Machine Learning in Graph-Based Hilbert Spaces

M. P. Benowitz*
Utah State University, Department of Physics, Logan, UT 84322-4415
(Dated: November 9, 2018)

This note aims to develop a new methodology for the computation of graph distances and kernels in polynomial time. We consider the complex L^2-space of functions of the adjacency matrix. We demonstrate the existence of this space by construction, extend the concept of eigenfunctions to graphs, and present a framework for the construction of Hilbert spaces encoding structural data of the graph.

I. INTRODUCTION

[1]

Graphs are ubiquitous structures. They capture the relations between objects – abstract or concrete. Real-world graphs arise as atoms and their bonds (molecular graphs), proteins and their molecular interactions (protein-protein interaction networks), metabolites and their protein interactions (metabolic networks), and as neurons and their metabolic interactions (neuronal networks). Developing efficient ways to compute similarities between graphs has become of considerable interest in computational biology, machine learning, and chemical informatics. For instance, quantifying how similar two molecules (or proteins) are is a powerful approach in drug (and protein) design. Among the most promising measures of graph similarity are *graph distances* [1] and *graph kernels* [2]. In this short paper, we extend these approaches to the arena of graph-based Hilbert spaces and show their computation in polynomial time.

This paper is organized as follows. In section II we: (1) demonstrate the existence of the complex L^2-space of functions of the adjacency matrix, (2) encode structural information of the graph through linear combinations of eigenfunctions, and (3) compute graph distances and kernels with respect to these linear combinations. In section III we conclude with why the methodology is at the intersection of algebraic graph theory and operator theory and how its analyticity leads to polynomial time complexity.

II. RESULTS

Definition 1. Let $\mathbf{A}$ be the adjacency matrix of a simple undirected graph G. Let λ be the eigenvalues of $\mathbf{A}$. Following the diagonalization of $\mathbf{A}$, we write the diagonal eigenvalue matrix $\mathbf{\Lambda}$ as the sum of its non-negative and negative components, $\mathbf{\Lambda}_+$ and $\mathbf{\Lambda}_-$, respectively. Let γ_i correspond to the diagonal of $\mathbf{\Lambda}_+$ and μ_i to the absolute values of the diagonal of $\mathbf{\Lambda}_-$.

Definition 2. Let the power-law

$$f(r) = r^{\theta/\pi}, \tag{1}$$

where $(\theta, r) \in \mathbb{R}$.

Fact 1. *The power-law as a function of the adjacency matrix is given as*

$$f(\boldsymbol{A}) = \sum_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \Big(f(\gamma_i) + e^{i\theta} f(\mu_i) \Big), \tag{2}$$

where $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^\top$ are the column and row vectors of $\boldsymbol{U}$ and $\boldsymbol{U}^{-1}$ (which diagonalize $\boldsymbol{A}$), respectively.

Definition 3. Let $L^2_{\mathbb{C}}(\mathbf{A})$ denote the complex square-integrable space of functions of the adjacency matrix.

Lemma 1 (**Power Space Theorem**)**.** *The power-law $f(\boldsymbol{A})$ is an element of $L^2_{\mathbb{C}}(\boldsymbol{A})$.*

Proof. Over the $[0, 2\pi]$ interval,

$$\langle f(\mathbf{A}), f(\mathbf{A}) \rangle = \frac{\pi}{2} \sum_i \mathbf{u}_i \mathbf{u}_i^\top \left(\frac{\gamma_i^4 - 1}{\ln \gamma_i} + \frac{\mu_i^4 - 1}{\ln \mu_i} \right). \tag{3}$$

The expression in parenthesis is an indeterminate of the form $0/0$ for $\lambda = 1$. In the limit as $\lambda \to 1$ the expression converges. The above then converges for all λ. Thus, $f(\mathbf{A}) \in L^2_{\mathbb{C}}(\mathbf{A})$. □

Remark 1. For simplicity, we use the $[0, 2\pi]$ interval to demonstrate the validity of the Power Space Theorem (PST). The above converges for any bounded interval in $\mathbb{R}$. We could argue all topological data is contained in the $[0,\ \mathrm{diam}(G) \cdot \pi]$ interval and define the above inner product as such. Alternatively, we can scale $f(r)$ by an attenuation coefficient $\beta^{-\theta/\pi}$ where $\beta > \lambda_{\max}$ and demonstrate convergence over $[0, \infty)$.

Corollary 1. *From the PST it follows:*

$$\langle \delta(\theta - k\pi), f(\mathbf{A}) \rangle = \mathbf{A}^k, \tag{4}$$

where $\delta(\theta - k\pi)$ is the Kronecker delta function and $k \in \mathbb{N}$.

Remark 2. The ij component of $\mathbf{A}^k$ counts the number of walks of length k from vertices i to j. The PST, therefore, generalizes the notion of discrete walks in G to continuous walks in $L^2_{\mathbb{C}}(\mathbf{A})$. The squared norm of $f(\mathbf{A})$ is then the sum of all continuous walks from i to j.

* mayabenowitz@gmail.com
[1] current address: 17901 NE 36th Way, Vancouver, WA

Definition 4. Let $\hat{D}$ be a linear differential operator. We say $f(\mathbf{A})$ is a ***graph eigenfunction of $\hat{\boldsymbol{D}}$*** if

$$\hat{D}f(\mathbf{A}) = \Lambda f(\mathbf{A}), \tag{5}$$

where $\Lambda = \sum_i \Lambda(\lambda_i)$ is a complex scalar. Likewise, we say Λ is a ***graph eigenvalue of $\hat{\boldsymbol{D}}$***.

Example 1. Let $\hat{D} = \dfrac{\partial}{\partial\theta}$. It then follows,

$$\hat{D}f(\mathbf{A}) = \frac{1}{\pi}\sum_i \ln\lambda_i f(\mathbf{A}). \tag{6}$$

Definition 5. Let $\mathcal{H}(G)$ be a Hilbert space of G spanned by the ***graph eigenbasis*** $\mathcal{G} = \{f(\lambda_1), ..., f(\lambda_n)\}$. An ***eigenfunction labeling*** of G assigns every vertex $v_i \in V$ an eigenfunction $f(\lambda_i) \in \mathcal{G}$.

Definition 6 (Degree Feature Extraction)**.** Let $d_i = \deg(v_i)$ and $\alpha_i = d_i/(\sum_i d_i)$. The ***degree span of $\mathcal{G}$*** is defined as

$$\mathrm{span}_{\mathcal{D}}(\mathcal{G}) = \left\{ \psi := \sum_i \alpha_i f(\lambda_i) \;\middle|\; \alpha_i \in \mathbb{Q},\ f(\lambda_i) \in \mathcal{G} \right\}.$$

We denote the ***degree Hilbert space*** of G as $\mathcal{H}_{\mathcal{D}}(G)$. Upon normalization it follows $\psi \in \mathcal{H}_{\mathcal{D}}(G)$.

Fact 2. *The distance between a pair of graphs with respect to $\mathcal{H}_{\mathcal{D}}(G)$ is given as*

$$d_{\mathcal{D}}(G_1, G_2) = ||\psi_1 - \psi_2||. \tag{7}$$

Fact 3. *The graph kernel with respect to $\mathcal{H}_{\mathcal{D}}(G)$ is given as*

$$K_{\mathcal{D}}(G_i, G_j) = \langle \psi_i, \psi_j \rangle . \tag{8}$$

III. CONCLUSION

It is worth emphasizing the techniques of this note lie at the interface of algebraic graph theory and operator theory. Building bridges between these fields opens the door to studying graphs with the tools of continuous mathematics. The advantage in machine learning applications is this: calculations are analytical and thus the time complexity of comparing graphs depends only on the complexity of the chosen matrix multiplication $M(n)$ and feature extraction $F(n)$ algorithms. In general, our method requires $\mathcal{O}\big(M(n) + F(n)\big)$ operations. Intriguingly, for many real-world graphs (e.g., small-world and scale-free networks) the method becomes significantly cheaper with sparse matrix multiplication algorithms.

IV. REFERENCES

[1] J. Bento and S. Ioannidis, *A Family of Tractable Graph Distances*, arXiv:1801.04301v1 (2018)

[2] M. Sugiyama, et al. *graphkernels: R and Python packages for graph comparison*, Bioinformatics, 34(3), 530-532 (2018).