Network Regression with Wasserstein Distances

Alex G. Zalles
Department of Computational Applied Mathematics and Operations Research
Rice University
Houston, TX 77005
agz2@rice.edu

Kai M. Hung
Department of Computer Science
Rice University
Houston, TX 77478
kai.hung@rice.edu

Ann E. Finneran
Department of Biosciences
Rice University
Houston, TX 77478
annie.finneran@rice.edu

Lydia Beaudrot
Department of Biosciences
Rice University
Houston, TX 77478
beaudrot@rice.edu

César A. Uribe
Department of Electrical & Computer Engineering
Rice University
Houston, TX 77478
cauribe@rice.edu

Abstract

We study the problem of network regression, where the graph topology is inferred for unseen predictor values. We build upon recent developments on generalized regression models on metric spaces based on Fréchet means and propose a network regression method using the Wasserstein metric. We show that when representing graphs as multivariate Gaussian distributions, the regression problem in the Wasserstein metric becomes a weighted Wasserstein barycenter problem. In the case of non-negative weights, such a weighted barycenter can be efficiently computed using fixed point iterations. Numerical results show that the proposed approach improves existing procedures by accurately accounting for graph size, randomness, and sparsity in synthetic experiments. Additionally, real-world experiments utilizing the proposed approach result in larger metrics of model fitness, cementing improved prediction capabilities in practice.

1 Introduction

Networks have been shown to be extremely useful in representing complex phenomena, and the modern explosion of data science has come with data usually represented as graphs. Therefore, the ability to perform inference tasks for graph data and other correlated observations has been a focus of recent research [5, 23], with successful applications in ecology [19, 27], brain imaging [25, 27], and trip scheduling [14, 27].

Regression is typically studied in Euclidean spaces, where regressors and outputs are real (possibly high-dimensional) multivariate values [3]. For example, in Linear Regression, relationships between predictor variables and their outputs are quantified through least-squares processes. In the context of graph prediction, prior work has successfully developed regression models using the Frobenius Norm as a metric for graphs [27]. Specifically, the distance between two graphs is defined as the Frobenius norm of the difference between their Laplacians. However, recent results suggest that different graph

Accepted to the Optimal Transport and Machine Learning Workshop at NeurIPS 2023.

representations, and thus different metrics, namely the Wasserstein distance, better encapsulate the graph structure for learning purposes [17].

A graph can represented as an appropriately defined multidimensional Gaussian distribution, and thus, the distance between two graphs is defined as the Wasserstein distance between their two Gaussian representations. It is hypothesized that Wasserstein distances allow for the prioritization of different graph structures, like local or global connectivity, which contributes to a more robust metric [18, 13]. Recent works have studied the average graph problems by finding their uniform weighted average through Wasserstein barycenters [13] or employing learning methods in Gromov-Wasserstein computations [4]. Nevertheless, formally defining regressors in the Wasserstein space is not trivial.

Recent research has enabled and developed the formal definition of regression models in general metric spaces using Fréchet means based on appropriately defined distance measures. Frechet means and variances extend notions of averages and standard deviations to metric spaces, which allows for the definition of regression models beyond Euclidean space [7, 12, 20]. Computationally, regression in Wasserstein space has been recently studied using quantile functions and empirical one-dimensional measures [10, 28].

In this paper, we focus on developing a network regression model where we leverage two main ideas: 1) Fréchet means for regression in Wasserstein spaces, and 2) Network (graph) representation as multidimensional Gaussian distributions. By combining the effective weight function as derived in Fréchet regression models [27, 20] and the Wasserstein metric [17, 18], we empirically show better performance of Wasserstein-based network regression models when compared with traditional Euclidean-based methods.

2 Regression on Network Metric Spaces

Following [27], we consider a random pair $(X, G) \sim F$, where $X \in \mathbb{R}^p$ and $G = (V, E, W)$ is a graph with node set $V = [1, \cdots, n]$, edge set $E \subseteq V \times V$, and $W \in \mathbb{R}^{n \times n}_{\geq 0}$ as the set of edge weights. Moreover, we assume G takes value in a metric space $(\mathcal{G}, d)$, and the marginal distributions F_X and F_G and conditional distributions $F_{X|G}$ and $F_{G|X}$ exist. The Fréchet mean and variance [12] are

$$w_\oplus = \underset{w \in \mathcal{G}}{\arg\min}\, \mathbb{E}[d^2(X, w)], \quad V_\oplus = \mathbb{E}[d^2(X, w_\oplus)].$$

In the classical Euclidean setting, for jointly distributed random variables X and Y, the conditional distribution is $\mathbb{E}[Y \mid X = x] = \arg\min_{y \in \mathbb{R}} \mathbb{E}[(Y - y)^2 \mid X = x]$. The authors in [27] propose a conditional Fréchet mean as a natural extension to network-valued and other metric space-valued responses, where $(Y - y)^2$ is replaced by $d^2(G, w)$ for some w in the metric space. Thus, the corresponding regression function of G given $X = x$ is defined as

$$m(x) := \underset{w \in \mathcal{G}}{\arg\min}\, \mathbb{E}[d^2(G, w)|X = x].$$

Moreover, by characterizing the regression function as a weighted least square problem, the authors in [27] propose a global Fréchet Regression model given $X = x$ as the weighted barycenter

$$m_G(x) := \underset{w \in \mathcal{G}}{\arg\min}\, \mathbb{E}[s_G(X, x) d^2(G, w)].$$

for a weight function $s_G(X, x) = 1 + (X + \mu)^T \Sigma^{-1}(x - \mu)$, which replicates Euclidean regression properties in metric spaces [20]. Here, $\mu = \mathbb{E}[X]$ and $\Sigma = \text{cov}(X)$. Similarly, when a finite set of i.i.d. pairs $(X_i, G_i) \sim F$ for $i = 1, \cdots, n$ is available, the model becomes the empirical regressor

$$\hat{m}_G(x) := \underset{w \in \mathcal{G}}{\arg\min}\, \frac{1}{n} \sum_{i=1}^{n} s_{iG}(X_i, x) d^2(G_i, w), \tag{1}$$

where $s_{iG}(X_i, x) = 1 + (X_i - \bar{X})^T \hat{\Sigma}^{-1}(x - \bar{X})$, $\bar{X} = n^{-1} \sum_{k=1}^{n} X_i$ is the sample mean, and $\hat{\Sigma} = n^{-1} \sum_{k=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$ is the sample covariance matrix. Here, s_{iG} represents the sample weight function. Through incorporating a smoothing kernel [21], we define the Local Regression model [27] to reduce bias from sampling effects on our data distribution, outlined in Appendix A.1. Next, we describe the two metrics d we will study the network regression problem with.

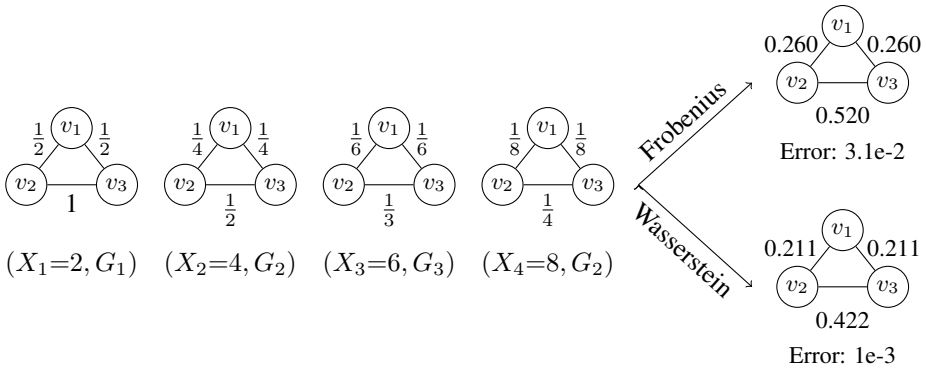

Figure 1: Graphs G_1, G_2, G_3, and G_4 are responses for covariates $X_1 = 2$, $X_2 = 4$, $X_3 = 6$, and $X_4 = 8$, with global results for an input value of $x = 5$. Error from the expected graph is below each output.

Regression on Metric Networks: The space of networks can be defined for many metrics, each encapsulating network difference separately [15]. We will focus on the Frobenius and the Wasserstein metrics and demonstrate their formulations for undirected graphs with real-valued edge weights. The Frobenius Norm is a baseline metric for initial network regression models [27]. However, recent results suggest that Wasserstein distances outperform d_F when comparing network structures [17].

Frobenius Norm [27]: All graphs with positive edge weights w_{ij} have a one-to-one correspondence with their Laplacians which are always positive semi-definite matrices by definition [2]. Given two graphs G_1 and G_2 with their corresponding graph Laplacians, L_1 and L_2 respectively, the Frobenius Norm, $d_F(G_1, G_2)$, between graphs G_1 and G_2 is defined as

$$d_F(G_1, G_2) = d_F(L_1, L_2) = \|L_1 - L_2\|_F = \{tr[(L_1 - L_2)^T (L_1 - L_2)]\}^{1/2}.$$

Wasserstein Distance [17]: The 2-Wasserstein distance between graphs G_1, G_2 is defined as

$$d_W(G_1, G_2) = W_2^2(\nu^{G_1}, \mu^{G_2}) = \inf_{T_\# \nu^{G_1} = \mu^{G_2}} \int_{\mathcal{X}} \|x - T(x)\|^2 d\nu^{G_1},$$

where $\nu^{G_i} = \mathcal{N}(0, L_i^\dagger)$ with $L_i^\dagger$ denoting the pseudo-inverse of Laplacian L_i [11]. Given that ν^{G_1} and μ^{G_2} are zero-mean Gaussian distributions, the authors in [17] showed their 2-Wasserstein Distance has the closed-form

$$W_2^2(\nu^{G_1}, \mu^{G_2}) = Tr(L_1^\dagger + L_2^\dagger) - 2Tr(\sqrt{L_1^{\dagger/2} L_2^\dagger L_1^{\dagger/2}}).$$

Figure 1 shows a toy example where a Wasserstein regressor outperforms the Frobenius distance regressor. Specifically, we have four random pairs $\{X_i, G_i\}_{i=1}^4$ independently observed, with weights shown next to the corresponding edges. Thus, each graph G_i has an associated covariate X_i. We seek to estimate G at $x = 5$ by finding the conditional expectation of G with response to $x = 5$ through the regression models defined above. The unknown ground truth model is $w_{1,2} = w_{1,3} = 1/X$ and $w_{2,3} = 2/X$, thus we expect $w_{1,2} = w_{1,3} = 0.2$ and $w_{2,3} = 0.4$.

3 Computational Aspects of Wasserstein Network Regressions

The empirical regressor in (1) takes the form of an affine combination of convex functions, where the weights are defined by the function s_{kG} since $\mathbb{E}[s_{kG}(X, x)] = 1$. Frobenius regression models thus require solving a convex quadratic problem [27]. Similarly, for Wasserstein regression models, the problem turns into the computation of a weighted Wasserstein barycenter problem.

In general, it follows from [29, Theorem 2.4] that the Wasserstein barycenter of a set of zero-mean Gaussian random distributions $\{\mathcal{N}(0, \Sigma_i)\}_{i=1}^k$ each with non-negative weights λ_i such that $\sum_{i=1}^k \lambda_i = 1$ is a zero-mean Gaussian distribution with covariance $S = \sum_{i=1}^k \lambda_i (S^{\frac{1}{2}} \Sigma_i S^{\frac{1}{2}})^{\frac{1}{2}}$. Moreover, from [29, Theorem 4.2], we know the sequence generated by the fixed-point iteration

$$S \leftarrow S^{-\frac{1}{2}} \left(\sum_{i=1}^k \lambda_i (S^{\frac{1}{2}} L_i^\dagger S^{\frac{1}{2}})^{\frac{1}{2}} \right)^2 S^{-\frac{1}{2}},$$

has the following property: $W_2^2(\mathcal{N}(0,S_t),\mathcal{N}(0,S)) \to 0$ as $t \to 0$. Therefore, as t grows, S_t approaches the covariance of the weighted barycenter of the set of Gaussian distributions. However, by representing each graph G_i as a multivariate Gaussian $\nu^{G_i} = \mathcal{N}(0, L_i^\dagger)$, graph Laplacians have a zero eigenvalue; thus, the conditions in [29, Theorem 4.2] do not hold.

The degeneracy issue of graph Laplacians can be solved by considering the modified fixed-point iteration proposed in [13] that shifts the covariances before iteration,

$$S \leftarrow S^{-\frac{1}{2}} \left(\sum_{i=1}^{k} \lambda_i \left(S^{\frac{1}{2}} \left(L_i + \frac{1}{n} \mathbf{1}_{n,n} \right)^{-1} S^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)^2 S^{-\frac{1}{2}},$$

and then shifts the resulting barycenter back. In our case, the weights are defined as $\lambda_i = s_{iG}(X_i, x)$. Another approach to tackle the non-degeneracy is to consider Entropy Regularized Wasserstein distances [16], where the fixed-point iteration is defined as

$$S = \frac{\epsilon}{4} \sum_{i=1}^{k} s_{iG}(X_i, x) \left(-I + \left(I + \frac{16}{\epsilon^2} S^{\frac{1}{2}} L_i^\dagger S^{\frac{1}{2}} \right)^{\frac{1}{2}} \right).$$

Note that even though $\mathbb{E}[s_{kG}(X, x)] = 1$, the weights as defined for $\hat{m}_G(x)$ can be negative. Convergence of the previously described fixed-point iterations is not guaranteed as one must consider General Fréchet Means [24]. However, in practice, convergence occurs for the studied scenarios. As discussed in [10], the computation of Wasserstein barycenters with possibly negative weights, which turns the problem into an affine combination instead of a convex combination, remains an open problem. We propose these fixed-point iteration methods for computational purposes and leave a connection between Fréchet Means and General Fréchet Means as a future extension.

4 Numerical Analysis

In this section, we show metric comparisons over simulated graphs with various topologies and real data. Appendix A.2 shows additional results for the improved performance of d_W over d_F in cases where networks are large and non-deterministic, with an example test in Fig. 2.

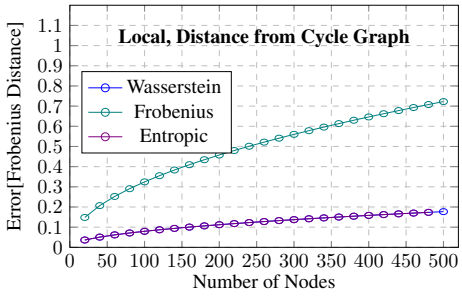

Figure 2: Scalability of the Toy Problem: The Frobenious distance between the predicted graph for $X = 5$ and ground truth when trained on cycle graphs with an increasing number of nodes for Wasserstein, Frobenius, and Entropic Wasserstein-based regressors.

Interpolating Topologies: We considered 5 named graphs: path, cycle, star, wheel, and complete graphs, each with 10 nodes and integer covariates from 1 to 5 in order of increasing connectivity. We then input values from 1 to 5 with a step size of 0.1 to our local models, summarizing the distance from each output graph to the named graphs in Fig. 3. We find that interpolating our sample graphs with d_W is more accurate as output graphs maintain smaller distances to graphs in the sample space. Additionally, graphs at covariates of step size 0.5 for Wasserstein predictions have smaller distances to their adjacent named graphs. In contrast, Frobenius predictions vary greatly from their adjacent named graphs, not maintaining structural patterns. These synthetic results encourage further exploration into non-synthetic applications. Results for each named graph on independent plots can be found in Appendix A.3.

Large Scale Real Data: Initial applications of these regression models focus on taxi data in the Manhattan region as a response to COVID-19 case numbers and weekend indicators, measuring distances between Laplacians with the Frobenius metric. We take rider data, which contains the number of passengers, the pickup and dropoff location, and the day the trip occurs. For a given day, we construct a graph where each node represents one of 13 Manhattan regions [8], and the edge weights represent the number of riders traveling between regions. From these graphs, we compute 172 graph Laplacians from April 12, 2020 to September 30, 2020. We then regress over these Laplacians in response to a binary weekend indicator, equaling 1 if the day is a weekend and 0 if not, and the daily number of COVID-19 cases [26], with an example result for local regression in Fig. 5 with edge weights represented by edge brightness.

Then, we calculate the Fréchet version of the R^2 coefficient, which has similar interpretations of model fitness [20], and is defined for global models as $R^2_{\oplus} = 1 - E[d^2(G, m_G(X))]/V_{\oplus}$. Global results are $\hat{R}^2_{\oplus} = 0.433$ for the Frobenius metric, $\hat{R}^2_{\oplus} = 0.453$ for the Power metric, an adaption of the Frobenius [27], and $R^2_{\oplus} = 0.607$ for the Wasserstein metric. We can see in this system the Wasserstein metric predicts graph structure to an extent the Frobenius doesn't.

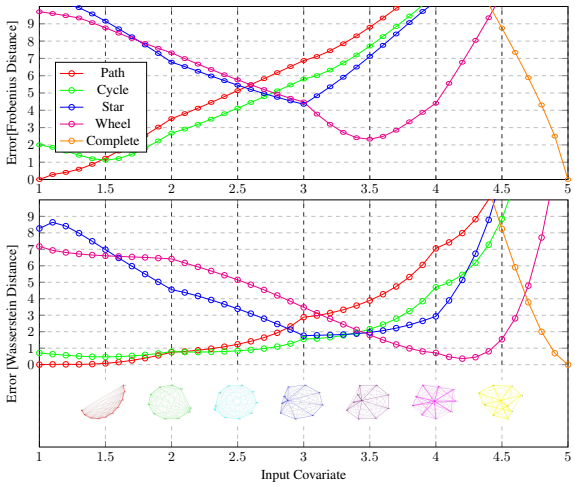

Figure 3: The error from our output graph and each named graph, where each color denotes a different named graph. The output graph is shown below each covariate at 0.5 steps, with a color corresponding to the named graph line. It should be minimal, with non-integer covariates expected to have outputs between their adjacent named graphs. The first plot shows results when modeling with Frobenius distances, and the second shows results with Wasserstein distances.

Additionally, we use 10-fold cross-validation [27] to compute the mean square prediction error with both the Frobenius and Wasserstein metrics. Prediction can occur with either distance, but error computation should be consistent to have comparable accuracy of results. Thus, we have two results: error of Frobenius and Wasserstein predictions measured with the Frobenius distance and error measured with the Wasserstein distance. When averaging over 100 iterations, the MSPE can be seen in Fig.4, leading to 3 main conclusions. Regardless of the metric used for error computation, our methods have smaller MSPE than methods in [27]. However, two particular results are encouraging. First, even when measuring error with the Frobenius distance, the Wasserstein metric is still an improvement over the power metric, which is an adaptation of the Frobenius distance [27] that we would assume to have a smaller error for a similar metric. Secondly, when we compute the error with Wasserstein distances, we see a large decrease, showing the extent of our improved predictions for this case.

Figure 4: Accuracy Relative to Frobenius

Distance used	% MSPE of Frobenius
Power Metric	96.4%
Wasserstein (Prediction) Frobenius (Error)	95.995%
Wasserstein (Prediction) Wasserstein (Error)	86.375%

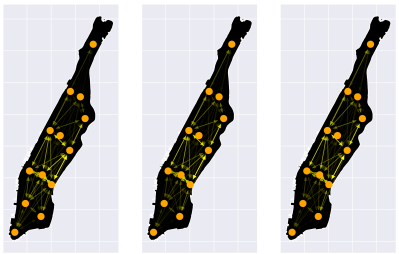

Figure 5: Graph prediction for Taxi Cab ridership on April 12, 2020, with Frobenius (left), Wasserstein (center), and true network (right). Wasserstein output recreates edge weights more accurately, as seen in the lower and upper regions.

5 Conclusion

We provided evidence for the superior performance of Wasserstein distances over the Frobenius norm in graph regression problems via experiments focused on network size, network structure, network variability, and analysis of real-world networks. In all of these instances, Wasserstein models' global and local instances obtain greater accuracy in response to these variables than their Frobenius counterparts. While our models are translatable to other metric spaces or distance measures, the accuracy of computation unique to Wasserstein Regression is vital to its applicability. We hope to motivate future efforts in network prediction by applying the Wasserstein metric to a wider breadth of real-world systems that are larger, more varied, and equally important in understanding occurrences in our world and the connections between them. The main focus of this application is utilizing General Fréchet Means and the Gromov-Wasserstein distance, which has been used in graph prediction previously [4] and generalizes the Wasserstein distance over graphs of differing size. Such extensions would theoretically support our methods over networks with a larger breadth of variance.

References

[1] S. K. Alexander. Food web analysis: An ecosystem approach. *The American Biology Teacher*, 44(3): 186–190, 1982. ISSN 00027685, 19384211. URL http://www.jstor.org/stable/4447458.

[2] E. Andreotti, D. Remondini, G. Servizi, and A. Bazzani. On the multiplicity of laplacian eigenvalues and fiedler partitions. *Linear Algebra and its Applications*, 544:206–222, 2018.

[3] E. R. Berndt and N. E. Savin. Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, 45(5):1263–1277, 1977. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1914072.

[4] L. Brogat-Motte, R. Flamary, C. Brouard, J. Rousu, and F. d'Alché Buc. Learning to predict graphs with fused gromov-wasserstein barycenters. *arXiv*, 2202.03813, 2022.

[5] I. Chami, S. Abu-El-Haija, B. Perozzi, C. Ré, and K. Murphy. Machine learning on graphs: A model and comprehensive taxonomy. *arXiv*, 2005.03675, 2022.

[6] S. Chanpuriya, C. Musco, K. Sotiropoulos, and C. Tsourakakis. On the power of edge independent graph models. In *Advances in Neural Information Processing Systems*, volume 34, pages 24418–24429. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/cc9b3c69b56df284846bf2432f1cba90-Paper.pdf.

[7] Y. Chen and H.-G. Müller. Uniform convergence of local fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *The Annals of Statistics*, 50(3):1573–1592, 2022.

[8] T. . L. Commision. Tlc trip record data. https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page, November 16, 2023.

[9] C. C. I. Dashboard. Change in mean sea levels. https://climatedata.imf.org/datasets/b84a7e25159b4c65ba62d3f82c605855_0/explore, December 16, 2022.

[10] J. Fan and H.-G. Müller. Conditional wasserstein barycenters and interpolation/extrapolation of distributions. *arXiv*, 2107.09218, 2021.

[11] A. Fontan and C. Altafini. On the properties of laplacian pseudoinverses. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5538–5543, 2021. doi: 10.1109/CDC45484.2021.9683525.

[12] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310, 1948.

[13] I. Haasler and P. Frossard. Bures-wasserstein means of graphs. *arXiv*, 2305.19738, 2023.

[14] X. Kong, K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li. Exploring human mobility for multi-pattern passenger prediction: A graph learning framework. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):16148–16160, 2022. doi: 10.1109/TITS.2022.3148116.

[15] I. Kyosev, I. Paun, Y. Moshfeghi, and N. Ntarmos. Measuring distances among graphs en route to graph clustering. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3632–3641, 2020. doi: 10.1109/BigData50022.2020.9378333.

[16] A. Mallasto, A. Gerolin, and H. Q. Minh. Entropy-regularized 2-wasserstein distance between gaussian measures. *arXiv*, 2006.03416, 2020.

[17] H. P. Maretic, M. E. Gheche, G. Chierchia, and P. Frossard. GOT: an optimal transport framework for graph comparison. *CoRR*, abs/1906.02085, 2019. URL http://arxiv.org/abs/1906.02085.

[18] H. P. Maretic, M. E. Gheche, G. Chierchia, and P. Frossard. Fgot: Graph distances based on filters and optimal transport. *arXiv*, 2109.04442, 2021.

[19] I. Noy-Meir. Stability of grazing systems: An application of predator-prey graphs. *Journal of Ecology*, 63 (2):459–481, 1975. ISSN 00220477, 13652745. URL http://www.jstor.org/stable/2258730.

[20] A. Petersen and H.-G. Müller. Fréchet regression for random objects with euclidean predictors. 2019.

[21] C. E. Rasmussen. *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-28650-9. doi: 10.1007/978-3-540-28650-9_4. URL https://doi.org/10.1007/978-3-540-28650-9_4.

[22] P. Ricci, L. Sion, F. Capezzuto, G. Cipriano, G. D'Onghia, S. Libralato, P. Maiorano, A. Tursi, and R. Carlucci. Modelling the trophic roles of the demersal chondrichthyes in the northern ionian sea (central mediterranean sea). *Ecological Modelling*, 444:109468, 2021. ISSN 0304-3800. doi: https://doi.org/10.1016/j.ecolmodel.2021.109468. URL https://www.sciencedirect.com/science/article/pii/S0304380021000405.

[23] E. Richard, S. Gaiffas, and N. Vayatis. Link prediction in graphs with autoregressive features. *Advances in Neural Information Processing Systems*, 25, 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf.

[24] C. Schötz. *The Fréchet Mean and Statistics in Non-Euclidean Spaces*. PhD thesis, Heidelberg University, 2021.

[25] J. Zhang, W. W. Sun, and L. Li. Generalized connectivity matrix response regression with applications in brain connectivity studies. *arXiv*, 1810.03192, 2021.

[26] Y. Zhou. Nyc coronavirus disease 2019 (covid-19) data. https://github.com/nychealth/coronavirus-data, November 16, 2023.

[27] Y. Zhou and H.-G. Müller. Network regression with graph laplacians. *Journal of Machine Learning Research*, 23(320):1–41, 2022. URL http://jmlr.org/papers/v23/22-0681.html.

[28] Y. Zhou and H.-G. Müller. Wasserstein regression with empirical measures and density estimation for sparse data. *arXiv*, 2308.12540, 2023.

[29] P. C. Álvarez Esteban, E. del Barrio, J. A. Cuesta-Albertos, and C. Matrán. A fixed-point approach to barycenters in wasserstein space. *arXiv*, 1511.05355, 2016.

A Appendix

A.1 Metric Regression Local Model Definition

Continued from Section 2, we have the definition of the local Fréchet regression model as

$$m_L(x) := \underset{w\in\Omega}{\arg\min}\, M_L(w,x), \quad M_L(\cdot,x) = \mathbb{E}[s_L(X,x)d^2(G,\cdot)].$$

with its corresponding sample version

$$\hat{m}_L(x) := \underset{w\in\Omega}{\arg\min}\, \hat{M}_L(w,x), \quad \hat{M}_L(\cdot,x) = n^{-1}\sum_{k=1}^{n} s_{kL}(X_k,x)d^2(G_k,\cdot).$$

where $s_{kL}(x,h) = \frac{1}{\hat{\mu}_0 - \hat{\mu}_1^T\hat{\mu}_2^{-1}\hat{\mu}_1} K_h(X_k - x)[1 - \hat{\mu}_1^T\hat{\mu}_2^{-1}(X_k - x)]$, with $\hat{\mu}_j = n^{-1}\sum_{k=1}^{n} K_h(X_k - x)(X_k - x)^j$ for $j = 0, 1, 2$. When applying this model to our toy example from Fig.1, we get the results summarized in Fig.6

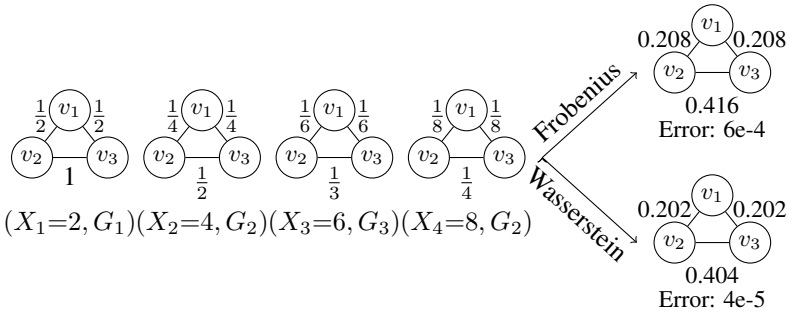

Figure 6: Graphs G_1, G_2, G_3, and G_4 are responses for covariates $X_1 = 2$, $X_2 = 4$, $X_3 = 6$, and $X_4 = 8$, with local results for an input value of $x = 5$. Error from the expected graph is below each output.

A.2 Wasserstein Resilience Examples

We want to verify that the Wasserstein distance is effective over graphs with many nodes, as the Frobenius distance struggles with graph swelling [15]. Thus, we work over graphs of set structure, specifically the path, cycle, star, and complete. For each, we use a sample set of 4 graphs with covariates 2, 4, 6, and 8, with their value denoting edge weights similarly to Fig.1. We then output the graph with covariate 5 and find the error from this graph to the true graph, iterating until said error for the Wasserstein model reaches a certain baseline. As can be seen in Fig.7, our models use the Frobenius, Wasserstein, and Entropic-Wasserstein distances and have two important findings: the Wasserstein error grows slower than the Frobenius, and the Wasserstein and Entropic-Wasserstein have nearly identical results when using a sufficiently small epsilon ($\epsilon = 1e^{-5}$).

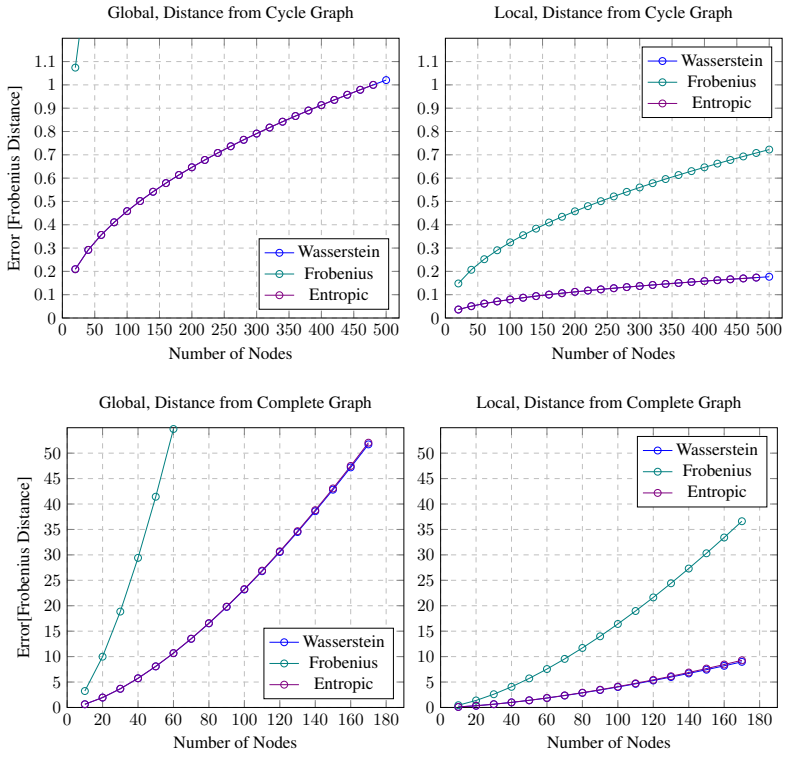

Figure 7: Error between the predicted and ground true graphs for cycle graphs above and complete graphs below. Graphs are shown until the error is greater than 1 for cycle and 50 for complete graphs, which occurs around 500 and 180 nodes, respectively. The final points are labeled with their number of nodes and errors.

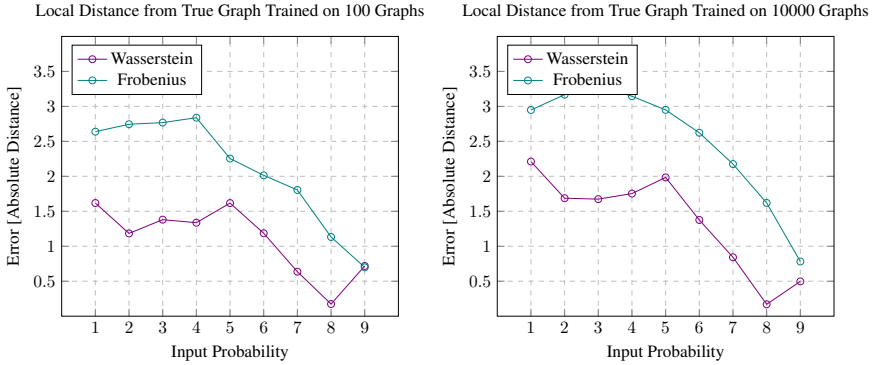

Figure 8: Graphs of the distance between the local predicted graphs and the true graph for Fiedler values from 1 to 9 for a graph with 10 nodes. The Wasserstein barycenter's distance from the expected graph is violet, and the Frobenius barycenter's distance is teal.

The Wasserstein Distance's improved predictions should also apply to less-deterministic graph structures. This randomness is simulated using Erdös-Renyi processes, where networks are generated with a set number of nodes and probability of edge existence [6]. Graphs of 10 nodes are generated, each with random probabilities, which

almost surely create connected graphs, ranging from $\frac{ln(n)}{n}$ to 1. Then, each graph's Fiedler value [2], λ_F, is computed and trained over, correlated with each network's Laplacian. Then, graphs for connectivity ranging from 1 to 9 are output, λ_F is recomputed, the absolute difference between the expected λ_F, the network's covariate, and its true λ_F is found. As shown in Fig.8, training over 100 and 10000 randomly generated graphs, Wasserstein performs significantly better, especially in less-connected graphs with Fielder values ranging from 1 to 5. This indicates that even in cases where our output networks are randomly generated, details about the graph structure are discoverable more accurately with the Wasserstein distance than the Frobenius.

A.3 Direct Metric Comparison for Varied Topologies

In this section, we separate Fig.3 into 5 different figures, comparing performance for each named graph.

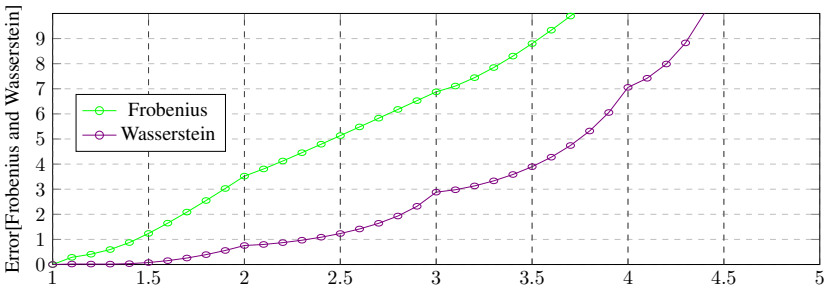

Figure 9: An excerpt of Fig.3 but only considering the Path Graph, comparing the performance of the Frobenius and Wasserstein predictions

For the path graph in Fig.9, the main performance difference occurs in the higher predictor inputs, where the Wasserstein predictions take longer for the error to increase rapidly than the Frobenius, indicating adherence to sample graph structure in the methods.

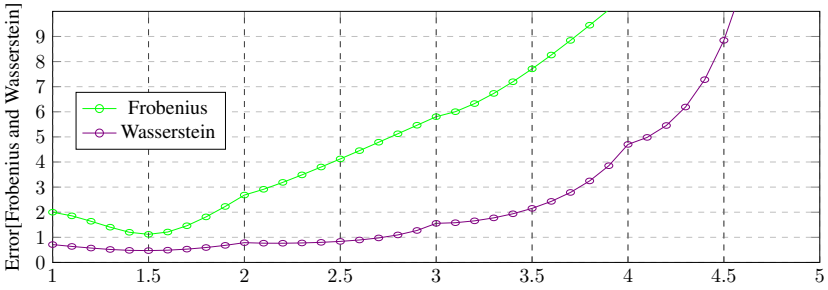

Figure 10: An excerpt of Fig.3 but only considering the Cycle Graph, comparing the performance of the Frobenius and Wasserstein predictions

When focusing on the cycle graph in Fig.10, the results are almost identical to the path, except we now see the trough of each line occurring at a predictor of 1.5. When we consider the similarity between the path and cycle graphs, this location makes sense, and reassuringly, the variance between the error at predictors 1.5 and 1 and 2 is smaller for the Wasserstein than Frobenius, again suggesting a greater adherence to graph structure and slower error growth.

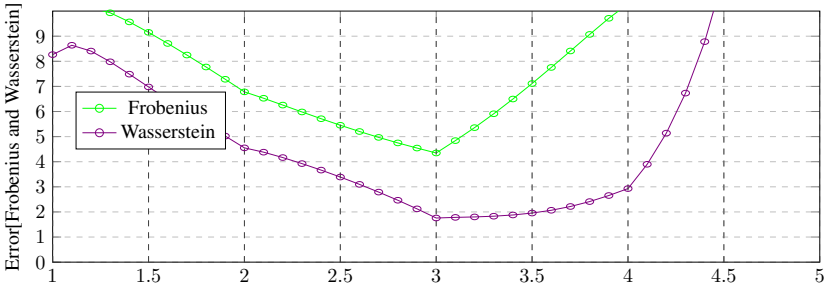

Figure 11: An excerpt of Fig.3 but only considering the Star Graph, comparing the performance of the Frobenius and Wasserstein predictions

In Fig.11, we see the star graph errors for the metrics have the trough at the true predictor for the star graph and continued trends from earlier of smaller errors and error growth from more distant graphs.

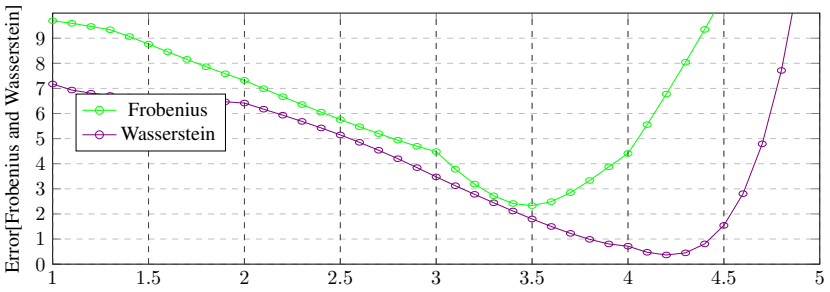

Figure 12: An excerpt of Fig.3 but only considering the Wheel Graph, comparing the performance of the Frobenius and Wasserstein predictions

The Wasserstein predictions improved performance for the wheel in Fig.12, in particular because the trough of each line occurs at different inputs for each distance. For the Frobenius, the trough happens around 0.5 before the true predictor for the wheel graph. However, the Wasserstein's trough occurs around 0.2 after the wheel graph's true predictor, indicating that our model's closest graph to the wheel graph is produced at a closer covariate when using the Wasserstein distance.

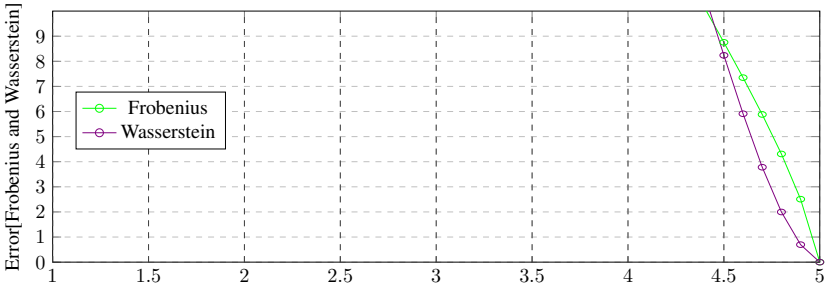

Figure 13: An excerpt of Fig.3 but only considering the Complete Graph, comparing the performance of the Frobenius and Wasserstein predictions

Finally, the performances are comparable for the complete graph in Fig.13, as the predictions only approach the complete graph at very large covariates.

A.4 Ecological Applications

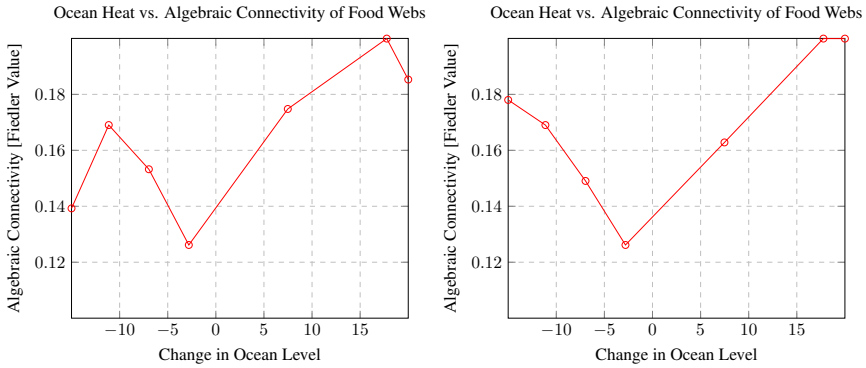

Figure 14: These graphs represent the relationship between the changing ocean level and the Fiedler Value of our networks. There is no clear correlation between the two variables for both the global case on the left and the local one on the right.

An important application of this network analysis lies in ecology and conservation. Environments can be modeled as food webs, which are graphical structures where nodes and predation patterns represent organisms

represented by edges. Food webs thus encapsulate the flow of energy through an ecosystem, and by analyzing the response of graph robustness to changing environmental factors, we can understand the future impact of global warming on the strength of our ecosystems. [1]

Our basic food webs come from [22], with sample marine life in 1995, 2005, and 2015. In these years, we also utilized ocean temperature data from the National Oceanic and Atmospheric Administration [9], creating a model where our sample set was these 3 Laplacians with their covariates. To adapt the food webs for our needs, we removed any self-loops in the Laplacians, assuming that species do not prey on themselves, and we then added the Laplacian to its transpose to transform the directed food web into a symmetric undirected version. We then trained our model and found the local regression model output on covariates less than, between, and greater than our sample covariates. The results are summarized in Fig.14

As can be seen, there is no direct correlation between the changing ocean level and the connectivity of our graphs. This can be for various reasons, but it indicates that further exploration in this field is necessary, specifically over larger data sets and with more impacting variables on the ecosystems, to understand the connections between our changing world and the organisms that live in it.