6. Appendix

6.1. Proofs and Derivations

Proof. **Theorem 1.** Applying eigenvalue perturbation theory we obtain that $\lambda'_p = \lambda_p + u_p^T(\Delta\hat{M})u_p$ where λ'_p is the eigenvalue of $\hat{M}'$ based on A' obtained after perturbing the graph. Using the fact that $\lambda_p = u_p^T\hat{M}u_p$, and the fact that singular values are equal to the absolute value of the corresponding eigenvalues we obtain the desired result. □

Proof. **Theorem 2.** Denote with e_i the vector of all zeros and a single one at position i. Then, we have $\Delta A = \Delta w_{ij}(e_i e_j^T + e_j e_i^T)$ and $\Delta D = \Delta w_{ij}(e_i e_i^T + e_j e_j^T)$. From eigenvalue perturbation theory (Stewart, 1990), we get: $\lambda'_y \approx \lambda_y + u_y^T(\Delta A - \lambda_y \Delta D)u_y$. Substituting $\Delta A/\Delta D$ concludes the proof. □

We include an immediate result to prove Theorem 3.

Lemma 4. *Consider the generalized eigenvalue problem $Au = \lambda Du$ and suppose that we have the changes in the respective matrices/vectors: $\Delta A, \Delta D$ and $\Delta\lambda$, then the change in the eigenvectors Δu can be expressed as:*

$$\Delta u = -(A - \lambda D)^{+}\Big(\Delta A - \Delta\lambda D - \lambda \Delta D\Big)u$$

Proof. **Theorem 3.** Let ΔA and ΔD be defined as in Theorem 2 and let $\Delta\lambda$ be the change in the eigenvalues as computed in Theorem 2. Plugging these terms in Lemma 4 and simplifying we obtain the result. □

We include an intermediate result to prove Lemmas 2 and 3.

Lemma 5. *λ is an eigenvalue of $D^{-1/2}AD^{-1/2} := A_{norm}$ with eigenvector $\hat{u} = D^{1/2}u$ if and only if λ and u solve the generalized eigen-problem $Au = \lambda Du$.*

Proof. **Lemma 5.** We have $Az = \lambda Dz \implies (Q^{-1}AQ^{-T})(Q^T z) = \lambda(Q^T z)$ for any real symmetric A and any positive definite D, where $D = QQ^T$ using the Cholesky factorization. Substituting the adjacency/degree matrix and using $Q = Q^T = D^{1/2}$ we obtain the result. □

Proof. **Lemma 2.** S is equal to a product of three matrices $S = D^{-1/2}\big(\hat{U}\big(\sum_{r=1}^T \hat{\Lambda}^r\big)\hat{U}^T\big)D^{-1/2}$ where $\hat{U}\hat{\Lambda}\hat{U}^T = D^{-1/2}AD^{-1/2} =: A_{norm}$ is the eigenvalue decomposition of A_{norm} (Qiu et al. (2018)). From Lemma 5 we have the fact that λ is an eigenvalue of $D^{-1/2}AD^{-1/2}$ with eigenvector $\hat{u} = D^{1/2}u$ if and only if λ and u solve the generalized eigen-problem $Au = \lambda Du$. Substituting $\hat{\Lambda} = \Lambda$ and $\hat{U} = D^{1/2}U$ in S, and use the fact that D is diagonal. □

Proof. **Lemma 3.** Following (Qiu et al., 2018), the singular values of S can be bounded by $\sigma_p(S) \leq \frac{1}{d_{min}}\left|\sum_{r=1}^T(\hat{\mu}_{\pi(p)})^r\right|$ where μ are the (standard) eigenvalues of A_{norm}. Using Lemma 5, the same bound applies using the generalized eigenvalues λ_p of A. Now using Theorem 2, we obtain $\tilde{\lambda}'_p$ an approximation of the p-th *generalized* eigenvalue of A'. Plugging it into the singular value bound we obtain: $\sigma_p(S) \leq \frac{1}{d_{min}}\left|\sum_{r=1}^T(\tilde{\lambda}'_{\pi(p)})^r\right|$ which concludes the proof. □

Please note that the permutation π does not need be computed/determined explicitly. In practice, for every $\tilde{\lambda}'_p$, we compute the term $\left|\sum_{r=1}^T(\tilde{\lambda}'_p)^r\right|$. Afterwards, these terms are simply sorted.

6.2. Analysis of Spectral Embedding Methods

Attacking spectral embedding. Finding the spectral embedding is equivalent to the following trace minimization problem:

$$\min_{Z \in \mathbb{R}^{|V|\times K}} Tr(Z^T L_{xy} Z) = \sum_{i=1}^{K}\lambda_i(L_{xy}) = \mathcal{L}_{SC} \quad (4)$$

subject to orthogonality constraints, where L_{xy} is the graph Laplacian. The solution is obtained via the eigen-decomposition of L, with $Z^* = U_K$ where U_K are the K-first eigen-vectors corresponding to the K-smallest eigenvalues λ_i. The Laplacian is typically defined in three different ways: the unnormalized Laplacian $L = D - A$, the normalized random walk Laplacian $L_{rw} = D^{-1}L = I - D^{-1}A$ and the normalized symmetric Laplacian $L_{sym} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}AD^{-1/2} = I - A_{norm}$, where A, D, A_{norm} are defined as before.

Lemma 6 ((von Luxburg, 2007))**.** *λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ is an eigenvalue of L_{sym} with eigenvector $w = D^{1/2}u$. Furthermore, λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ and u solve the generalized eigen-problem $Lu = \lambda Du$.*

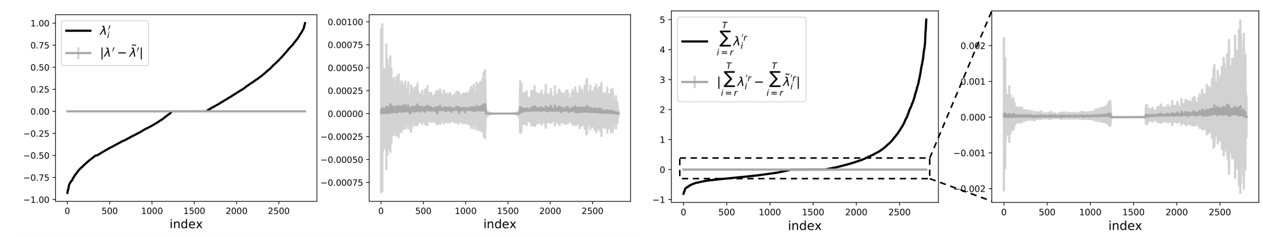
From Lemma 6 we see that we can attack both normalized versions of the graph Laplacian with a single attack strategy since they have the same eigenvalues. It also helps us to do that efficiently similar to our previous analysis (Theorem. 3).

Theorem 4. *Let L_{rw} (or equivalently L_{sym}) be the initial graph Laplacian before performing a flip and λ_y and u_y be any eigenvalue and eigenvector of L_{rw}. The eigenvalue λ'_y of L'_{rw} obtained after flipping a single edge (i,j) is*

$$\lambda'_y \approx \lambda_y + \Delta w_{ij}((u_{yi} - u_{yj})^2 - \lambda_y(u_{yi}^2 + u_{yj}^2)) \quad (5)$$

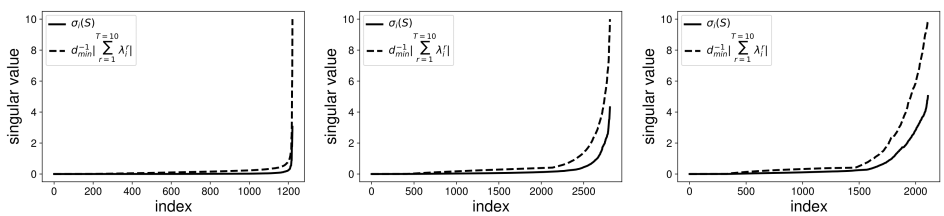
where u_{yi} is the i-th entry of the vector u_y.

Proof. From Lemma 6 we can estimate the change in L_{rw} (or equivalently L_{sym}) by estimating the eigenvalues solving the generalized eigen-problem $Lu = \lambda Du$. Let



(a) Eigenvalues: $|\lambda' - \tilde{\lambda}'|$ (b) Sums of powers of eigenvalues: $\left|\sum_{r=1}^{T} \lambda_i'^r - \sum_{r=1}^{T} \tilde{\lambda}_i'^r\right|$.

Figure 6: Comparison between the true eigenvalues λ' after performing a flip (i.e. doing a full eigen-decomposition) and our approximation $\tilde{\lambda}'$. Since the difference is several orders of magnitude smaller than the eigenvalues (sums of powers of eigenvalues resp.) themselves, we show a "zoomed-in" view (note the difference in the scale on the y-axis). In each subplot on the right side we show the average absolute difference and the standard deviation across the 5K randomly selected flips.



(a) PolBlogs (b) Cora (c) Citeseer

Figure 7: The singular value of S and our upper bound $d_{min}^{-1}|\sum_{r=1}^{T} \lambda_i^r| \geq \sigma_i(S)$ for different graphs.

$\Delta L = L' - L$ be the change in the unnormalized graph Laplacian after performing a single edge flip (i, j) and ΔD be the corresponding change in the degree matrix. Let e_i be defined as before. Then $\Delta L = (1-2A_{ij})(e_i-e_j)(e_i-e_j)^T$ and $\Delta D = (1 - 2A_{ij})(e_i e_i^T + e_j e_j^T)$. Based on the theory of eigenvalue perturbation we have $\lambda_y' \approx \lambda_y + u_y^T(\Delta L - \lambda_y \Delta D)u_y$. Finally, we substitute ΔL and ΔD. □

Using now Theorem 4 and Eq. 4 we finally estimate the loss of the spectral embedding after flipping an edge $\mathcal{L}_{SC}(L'_{rw}, Z) \approx \sum_{p=1}^{K} \lambda_p'$. Note that here we are summing over the K-first *smallest* eigenvalues. We see that spectral embedding and the random walk based approaches are indeed very similar.

Theorem 5. *Let L be the initial unnormalized graph Laplacian before performing a flip and λ_y and u_y be any eigenvalue and eigenvector of L. The eigenvalue λ_y' of L' obtained after flipping a single edge (i, j) can be approximated by:*

$$\lambda_y' \approx \lambda_y - (1 - 2A_{ij})(u_{yi} - u_{yj})^2 \tag{6}$$

Proof. Let $\Delta A = A' - A$ be the change in the adjacency matrix after performing a single edge flip (i, j) and ΔD be the corresponding change in the degree matrix. Let e_i be defined as before. Then $\Delta L = L' - L = (D + \Delta D) - (A + \Delta A) - (D - A) = \Delta D - \Delta A = (1 - 2A_{ij})(e_i e_i^T + e_j e_j^T - (e_i e_j^T + e_j e_i^T))$. Based on the theory of eigenvalue perturbation we have $\lambda_y' \approx \lambda_y + u_y^T(\Delta L)u_y$. Substituting ΔL and re-arranging we get the above results. □

6.3. Approximation Quality

Approximation quality of the eigenvalues. We randomly select 5K candidate edge flips (Cora) and we compare the true eigenvalues λ' after performing a flip (i.e. doing a full eigen-decomposition) and our approximation $\tilde{\lambda}'$ obtained from Theorem 2. We can see in Fig. 6a that the average absolute difference $|\lambda' - \tilde{\lambda}'|$ across the 5K randomly selected flips and the standard deviation are negligible: several orders of magnitude smaller than the eigenvalues themselves. The difference between the terms $|\sum_{r=1}^{T} \lambda_i'^r - \sum_{r=1}^{T} \tilde{\lambda}_i'^r|$ used in Lemma 3 is similarly negligible as shown in Fig. 6b.

Upper bound on the singular values. Lemma 3 shows that $\mathcal{L}_{DW3}$ is an upper bound on $\mathcal{L}_{DW1}$ (excluding the elementwise logarithm). For a better understanding of the tightness of the bound we visualize the true singular values $\sigma_i(S)$ of the matrix S and their respective upper bounds $d_{min}^{-1}|\sum_{r=1}^{T} \lambda_i^r| \geq \sigma_i(S)$ obtained by applying Lemma 3 for all datasets. As we can see in Fig. 7, the gap is different across the different graphs and it is relatively small overall.

These results together (Fig. 6 and Fig. 7) demonstrate that we have obtained a good approximation of both the eigenvalues and the singular values, leading to a good overall approximation of the loss.