{ij} = m_{ji}$. Regardless of Λ, m_{ii} is always the unique largest element in the i-th column and row of M.

Proof. (a) Since $L + \alpha\Lambda$ is symmetric, $M = (L + \alpha\Lambda)^{-1}$ is symmetric.
(b) Note that

$$\begin{aligned} M &= (L + \alpha\Lambda)^{-1} = (D + \alpha\Lambda - W)^{-1} \\ &= (I - (D + \alpha\Lambda)^{-1}W)^{-1}(D + \alpha\Lambda)^{-1} \\ &= \left(\sum_{k=0}^{\infty}[((D + \alpha\Lambda)^{-1})W]^k\right)(D + \alpha\Lambda)^{-1}, \end{aligned}$$

from which we can see that M is positive since the graph is connected.
(c) Now we show that m_{jj} is the unique largest in its column. Assume, to the contrary, there exists i, j, $i \neq j$, such that $m_{jj} \leq m_{ij}$. Denote $k = \arg\max_{i \neq j} m_{ij}$. Note that M is symmetric and $M > 0$. Let $B = (b_{ij}) := D + \alpha\Lambda - W$. Note that B is symmetric and strictly diagonally dominant, i.e., $\forall k, b_{kk} > \sum_{i \neq k} |b_{ki}|$. By $BM = I$, we have $0 = B(k, :)M(:, j) = \sum_i b_{ki}m_{ij} = b_{kk}m_{kj} + \sum_{i \neq k} b_{ki}m_{ij} \geq b_{kk}m_{kj} - (\sum_{i \neq k} |b_{ki}|)m_{kj} = (b_{kk} - \sum_{i \neq k} |b_{ki}|)m_{kj} > 0$, which contradicts the assumption. □

Proof of Theorem 2.1:
$M = C + E$, where $C = \dfrac{1}{\alpha \sum_i \lambda_i}\mathbf{1}\mathbf{1}^\top$, and $E = \Lambda^{-\frac{1}{2}}\left(\sum_{i=2}^{n}\dfrac{1}{\gamma_i + \alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)\Lambda^{-\frac{1}{2}}$.

Proof. By definition,

$$\begin{aligned} M &= (L + \alpha\Lambda)^{-1} \\ &= \Lambda^{-\frac{1}{2}}(\Lambda^{-\frac{1}{2}}L\Lambda^{-\frac{1}{2}} + \alpha I)^{-1}\Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}}\left(\sum_{i=1}^{n}(\gamma_i + \alpha)\mathbf{u}_i\mathbf{u}_i^\top\right)^{-1}\Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}}\left(\sum_{i=1}^{n}\frac{1}{\gamma_i + \alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)\Lambda^{-\frac{1}{2}} \\ &= \frac{1}{\alpha\sum_i \lambda_i}\mathbf{1}\mathbf{1}^\top + \Lambda^{-\frac{1}{2}}\left(\sum_{i=2}^{n}\frac{1}{\gamma_i + \alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)\Lambda^{-\frac{1}{2}}. \end{aligned}$$

□

Proof of Corollary 2.2: $\lim\limits_{\alpha \to 0} E = \Lambda^{-\frac{1}{2}}\bar{L}^\dagger\Lambda^{-\frac{1}{2}}$.

Proof. It follows from $\bar{L}^\dagger = \sum_{i=2}^{n}\frac{1}{\gamma_i}\mathbf{u}_i\mathbf{u}_i^\top$. □

Proof of Statements in Sec. 2.1:
Ranking by $(h_{ij})_{i=1,\dots,n}$ is equivalent as ranking by the j-th column of $D^{-\frac{1}{2}}L_{sym}^\dagger D^{-\frac{1}{2}}$.

Proof. Let e_i denote the i-th unit vector in $\mathbb{R}^n$. The hitting time that a random walk from vertex i to hit vertex j can be computed by [1]:

$$\begin{aligned} H_{ij} &= d(\mathcal{V})\langle\frac{1}{\sqrt{d_j}}e_j, L_{sym}^\dagger(\frac{1}{\sqrt{d_j}}e_j - \frac{1}{\sqrt{d_i}}e_i)\rangle \\ &= d(\mathcal{V})\left(\frac{1}{d_j}e_j{}^\top L_{sym}^\dagger e_j - \frac{1}{\sqrt{d_id_j}}e_i{}^\top L_{sym}^\dagger e_j\right). \end{aligned}$$

Thus given j, ranking by $(h_{ij})_{i=1,\dots,n}$ is determined by $-\frac{1}{\sqrt{d_id_j}}e_i{}^\top L_{sym}^\dagger e_j$. Denote by $B = (b_{ij}) := D^{-\frac{1}{2}}L_{sym}^\dagger D^{-\frac{1}{2}}$. Then $b_{ij} = \frac{1}{\sqrt{d_id_j}}e_i{}^\top L_{sym}^\dagger e_j$. This shows that ranking by $(h_{ij})_{i=1,\dots,n}$ in ascending order is the same as ranking by $(b_{ij})_{i=1,\dots,n}$ in descending order. Note that a smaller h_{ij} means vertices i and j are closer on the graph. □

Proof of Lemma 3.1: (a) [2] $\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k) = \sum_{j\in\bar{\mathcal{S}}_k} a_{1j}$, (b) $\lim_{\alpha\to 0}\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k) = \lambda(\bar{\mathcal{S}}_k)/\lambda(\mathcal{V})$, $1\le k\le n$.

Proof. (a) Recall that $\mathfrak{f}$ is the first column of $M = (L+\alpha\Lambda)^{-1}$. We have $(L+\alpha\Lambda)\mathfrak{f} = e_1$, which can be written as:

$$\sum_{j\neq 1} w_{1j}(\mathfrak{f}_1 - \mathfrak{f}_j) = 1 - \alpha\lambda_1\mathfrak{f}_1, \quad (1)$$

$$\sum_{j\neq i} w_{ij}(\mathfrak{f}_i - \mathfrak{f}_j) = -\alpha\lambda_i\mathfrak{f}_i, \quad i\neq 1. \quad (2)$$

By Eq. (1) and Eq. (2), we have

$$\begin{aligned}\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k) &= \sum_{i=1}^{k}\sum_{j\neq i} w_{ij}(\mathfrak{f}_i - \mathfrak{f}_j) = 1 - \sum_{i=1}^{k}\alpha\lambda_i\mathfrak{f}_i\\ &= 1 - \sum_{j\in\mathcal{S}_k} a_{1j} = \sum_{j\in\bar{\mathcal{S}}_k} a_{1j}. \quad (3)\end{aligned}$$

Note that in Eq. (3), since $A = (a_{ij}) = (L+\alpha\Lambda)^{-1}\alpha\Lambda$, $\alpha\lambda_i\mathfrak{f}_i = a_{1i}$. We also use the fact that $\sum_j a_{1j} = 1$.
(b) By Theorem 2.1.,

$$\begin{aligned}A &= (L+\alpha\Lambda)^{-1}\alpha\Lambda\\ &= \frac{1}{\sum_i\lambda_i}\mathbf{1}\mathbf{1}^\top\Lambda + \alpha\Lambda^{-\frac{1}{2}}\left(\sum_{i=2}^{n}\frac{1}{\gamma_i+\alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)\Lambda^{\frac{1}{2}}.\end{aligned}$$

Therefore, $\lim_{\alpha\to 0} A = \frac{1}{\sum_i\lambda_i}\mathbf{1}\mathbf{1}^\top\Lambda$. By Eq. (3), we have $\lim_{\alpha\to 0}\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k) = \lambda(\bar{\mathcal{S}}_k)/\lambda(\mathcal{V})$. □

Proof of Theorem 3.4: $\mathcal{R}_{\mathfrak{f}}(\mathcal{S}_c) < 1/(c-1)$.

Proof. Since $\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k) = \sum_{j\in\bar{\mathcal{S}}_k} a_{1j}$ strictly decreases when k increases, $\forall k<c$, $\mathcal{L}_{\mathfrak{f}}(\mathcal{S}_c) < \mathcal{L}_{\mathfrak{f}}(\mathcal{S}_k)$. □

Proof of Theorem 3.5:
(a) If $d_i = b$, $\forall i$, for some constant b, then $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{i}}(\mathcal{S}_c) = \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)$.
(b) Suppose for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{|\mathcal{S}_c\backslash\mathcal{S}_k|} > \frac{d(\bar{\mathcal{S}}_c)}{|\bar{\mathcal{S}}_c|}$. Then $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{i}}(\mathcal{S}_c) > \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)$.
(c) Suppose for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{|\mathcal{S}_c\backslash\mathcal{S}_k|} < \frac{d(\bar{\mathcal{S}}_c)}{|\bar{\mathcal{S}}_c|}$. Then $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{i}}(\mathcal{S}_c) < \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)$.

Proof. (a) It follows from $d(\bar{\mathcal{S}}_k) = b|\bar{\mathcal{S}}_k|$, for $1\le k\le c$.
(b) Since for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\bar{\mathcal{S}}_c)} > \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|}{|\bar{\mathcal{S}}_c|}$, we have $\frac{d(\bar{\mathcal{S}}_k)}{d(\bar{\mathcal{S}}_c)} = \frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)+d(\bar{\mathcal{S}}_c)}{d(\bar{\mathcal{S}}_c)} > \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|+|\bar{\mathcal{S}}_c|}{|\bar{\mathcal{S}}_c|} = \frac{|\bar{\mathcal{S}}_k|}{|\bar{\mathcal{S}}_c|}$.
(c) The proof is similar to that of (b). □

Proof of Lemma 3.6:

$$\lim_{d(\mathcal{S}_c)/d(\bar{\mathcal{S}}_c)\to 0}\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c) = \frac{1}{c-1}.$$

Proof. $\frac{1}{\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)} = \frac{\sum_{k=1}^{c-1} d(\bar{\mathcal{S}}_k)}{d(\bar{\mathcal{S}}_c)} = \frac{\sum_{k=1}^{c-1}(d(\mathcal{S}_c\backslash\mathcal{S}_k)+d(\bar{\mathcal{S}}_c))}{d(\bar{\mathcal{S}}_c)} = c-1+\frac{\sum_{k=1}^{c-1} d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\bar{\mathcal{S}}_c)}$. As $\frac{d(\mathcal{S}_c)}{d(\bar{\mathcal{S}}_c)}\to 0$, we have $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\bar{\mathcal{S}}_c)}\to 0$ for $k<c$, which completes the proof. □

Proof of Lemma 3.7: $\lim_{d(\mathcal{S}_c)/d(\bar{\mathcal{S}}_c)\to\infty}\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c) = 0$, if $d_1 < td(\mathcal{S}_c)$ for a fixed scalar t, $0<t<1$.

Proof. $\frac{1}{\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)} = \frac{\sum_{k=1}^{c-1} d(\bar{\mathcal{S}}_k)}{d(\bar{\mathcal{S}}_c)} \ge \frac{d(\bar{\mathcal{S}}_1)}{d(\bar{\mathcal{S}}_c)} \ge \frac{d(\bar{\mathcal{S}}_1)+d_1-d_1}{d(\bar{\mathcal{S}}_c)} \ge \frac{d(\mathcal{S}_c)-d_1}{d(\bar{\mathcal{S}}_c)} \ge \frac{(1-t)d(\mathcal{S}_c)}{d(\bar{\mathcal{S}}_c)}\to\infty$, as $\frac{d(\mathcal{S}_c)}{d(\bar{\mathcal{S}}_c)}\to\infty$. □

Proof of Theorem 3.8: Suppose for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{|\mathcal{S}_c\backslash\mathcal{S}_k|} < \frac{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}}{|\bar{\mathcal{S}}_c|}$. Then $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{i}}(\mathcal{S}_c) < \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c)$.

Proof. Since for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}} < \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|}{|\bar{\mathcal{S}}_c|}$, we have $\frac{d(\mathcal{S}'\backslash\mathcal{S}_k)+\tau\hat{d}}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}} = \frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)+d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}} < \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|+|\bar{\mathcal{S}}_c|}{|\bar{\mathcal{S}}_c|} = \frac{|\bar{\mathcal{S}}_k|}{|\bar{\mathcal{S}}_c|}$. Therefore, for $1\le k<c$, $\frac{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}}{d(\mathcal{S}'\backslash\mathcal{S}_k)+\tau\hat{d}} > \frac{|\bar{\mathcal{S}}_c|}{|\bar{\mathcal{S}}_k|}$. This proves $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{i}}(\mathcal{S}_c) < \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c)$. □

Proof of Lemma 3.9:

$$\lim_{\max_{i\in\mathcal{S}_c} d_i/\hat{d}\to 0}\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c) = \frac{1}{c-1}.$$

Proof. $\frac{1}{\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c)} = \frac{d(\mathcal{S}'\backslash\mathcal{S}_k)+\tau\hat{d}}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}} = \frac{\sum_{k=1}^{c-1}(d(\mathcal{S}_c\backslash\mathcal{S}_k)+d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d})}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}} = c-1+\frac{\sum_{k=1}^{c-1} d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}}$. As $\frac{\max_{i\in\mathcal{S}_c} d_i}{\hat{d}}\to 0$, we have $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\mathcal{S}'\backslash\mathcal{S}_c)+\tau\hat{d}}\to 0$ for $k<c$, which completes the proof. □

Proof of Theorem 3.10: Suppose for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{|\mathcal{S}_c\backslash\mathcal{S}_k|} > \frac{d(\bar{\mathcal{S}}_c)}{|\mathcal{S}^*\backslash\mathcal{S}_c|+d(\bar{\mathcal{S}}_*)/\hat{d}}$. Then $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c) > \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c)$.

Proof. Since for $1\le k<c$, $\frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)}{d(\bar{\mathcal{S}}_c)} > \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|\hat{d}}{|\mathcal{S}^*\backslash\mathcal{S}_c|\hat{d}+d(\bar{\mathcal{S}}_*)}$, we have $\frac{d(\bar{\mathcal{S}}_k)}{d(\bar{\mathcal{S}}_c)} = \frac{d(\mathcal{S}_c\backslash\mathcal{S}_k)+d(\bar{\mathcal{S}}_c)}{d(\bar{\mathcal{S}}_c)} > \frac{|\mathcal{S}_c\backslash\mathcal{S}_k|\hat{d}+|\mathcal{S}^*\backslash\mathcal{S}_c|\hat{d}+d(\bar{\mathcal{S}}_*)}{|\mathcal{S}^*\backslash\mathcal{S}_c|\hat{d}+d(\bar{\mathcal{S}}_*)} = \frac{|\mathcal{S}^*\backslash\mathcal{S}_k|\hat{d}+d(\bar{\mathcal{S}}_*)}{|\mathcal{S}^*\backslash\mathcal{S}_c|\hat{d}+d(\bar{\mathcal{S}}_*)}$. Therefore, for $1\le k<c$, $\frac{d(\bar{\mathcal{S}}_c)}{d(\bar{\mathcal{S}}_k)} < \frac{|\mathcal{S}^*\backslash\mathcal{S}_c|\hat{d}+d(\bar{\mathcal{S}}_*)}{|\mathcal{S}^*\backslash\mathcal{S}_k|\hat{d}+d(\bar{\mathcal{S}}_*)}$. This proves $\lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{d}}(\mathcal{S}_c) < \lim_{\alpha\to 0}\mathcal{R}_{\mathfrak{h}}(\mathcal{S}_c)$. □

References

[1] U. von Luxburg, A. Radl, and M. Hein. Hitting and commute times in large random neighborhood graphs. *Journal of Machine Learning Research*, 15:1751–1798, 2014. 1

[2] X.-M. Wu, Z. Li, and S.-F. Chang. Analyzing the harmonic structure in graph-based learning. In *NIPS*, 2013. 2

[3] X.-M. Wu, Z. Li, and S.-F. Chang. New insights into laplacian similarity search. In *CVPR*, 2015. 1