SEMI-PARAMETRIC GRAPH KERNEL-BASED RECONSTRUCTION[†]

Vassilis N. Ioannidis[1,2], Athanasios N. Nikolakopoulos[2], and Georgios B. Giannakis[1,2]

Dept. of ECE[1] and Digital Technology Center[2], University of Minnesota, USA
Emails: {ioann006, anikolak, georgios}@umn.edu

ABSTRACT

Signal reconstruction over graphs arises naturally in diverse science and engineering applications. Existing methods employ either parametric or nonparametric approaches based on graph kernels. Although the former are adequate when the signals of interest adhere to postulated models, their performance degrades rapidly under model mismatch. Nonparametric alternatives on the other hand are flexible, but not as parsimonious in capturing prior information. Targeting a hybrid "sweet spot," the present contribution advocates an efficient semi-parametric approach capable of incorporating known signal structure without sacrificing the flexibility of the overall model. Numerical tests on synthetic as well as real data corroborate that the novel method leads to markedly improved signal reconstruction performance.

***Index Terms*—** graph kernel, graph signal processing, semi-parametric inference

1. INTRODUCTION

Numerous applications arising in diverse disciplines involve inference over networks. Modeling nodal attributes as functions (or signals) that take values over the vertices of the underlying graph, allows the associated inference tasks to benefit from leveraging node dependencies captured by the graph topology. This approach has been shown to yield promising results in several application areas [1, 2].

In many real settings, often one needs to work with limited nodal observations due to inherent restrictions particular to the inference task at hand. In social networks, for example, individuals may be reluctant to share personal information; in sensor networks the nodes may report observations sporadically in order to save energy; in brain networks acquiring nodal samples may involve invasive procedures (e.g. electrocorticography). In this context, a frequently encountered challenge that has received growing attention recently [1], pertains to reconstructing the graph signal on unobserved vertices, given its values over a limited subset of observed ones.

Traditional approaches – falling under the umbrella term of signal processing on graphs [1,3] – either adopt a *bandlimited* model, which postulates that the signal of interest lies in a B-dimensional subspace related to the graph topology [4–6], or they assume that the signal can be sparsely represented by an overcomplete dictionary [7]. Nonparametric techniques rely on *graph kernels* [8, 9] which allow them to also accommodate nonbandlimited signals, upon selecting an appropriate kernel [9]. The performance of algorithms in the first category is limited by how well the signals actually adhere to the selected model. Nonparametric models on the other hand, offer flexibility and robustness but they cannot readily incorporate information available a priori – a fact that could limit their performance especially in face of scarce nodal samples.

To address the aforementioned limitations, this paper develops a *semi-parametric* approach whereby the signal of interest is modeled as the superposition of a parametric and a nonparametric component. While the former leverages side information, the latter accounts for deviations from the parametric part, and can also promote smoothness using *graph kernels*. We introduce two efficient estimators with complementary strengths, and adopt a dimensionality reduction technique to ensure affordable complexity in large-scale applications. Finally, we perform tests using both synthetic and real data, to corroborate the merits of our approach for reliable reconstruction even based on a minimal number of samples.

Notation: Scalars are denoted using lowercase, column vectors by bold lowercase, and matrices by bold uppercase fonts. Superscripts ^{T} and † respectively denote transpose and pseudo-inverse; $\mathbf{I}_N$ the $N \times N$ identity matrix; $\mathbf{1}_N$ the $N \times 1$ all-one vector; and $\mathrm{diag}\,\{\mathbf{x}\}$ corresponds to a diagonal matrix with the entries of $\mathbf{x}$ on its diagonal. Finally, if $\mathbf{A}$ is a matrix and $\mathbf{x}$ a vector, then $||\mathbf{x}||^2_{\mathbf{A}} := \mathbf{x}^T\mathbf{A}^{-1}\mathbf{x}$ and $||\mathbf{x}||^2_2 := \mathbf{x}^T\mathbf{x}$.

2. MODELING AND PROBLEM FORMULATION

Consider an undirected graph $\mathcal{G} := (\mathcal{V}, \mathbf{A})$, where $\mathcal{V} := \{v_1, \ldots, v_N\}$ is the vertex set, and $\mathbf{A}$ is the symmetric entry-wise nonegative $N \times N$ adjacency matrix, whose (n, n')-th entry denotes the edge weight between vertices v_n and $v_{n'}$. We assume that $\mathcal{G}$ has no self-loops, meaning $(\mathbf{A})_{n,n} = 0,\ \forall v_n \in \mathcal{V}$. The Laplacian matrix of $\mathcal{G}$ is $\mathbf{L} := \mathbf{D} - \mathbf{A}$, with $(\mathbf{D})_{n,n} := \sum_{m=1}^{N}(\mathbf{A})_{n,m}$ and $(\mathbf{D})_{n,n'} := 0$ if $n \neq n'$; matrix $\mathbf{L}$ is known to be positive semidefinite [8].

A real-valued signal on a graph is a function $f : \mathcal{V} \to \mathbb{R}$

[†]The work in this paper was supported by grants NSF 1514056 and ARO W911NF-15-1-0492

that can be compactly represented by the $N \times 1$ vector $\mathbf{f} := [f(v_1), \ldots, f(v_N)]^T$. At each sampled node v_{n_s}, a measurement $y_s = f(v_{n_s}) + e_s$, $s = 1, \ldots, S$ is collected, where $\{e_s\}_{s=1}^S$ represents noise, and $1 \leq n_1 < \ldots < n_S \leq N$ are the indices of the observed vertices. Upon defining $\mathbf{e} := [e_1, \ldots, e_S]^T$, and $\mathbf{y} := [y_1, \ldots, y_S]^T$ it follows that

$$\mathbf{y} = \mathbf{S}\mathbf{f} + \mathbf{e} \tag{1}$$

where $\mathbf{S}$ is an $S \times N$ sampling matrix with all zeros except for the entries (s, n_s), $s = 1, \ldots, S$, which contain ones.

Function f is modeled as the superposition $f = f_{\text{P}} + f_{\text{NP}}$, or, in vector form

$$\mathbf{f} = \mathbf{f}_{\text{P}} + \mathbf{f}_{\text{NP}} \tag{2}$$

where $\mathbf{f}_{\text{P}} := [f_{\text{P}}(v_1), \ldots, f_{\text{P}}(v_N)]^T$, and $\mathbf{f}_{\text{NP}} := [f_{\text{NP}}(v_1), \ldots, f_{\text{NP}}(v_N)]^T$. The parametric term $f_{\text{P}}(v) := \sum_{m=1}^M \beta_m b_m(v)$ captures the known signal structure via the basis $\mathcal{B} := \{b_m\}_{m=1}^M$, while the nonparametric term f_{NP} belongs to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, which accounts for deviations from the span of $\mathcal{B}$. The goal of this paper is efficient and reliable estimation of $\mathbf{f}$ given $\mathbf{y}$, $\mathbf{S}$, $\mathcal{B}$, $\mathcal{H}$ and $\mathbf{A}$.

Remark 1 Decomposing f as in (2) is well motivated in certain applications. Consider for instance an employment-oriented social network like LinkedIn, and let the goal be to estimate the salaries of all users given information about the salaries of a few. Clearly, besides network connections, exploiting available information regarding the users' education level and work experience could benefit the reconstruction task. Another application where this decomposition fits nicely, is in recommender systems. Inferring preference scores for every item, given the users' feedback about particular items, could be cast as a signal reconstruction problem over the item correlation graph. Exploiting side information about the items, is known to alleviate limitations of pure collaborative filtering techniques, leading to considerably improved recommendation performance [10, 11]. In our setup, the item attributes can be used to create a parametric base capturing the user's coarse level preferences.

3. SEMI-PARAMETRIC RECONSTRUCTION

This section introduces our semi-parametric approach. Specifically, Sec. 3.1 reviews the RKHS for graph functions and Sec. 3.2 presents two semi-parametric estimators.

3.1. Reproducing kernel Hilbert spaces on graphs

An RKHS is a space of functions $h : \mathcal{V} \to \mathbb{R}$ expressed in terms of a kernel function $\kappa : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ as

$$\mathcal{H} := \left\{ h : h(v) = \sum_{n=1}^N \alpha_n \kappa(v, v_n),\ \alpha_n \in \mathbb{R} \right\} \tag{3}$$

Kernel name	**Function**	**Parameters**
Diffusion [8, 14]	$r(\lambda) = \exp\{\sigma^2\lambda/2\}$	σ^2
Laplacian [1, 8]	$r(\lambda) = 1 + \sigma^2\lambda$	σ^2
Bandlimited [9]	$r(\lambda) = \begin{cases} 1/\beta, & \lambda \leq \lambda_{\max} \\ \beta, & \text{otherwise} \end{cases}$	$\beta > 0, \lambda_{\max}$

Table 1: Common spectral weight functions.

where $\kappa(v_n, v_{n'})$ captures the similarity between vertices v_n and $v_{n'}$ [8]. Upon defining the $N \times N$ positive definite matrix with entries $(\mathbf{K})_{n,n'} := \kappa(v_n, v_{n'})$, and $\mathbf{h} := [h(v_1), h(v_2), \ldots, h(v_N)]$, we can write

$$\mathbf{h} = \mathbf{K}\boldsymbol{\alpha} \tag{4}$$

where $\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$. The RKHS norm of a function h is given by $\|h\|_{\mathcal{H}}^2 := \sum_{n=1}^N \sum_{n'=1}^N \alpha_n \alpha_{n'} \kappa(v_n, v_{n'})$ or in vector form by

$$\|h\|_{\mathcal{H}}^2 = \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} \tag{5}$$

and is usually employed as a regularization term to control overfitting [8, 9, 12].

Laplacian kernels have been widely used [8, 9, 13, 14] to promote *smoothness* with respect to the underlying graph topology, by penalizing functions that exhibit pronounced variations among neighboring vertices. For a given Laplacian matrix with eigendecomposition $\mathbf{L} = \mathbf{U}\,\text{diag}\{\boldsymbol{\lambda}\}\mathbf{U}^T$, a family of graph kernels is defined as [8]

$$\mathbf{K} := r^{-1}(\mathbf{L}) := \mathbf{U}\,\text{diag}\{r^{-1}(\boldsymbol{\lambda})\}\mathbf{U}^T \tag{6}$$

where $r : \mathbb{R} \to \mathbb{R}_+$ is chosen to be a monotonically increasing function. Table 1 summarizes common choices of $r(\cdot)$ which can be selected to promote a certain structure in the so-called graph Fourier transform of $\mathbf{h}$ [1, 8, 9].

3.2. Kernel-based semi-parametric reconstruction

Since $f_{\text{NP}} \in \mathcal{H}$, vector $\mathbf{f}_{\text{NP}}$ can be represented as in (4). By defining $\boldsymbol{\beta} := [\beta_1, \ldots, \beta_M]^T$, and the $N \times M$ matrix $\mathbf{B}$ with entries $(\mathbf{B})_{n,m} := b_m(v_n)$, the parametric term can be written in vector form as $\mathbf{f}_{\text{P}} := \mathbf{B}\boldsymbol{\beta}$. The semi-parametric estimates can be found as the solution of the following optimization problem

$$\begin{aligned} \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\} = \arg\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \ & \frac{1}{S} \sum_{s=1}^S \mathcal{L}(y_s, f(v_{n_s})) + \mu \|f_{\text{NP}}\|_{\mathcal{H}}^2 \\ \text{s.t.} \quad & \mathbf{f} = \mathbf{f}_{\text{P}} + \mathbf{f}_{\text{NP}} \\ & \mathbf{f}_{\text{P}} = \mathbf{B}\boldsymbol{\beta} \\ & \mathbf{f}_{\text{NP}} = \mathbf{K}\boldsymbol{\alpha} \end{aligned} \tag{7}$$

where the fitting loss $\mathcal{L}$ quantifies the deviation of f from the data, and $\mu > 0$ is the regularization scalar that controls overfitting the nonparametric term. Using (7), we can express our semi-parametric estimate as $\hat{\mathbf{f}} = \mathbf{B}\hat{\boldsymbol{\beta}} + \mathbf{K}\hat{\boldsymbol{\alpha}}$.

Solving (7) entails minimization over $N + M$ variables. Clearly, when dealing with large-scale graphs this could lead to prohibitively large computational cost. To ensure applicability in big-data scenaria we leverage the dimensionality reduction effected through the semi-parametric version of the representer theorem [12, 15], which establishes that

$$\hat{\mathbf{f}} = \mathbf{B}\hat{\boldsymbol{\beta}} + \mathbf{K}\mathbf{S}^T\hat{\bar{\boldsymbol{\alpha}}} \tag{8}$$

where $\hat{\bar{\boldsymbol{\alpha}}} := [\hat{\bar{\alpha}}_1, \ldots, \hat{\bar{\alpha}}_S]^T$. Estimates $\hat{\bar{\boldsymbol{\alpha}}}$, $\hat{\boldsymbol{\beta}}$ are found solving the optimization problem

$$\{\hat{\bar{\boldsymbol{\alpha}}}, \hat{\boldsymbol{\beta}}\} = \arg\min_{\bar{\boldsymbol{\alpha}}, \boldsymbol{\beta}} \frac{1}{S}\sum_{s=1}^{S} \mathcal{L}(y_s, f(v_{n_s})) + \mu\|f_{\text{NP}}\|^2_{\mathcal{H}} \tag{9}$$
$$\text{s.t.} \quad \mathbf{f} = \mathbf{f}_{\text{P}} + \mathbf{f}_{\text{NP}}$$
$$\mathbf{f}_{\text{P}} = \mathbf{B}\boldsymbol{\beta}$$
$$\mathbf{f}_{\text{NP}} = \mathbf{K}\mathbf{S}^T\bar{\boldsymbol{\alpha}}$$

where $\bar{\boldsymbol{\alpha}} := [\bar{\alpha}_1, \ldots, \bar{\alpha}_S]^T$. The RKHS norm in (9) is expressed as $\|f_{\text{NP}}\|^2_{\mathcal{H}} = \bar{\boldsymbol{\alpha}}^T\bar{\mathbf{K}}\bar{\boldsymbol{\alpha}}$, with $\bar{\mathbf{K}} := \mathbf{S}\mathbf{K}\mathbf{S}^T$. Relative to (7) the number of optimization variables in (9) is reduced to the more affordable $S + M$, with $S \ll N$.

We will consider two loss functions with complementary benefits: the *square* loss and the ϵ-*insensitive* loss. The square loss function is

$$\mathcal{L}(y_s, f(v_{n_s})) := \|y_s - f(v_{n_s}))\|_2^2 \tag{10}$$

and (9) then admits the following closed-form solution

$$\hat{\bar{\boldsymbol{\alpha}}} = (\mathbf{P}\bar{\mathbf{K}} + \mu\mathbf{I}_S)^{-1}\mathbf{P}\mathbf{y} \tag{11a}$$
$$\hat{\boldsymbol{\beta}} = (\bar{\mathbf{B}}^T\bar{\mathbf{B}})^{-1}\bar{\mathbf{B}}^T(\mathbf{y} - \bar{\mathbf{K}}\hat{\bar{\boldsymbol{\alpha}}}) \tag{11b}$$

where $\bar{\mathbf{B}} := \mathbf{S}\mathbf{B}$ and $\mathbf{P} := \mathbf{I}_S - \bar{\mathbf{B}}(\bar{\mathbf{B}}^T\bar{\mathbf{B}})^{-1}\bar{\mathbf{B}}^T$. The complexity of (11) is $\mathcal{O}(S^3 + M^3)$.

The ϵ-*insensitive* loss function is given by

$$\mathcal{L}(y_s, f(v_{n_s})) = \max(0, |y_s - f(v_{n_s})| - \epsilon) \tag{12}$$

where ϵ is tuned, e.g. via cross-validation, to minimize the generalization error and has well-documented merits in signal estimation from quantized data [16]. Substituting (12) into (9) yields a convex non-smooth quadratic problem that can be solved efficiently for $\bar{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ using e.g. interior-point methods [15].

4. NUMERICAL TESTS

This section describes tests on synthetic and real graph functions to illustrate the effective reconstruction performance of our semi-parametric graph kernel estimators, SP-GK and SP-GK(ϵ) resulting from using (10) and (12) in (9) respectively.

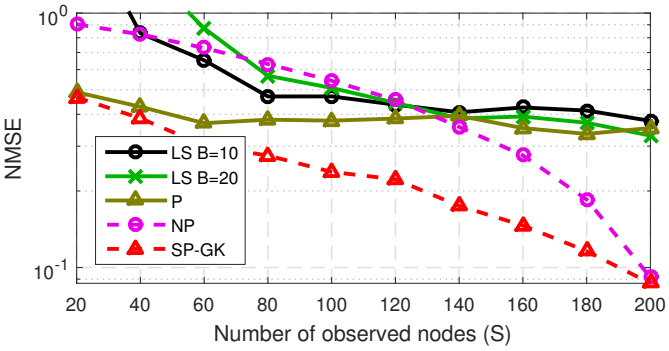

Fig. 1: NMSE of the synthetic signal estimates. ($\mu = 5 \times 10^{-4}$, $\sigma = 5 \times 10^{-4}$, $\text{SNR}_e = 5$dB).

Our approach is compared against the *parametric* (P) that considers only the parametric term in (2); the *nonparametric* (NP) [8, 14] that considers only the nonparametric term in (2); and the *least-squares* estimators (LS) from [4, 17], which assume a bandlimited model with bandwidth B. For all the experiments we use the diffusion kernel (cf. Table 1) with parameter σ.

We assess the performance of the proposed estimators via Monte Carlo simulation by comparing the normalized mean-square error (NMSE)

$$\text{NMSE} = \mathbb{E}\left[\frac{\|\hat{\mathbf{f}} - \mathbf{f}\|^2}{\|\mathbf{f}\|^2}\right] \tag{13}$$

averaged over choices of sample indices $\{n_s\}_{s=1}^S$ and, for synthetic data experiments, also over noise and signal realizations.

4.1. Synthetic signals

An Erdős-Rènyi graph with probability of edge presence 0.6 and $N = 200$ nodes was generated, and $\mathbf{f}$ was formed by superimposing a bandlimited [4, 17] with a piecewise constant signal [18]; that is

$$\mathbf{f} = \sum_{i=1}^{10} \gamma_i \mathbf{u}_i + \sum_{i=1}^{6} \delta_i \mathbf{1}_{\mathcal{V}_c} \tag{14}$$

where $\{\gamma_i\}_{i=1}^{10}$ and $\{\delta_i\}_{i=1}^{6}$ are standardized Gaussian distributed, $\{\mathbf{u}_i\}_{i=1}^{10}$ are the eigenvectors associated with the 10 smallest eigenvalues of the Laplacian matrix, $\{\mathcal{V}_i\}_{i=1}^{6}$ are the vertex sets of 6 clusters obtained via spectral clustering [19], and $\mathbf{1}_{\mathcal{V}_i}$ is the indicator vector with entries $(\mathbf{1}_{\mathcal{V}_i})_n := 1$ if $v_n \in \mathcal{V}_i$, and 0 otherwise. The parametric basis $\mathcal{B} = \{\mathbf{1}_{\mathcal{V}_i}\}_{i=1}^{6}$ was used by the estimators capturing the prior knowledge, and S vertices were sampled uniformly at random.

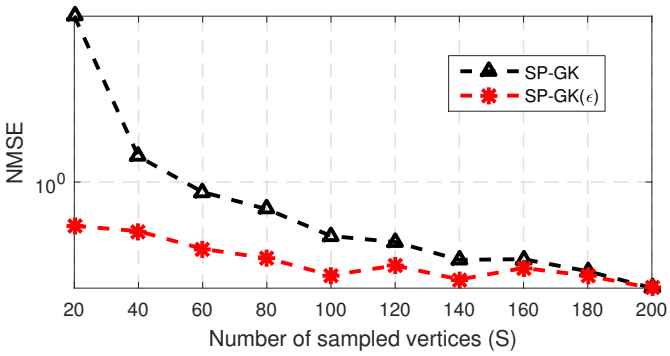

Fig. 2: NMSE of the synthetic signal estimates. ($\mu = 5 \times 10^{-4}$, $\sigma = 5 \times 10^{-4}$, $\epsilon = 10^{-4}$, and $\text{SNR}_o = -5$dB).

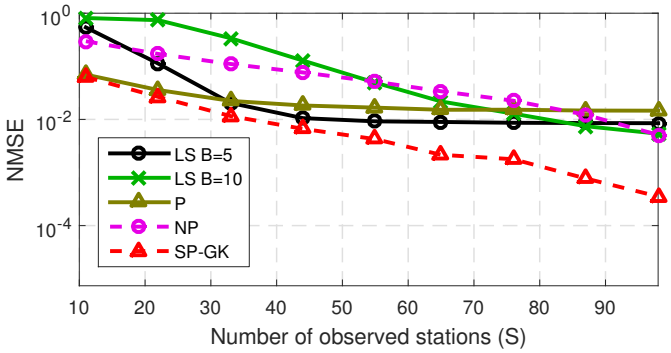

Fig. 3: NMSE of the mean temperature estimates over 2010. ($\mu = 5 \times 10^{-5}$, $\sigma = 1.3$, and $C = 4$).

In the first experiment, white Gaussian noise e_s of variance σ_e^2 is added to each sample f_s to yield signal-to-noise ratio $\text{SNR}_e := \|\mathbf{f}\|_2^2/(N\sigma_e^2)$. Fig. 1 reports the NMSE of all competing methods and showcases the benefits of our semi-parametric estimator. Observe that the limited flexibility of the parametric approaches, LS and P, affects their ability to capture the true signal structure. The nonparametric approach (NP) is performing better, but only when the amount of available samples increases. Both our semi-parametric estimators were found to outperform all competing approaches, exhibiting reliable reconstruction even with few samples.

Note here that since the performance of SP-GK(ϵ) and SP-GK was very close, we have chosen to include only SP-GK in Fig. 1, to avoid "clotting" the plot. To illustrate the differences of our semi-parametric estimators, we conduct a second experiment which compares the performance of SP-GK and SP-GK(ϵ) in the presence of outlying noise. Each sample f_s is contaminated with Gaussian noise o_s of large variance σ_o^2 with probability $p = 0.1$. Fig. 2 demonstrates the robustness of SP-GK(ϵ) which is attributed to the $\epsilon-$insensitive loss function (12).

4.2. Real signals

The second dataset is provided by the National Climatic Data Center [20], and comprises temperature measurements at $N = 109$ stations across the continental United States in 2010. The geographical coordinates of the measuring stations have been used to construct a graph

$$(\mathbf{A})_{n,n'} = \frac{\exp\{-d_{n,n'}^2\}}{\sqrt{\sum_{j\in\mathcal{N}_n^k}\exp\{-d_{n,j}^2\}\sum_{l\in\mathcal{N}_{n'}^k}\exp\{-d_{n',l}^2\}}} \quad (15)$$

where $\mathcal{N}_n^k$ is the set containing the $k = 7$ nearest neighbors of station n, and $d_{n,n'}$ is the geographical distance between stations n and n'. The neighborhoods are defined based on $d_{n,n'}$, which is justified since geographically close stations tend to measure similar temperature values. To illustrate the benefits of leveraging side information, we cluster the stations into C vertex sets $\{\mathcal{V}_c\}_{c=1}^C$ according to their altitude, and we construct $\mathcal{B}$ using the indicator vectors $\{\mathbf{1}_{\mathcal{V}_c}\}_{c=1}^C$. We sample the temperatures at S stations, chosen uniformly at random, and we reconstruct the signal across all N nodes.

Fig. 3 reports the performance of the different graph inference methods, and illustrates the advantage of the proposed approach. SP-GK[1] leverages the altitude information and achieves NMSE$\leq 10^{-2}$ with as few as $S = 35$ samples, whereas NP requires at least $S = 85$ for the same NMSE. Moreover, we observe that the performance of the pure parametric method LS – which assumes a bandlimited model – does not improve after a certain number of samples. This was expected since the actual signal does not adhere to the modeling assumptions.

5. CONCLUSION

In this work, we introduced a novel semi-parametric method for reconstruction of functions over graphs. We decompose the graph function into a nonparametric component that promotes smoothness via graph kernels, and a parametric component that represents prior information via a known basis. We derive two estimators corresponding to different loss functions, the square loss that admits closed form solution, and the ϵ-insensitive loss that promotes robustness, and by employing the representer theorem we achieve affordable complexity. Numerical tests on synthetic and real data-sets demonstrate the competitive performance of SP-GK against existing alternatives.

Interesting future directions include a multi-kernel extension, a data-driven approach for selecting a suitable kernel function, as well as the application of the semi-parametric approach to real-world scenarios.

[1] SP-GK(ϵ) performed similarly to SP-GK and was not included in the plot.

6. REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig. Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*, Springer New York, 2009.

[3] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Sig. Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[4] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Sig. Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.

[5] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Global Conf. Sig. Inf. Process.*, Austin, Texas, 2013, IEEE, pp. 491–494.

[6] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Sig. Process.*, vol. 64, no. 16, pp. 4363–4378, Aug. 2016.

[7] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Sig. Process.*, vol. 62, no. 15, pp. 3849–3862, Aug. 2014.

[8] A. J. Smola and R. I. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*, pp. 144–158. Springer, 2003.

[9] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Sig. Process.*, vol. 65, no. 3, pp. 764–778, 2017.

[10] A. N. Nikolakopoulos and J. D. Garofalakis, "Top-n recommendations in the presence of sparsity: An ncd-based approach," in *Web Intelligence*. IOS Press, 2015, vol. 13, pp. 247–265.

[11] A. N. Nikolakopoulos, M. A. Kouneli, and J. D. Garofalakis, "Hierarchical itemspace rank: Exploiting hierarchy to alleviate sparsity in ranking-based recommendation," *Neurocomputing*, vol. 163, pp. 126–136, 2015.

[12] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Computational Learning Theory*. Springer, 2001, pp. 416–426.

[13] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE J. Sel. Topics Sig. Process.*, vol. 11, no. 6, pp. 1–14, Sep. 2017.

[14] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. Int. Conf. Mach. Learn.*, Sydney, Australia, Jul. 2002, pp. 315–322.

[15] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2002.

[16] V. Vapnik, *The nature of statistical learning theory*, Springer, 2013.

[17] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Sig. Process.*, vol. 64, no. 18, pp. 4845–4860, Sep. 2016.

[18] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal representations on graphs: Tools and applications," arXiv preprint arXiv:1512.05406 [Online]. Available: `http://arxiv.org/abs/1512.05406`, 2015.

[19] U. Von Luxburg, "A tutorial on spectral clustering," vol. 17, no. 4, pp. 395–416, Dec. 2007.

[20] "1981-2010 U.S. climate normals," [Online]. Available: `https://www.ncdc.noaa.gov`.