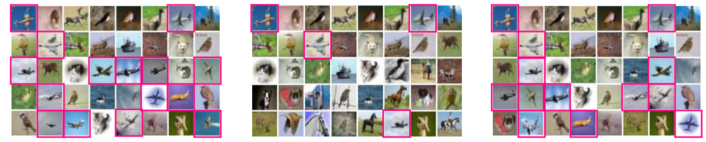
New Insights into Laplacian Similarity Search

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.

(a) $\Lambda = I$, AP = 0.14 (b) $\Lambda = D$, AP = 0.67 (c) $\Lambda = H$, AP = 0.67

Figure 1: Top 40 retrieved images on extended YaleB, with false images highlighted in blue box (query on top left comes from the sparsest cluster).



(a) $\Lambda = I$, AP = 0.27 (b) $\Lambda = D$, AP = 0.17 (c) $\Lambda = H$, AP = 0.27

Figure 2: Top 40 retrieved images on CIFAR-10, with positive images highlighted in magenta box (query on top left comes from the densest cluster).

Similarity metrics are important building blocks of many visual applications such as image retrieval, image segmentation, and manifold learning. Well-known similarity metrics include personalized PageRank, hitting and commute times, and the pseudo-inverse of graph Laplacian. Despite their popularity, the understanding of their behaviors is far from complete, and their use in practice is mostly guided by empirical trials and error analysis. This paper bridges this gap by investigating the fundamental design of similarity metrics on graphs.

We consider a family of similarity metrics in the following form:

$$M = [m_{ij}] \in \mathbb{R}^{n\times n} = (L + \alpha\Lambda)^{-1},$$

where L is the graph Laplacian, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ is a positive diagonal matrix, and α is a positive balancing factor. Note that while L is degenerate, $L + \alpha\Lambda$ is invertible, where Λ acts as a *regularizer* and α controls the degree of regularization. As M is derived based on the graph Laplacian L, we call M the Laplacian-based similarity metric. It is shown that M respects the graph structures when α is small [3], in contrast to metrics such as commute times which reflect only local information in large graphs [1].

A Unifying View. As one contribution of this paper, we show that M converges to a meaningful limit when $\alpha \to 0$ and reproduces popular metrics including hitting times and the pseudo-inverse of graph Laplacian with different regularizer Λ. To see this, we decompose M into a ranking matrix E plus a constant matrix C, as follows:

$$M = \underbrace{\frac{1}{\alpha\sum_i \lambda_i}\mathbf{1}\mathbf{1}^\top}_{C} + \underbrace{\Lambda^{-\frac{1}{2}}\left(\sum_{i=2}^{n}\frac{1}{\gamma_i+\alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)\Lambda^{-\frac{1}{2}}}_{E},$$

where $0 = \gamma_1 < \gamma_2 \le \cdots \le \gamma_n$ and $\mathbf{u}_1, \cdots, \mathbf{u}_n$ are the eigenvalues and the orthonormal eigenvectors of $\Lambda^{-\frac{1}{2}} L \Lambda^{-\frac{1}{2}}$.

1) **Regularizer I.** For $\Lambda = I$ (identity matrix), we have $\lim_{\alpha\to 0} E = L^\dagger$, where $L^\dagger$ denotes the pseudo-inverse of the graph Laplacian. When α is sufficiently small, ranking by M is essentially the same as ranking by the pseudo-inverse of the graph Laplacian, which is widely used in clustering and recommendation.

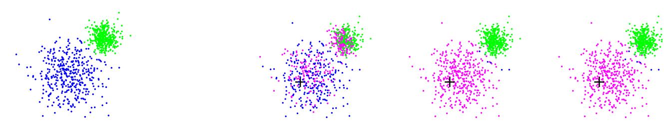
2) **Regularizer D.** For $\Lambda = D = \mathrm{diag}(d_1, \cdots, d_n)$ (degree matrix), we have $\lim_{\alpha\to 0} E = D^{-\frac{1}{2}} L_{sym}^\dagger D^{-\frac{1}{2}}$, where L_{sym} denotes the normalized graph Laplacian. Given a vertex j, let us consider the hitting times h_{ij} from every vertex i to hit j. When α is sufficiently small, ranking by the j-th column of M is essentially the same as ranking by the hitting times $(h_{ij})_{i=1,\ldots,n}$, which is a popular metric in machine learning and social network.

Model Selection. This paper is the first to reveal the important impact of selecting the regularizer Λ in retrieving the local cluster from a seed. We find that regularizer I and D behave complementarily, and each has its own strength and weakness. If the cluster of a query is sparser than surrounding clusters, D performs much better than I (Fig. 3 (c&g)); while if the cluster is denser than surrounding clusters, I is preferred (Fig. 3 (b&f)). Such behaviors of I and D could be intuitively explained under the partially absorbing random walk [2]. Since in practice there is no reliable way to determine the local density in order to select the right model, we propose a new design of Λ that is able to automatically switch between the I mode and the D mode.

Regularizer H. We propose to set $\Lambda = H := \mathrm{diag}(h_1, h_2, \ldots, h_n)$ with

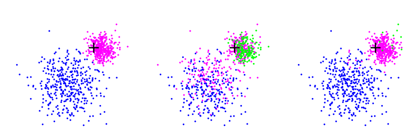
$$h_i = \min(\hat{d}, d_i),\ i = 1, \ldots, n,$$

where $\hat{d}$ is the τ-th largest entry in $(d_1, d_2, \ldots, d_n)$ (e.g., the median degree). H is essentially a mix of I and D – it equals to D at vertices with degree smaller than $\hat{d}$, and stays constant otherwise. With such setting, H behaves like D on sparse clusters where the vertices are of relatively low degrees, and behaves like I while on dense clusters (Fig. 3 (d&h)).



(a) Two Gaussians (b) $\Lambda = I$ (c) $\Lambda = D$ (d) $\Lambda = H$

	Sparse	Dense	All
I	.540	**.999**	.770
D	**.989**	.789	.889
H	**.972**	**.952**	**.962**



(e) MAP (f) $\Lambda = I$ (g) $\Lambda = D$ (h) $\Lambda = H$

Figure 3: Two 20-dimensional Gaussians with variances 1 and 0.16, and 400 points in each. The black cross denotes a query. The top 400 ranked points are highlighted in magenta. (e) Mean average precision (MAP).

In this paper, we develop a unified analysis to characterize the behavior of any regularizer Λ by evaluating the local divergence ratio, which allows for comparing different regularizer including I, D, and H. Our theoretical arguments are justified by rigorous analysis and verified by extensive experiments on image retrieval. The key message of this paper can be summarized visually in the image retrieval results in Figs. 1 and 2. As expected, $\Lambda = D$ and $\Lambda = I$ show distinctive yet complementary behaviors, while $\Lambda = H$ automatically biases to the better of the two. While we only report experiments on image retrieval, our theories and results apply to any visual application that relies on similarity measures and we expect them to guide more visual applications in the future.

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

[2] X.-M. Wu, Z. Li, A.M.-C. So, J. Wright, and S.-F. Chang. Learning with Partially Absorbing Random Walks. In *NIPS*, 2012.

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

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.