



HAL
open science

MODELING SIGNALS OVER DIRECTED GRAPHS THROUGH FILTERING

Harry Sevi, Gabriel Rilling, Pierre Borgnat

► **To cite this version:**

Harry Sevi, Gabriel Rilling, Pierre Borgnat. MODELING SIGNALS OVER DIRECTED GRAPHS THROUGH FILTERING. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov 2018, Anaheim, United States. pp.718-722, 10.1109/GlobalSIP.2018.8646534. hal-02399335

HAL Id: hal-02399335

<https://hal.science/hal-02399335v1>

Submitted on 9 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MODELING SIGNALS OVER DIRECTED GRAPHS THROUGH FILTERING

Harry Sevi^{1,2}, Gabriel Rilling¹, Pierre Borgnat²

¹ CEA, LIST, Laboratoire d'Analyse de Données et Intelligence des Systèmes, 91400 Gif-sur-Yvette, France

² Univ Lyon, Ens de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France

ABSTRACT

In this paper, we discuss the problem of modeling a graph signal on a directed graph when observing only partially the graph signal. The graph signal is recovered using a learned graph filter. The novelty is to use the random walk operator associated to an ergodic random walk on the graph, so as to define and learn a graph filter, expressed as a polynomial of this operator. Through the study of different cases, we show the efficiency of the signal modeling using the random walk operator compared to existing methods using the adjacency matrix or ignoring the directions in the graph.

Index Terms— Graph signal processing, random walks, filtering.

1. INTRODUCTION

The interest in the analysis and exploitation of data represented by graphs is becoming increasingly significant, partly due to the affluence of such data in various fields such as neurosciences, biology or social networks. To better process, analyze and exploit these graph data, developing relevant and efficient mathematical and computational methods is helpful.

In recent years, a new mathematical framework has been developed called graph signal processing [1, 2, 3]. The aim of this framework is to generalize the fundamental concepts of classical signal processing, such as filtering, Fourier transform or wavelets, and it considers graph data as a signal on graph.

A main challenge of graph signal processing is the determination of a suitable Fourier basis for signals on a directed graph. A few approaches have emerged, either based on Jordan's decomposition of the adjacency matrix [4] or recently constructed through non-convex optimization problems [5, 6]. Recently, we have proposed an appropriate frequency analysis based on the eigenvectors of the random walk operator, showing that they constitute a new Fourier-type basis for signals defined on directed graphs [7].

We focus on a problem of signal modeling, where we learn a model that allows the reconstruction of a signal's miss-

ing values given a small number of known values. This problem is related to the semi-supervised learning framework [8], except we assume the whole signal known to learn the model. Hence, our model would be better suited for instance for a lossy compression application. Here, the focus is rather on the comparison of different models based on different graph signal processing frameworks. The problem can be studied using the graph Laplacian [9] whose main limitation is to apply only to undirected graphs with real and non-negative edges, or a graph signal processing framework based on the adjacency matrix [2].

We develop signal modeling on directed graphs, leveraging directed graph signal processing using random walks [7]. We show how to learn the parameters of a polynomial filter based on the random walk operator to model a signal, then how it allows us to recover it when it is only partially known, on a small number of vertices. We also consider other reference operators combining the random walk and its reversed version. We study numerically, on an example based on the political blogs dataset [10], the efficiency and advantage of the proposed approaches, and compare them to existing works.

2. RANDOM WALKS ON DIRECTED GRAPHS

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a weighted directed graph with cardinality $|\mathcal{V}| = N$. The graph \mathcal{G} is represented by its weighted adjacency matrix $\mathbf{A} = \{a_{ij}\}_{1 \leq i, j \leq N} \in \mathbb{R}_+^{N \times N}$. We assume that \mathcal{G} is strongly connected.

A random walk [11] on \mathcal{G} is a (possibly non-reversible) Markov chain $\mathcal{X} = (X_n)_{n \geq 0}$ on the finite state space \mathcal{V} . From a graph theory point of view, the random walk operator $\mathbf{P} \in \mathbb{R}_+^{N \times N}$ is defined as $\mathbf{P} = \mathbf{D}^{-1} \mathbf{A}$, where $\mathbf{D} = \text{diag}\{d_1, \dots, d_N\}$, the diagonal matrix of the out-degrees of the vertices and \mathbf{A} the adjacency matrix, with $d_i = \sum_{j=1}^N a_{ij}$, for all $i = 1, \dots, N$.

We assume throughout this paper that the random walk \mathcal{X} is ergodic, that is irreducible and aperiodic [11]. Consequently, the random walk \mathcal{X} admits a unique stationary distribution π [11]. We define the time reversed ergodic random walk \mathcal{X}^* with stationary distribution π [12] characterized by the following operator

This work was supported by the ANR-14-CE27-0001 GRAPHSIP grant and the ACADEMICS Grant in the Scientific Breakthrough Program of the IDEXLYON.

$$\mathbf{P}^* = \mathbf{\Pi}^{-1} \mathbf{P}^\top \mathbf{\Pi},$$

where $\mathbf{\Pi} = \text{diag}\{\pi_1, \dots, \pi_N\}$ is the diagonal matrix of the stationary distribution.

From the non-reversible \mathcal{X} , with transition matrix \mathbf{P} and unique stationary measure π , it is known how to construct the additive reversibilization [13] of \mathcal{X} , whose transition matrix is the average between \mathbf{P} and its time reversed \mathbf{P}^* . It is expressed as

$$\bar{\mathbf{P}} = \frac{\mathbf{P} + \mathbf{P}^*}{2}. \quad (1)$$

$\bar{\mathbf{P}}$ is reversible with unique stationary distribution π .

A novelty here is to define and consider the more general class $\bar{\mathcal{P}}$ of convex combinations between \mathbf{P} and \mathbf{P}^* :

$$\bar{\mathcal{P}} = \left\{ \bar{\mathbf{P}}_\alpha : \bar{\mathbf{P}}_\alpha = (1 - \alpha)\mathbf{P} + \alpha\mathbf{P}^* \mid \alpha \in [0, 1] \right\}. \quad (2)$$

The elements $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$ are non-reversible except for $\alpha = 1/2$, namely $\bar{\mathbf{P}}_{1/2} = \bar{\mathbf{P}}$. Still, they share the same stationary distribution π , and we will show that it can be useful to consider them as reference to defined filters on directed graphs.

3. PARAMETRIC MODELING OF GRAPHS SIGNALS

In this section, we consider a way of modeling the relationships between values of a graph signal using a graph filter. The model is expressed as a graph filter that takes a few values of the graph signal and reconstructs the other values. A possible application of such a model could be the lossy compression of the signal, where knowing only the graph, the coefficients of the graph filter and a few signal values enable the reconstruction of the whole signal.

3.1. Problem formulation

A graph filter \mathbf{H} is expressed as a polynomial finite sum of a reference operator \mathbf{R} , that is

$$\mathbf{H} = \sum_{k=0}^K \theta_k \mathbf{R}^k, \quad \theta_k \in \mathbb{R}, \quad \forall k = 0, \dots, K. \quad (3)$$

The coefficients of the graph filter are learned by solving the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} = \{\theta_k\}_{k=0}^K \in \mathbb{R}^{K+1}}{\text{argmin}} \left[\mathbb{E} \left[\left\| \mathbf{f}_0 - \sum_{k=0}^K \theta_k \mathbf{R}^k \mathbf{Y} \right\|_\mu^2 \right] \right], \quad (4)$$

where \mathbf{f}_0 is the original graph signal, $\mathbf{Y} = [Y_1, \dots, Y_N]^\top \in \ell^2(\mathcal{V}, \mu)$ is a multivariate random variable indexed on the vertex set \mathcal{V} . The random variables $Y_k, k = 1, \dots, N$ can be

expressed as $Y_k = \varepsilon_k f_{0k}$, where the ε_k are assumed independent and Bernoulli distributed with parameters δ_k : $\varepsilon_k \sim \text{Ber}(\delta_k)$. The random variable $\bar{\varepsilon}$ defined as

$$\bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^N \varepsilon_i, \quad (5)$$

is the proportion of known values of \mathbf{f}_0 in \mathbf{Y} . \mathbf{Y} can be viewed as a random sampling of the graph signal \mathbf{f}_0 and $\mathbf{H}\mathbf{Y}$ as a reconstruction of \mathbf{f}_0 from these samples. The optimization problem (4) corresponds to the minimization of the reconstruction error given such a set of random samples.

3.2. Solution of the problem

The solution of (4) can be expressed as

$$\hat{\boldsymbol{\theta}} = \mathbf{Z}^{-1} \mathbf{M} \mathbf{Q}^\top \mathbf{f}_0,$$

with $\mathbf{M} = \text{diag}\{\mu_1, \dots, \mu_N\}$, \mathbf{Z} the matrix where each entry $Z_{k\ell}$ is expressed as

$$Z_{k\ell} = \text{Tr} \left(([\mathbf{R}^k]^\top \mathbf{M} \mathbf{R}^\ell \mathbb{E}(\mathbf{Y} \mathbf{Y}^\top)) \right), \quad \forall \{k, \ell\} \in \llbracket 0, K \rrbracket^2,$$

and $\mathbf{Q} = [\mathbf{q}_0, \dots, \mathbf{q}_K] \in \mathbb{R}^{N \times (K+1)}$ where each vector \mathbf{q}_j is

$$\mathbf{q}_j = \mathbf{R}^j \mathbb{E}(\mathbf{Y}), \quad \forall j \in \llbracket 0, K \rrbracket.$$

The resulting estimation of the filter is

$$\hat{\mathbf{H}} = \sum_{k=0}^K \hat{\theta}_k \mathbf{R}^k, \quad \hat{\theta}_k \in \mathbb{R}, \quad \forall k = 0, \dots, K. \quad (6)$$

4. AN EXPLORATION OF MODELING ACCURACY WITH DIFFERENT GRAPH OPERATORS

The reconstruction quality using the previous approach depends on several factors: the choice of reference operator \mathbf{R} , the order of the filter K and the random sampling strategy defined by the Bernoulli parameters δ_k . Here, we evaluate primarily the influence of \mathbf{R} and compare two possible sampling strategies.

As an example, we consider the dataset of the political blogs of the 2004 US presidential campaign [10]. The dataset consists of 1224 political blogs where each political blog has a political orientation, either republican or democrat. This dataset is modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where each vertex $v \in \mathcal{V}$ corresponds to a blog and an edge between two vertices $\{v_i, v_j\}$ indicates the presence of hyperlinks from the blog associated to the vertex v_i to the vertex v_j . This graph is unweighted. The political orientations of blogs are modeled by a graph signal $\mathbf{f}_0 = \{f_0(v_1), \dots, f_0(v_N)\}$ where each vertex v_i has a label $f_0(v_i) \in \{-1, 1\}$ characterizing its political orientation.

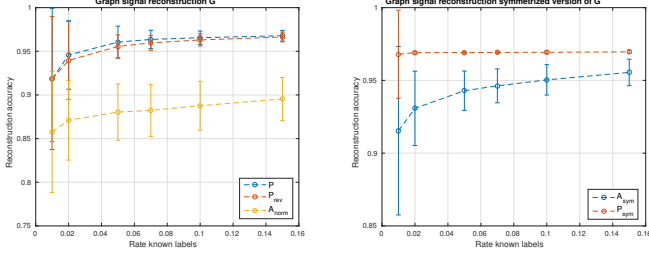


Fig. 1. (Case 1) Left: Reconstruction of the graph signal f'_0 on the subgraph \mathcal{G}' . Right: Reconstruction of the graph signal f'_0 on the symmetrized version of \mathcal{G}' .

The directed graph \mathcal{G} associated to the political blogs is not strongly connected. In our theoretical framework, we only consider strongly connected graphs. Hence, we will first consider the largest strongly connected subgraph of \mathcal{G} , denoted by $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ which is made up of $|\mathcal{V}'| = N' = 793$ vertices (roughly 65% of the vertex set \mathcal{V}) and its associated graph signal f'_0 . In 4.3, we will extend the method to \mathcal{G} .

As the chosen signal takes values only in $\{-1, 1\}$, we will compare the reconstruction accuracy using a proportion of correctly reconstructed labels obtained via

$$\hat{f} = \text{sign}(\hat{\mathbf{H}}\mathbf{Y}).$$

4.1. Signal modeling on \mathcal{G}'

In this section, we solve the problem (4) by learning a polynomial graph filter with $K = 10$ on \mathcal{G}' . We consider various possible reference operators: those for directed graphs, in the set $\mathcal{C} = \{\mathbf{A}_{norm}, \mathbf{P}, \bar{\mathbf{P}}\}$ where \mathbf{A}_{norm} is a normalized version of the the adjacency matrix whose spectral norm is equal to one, \mathbf{P} the random walk operator and $\bar{\mathbf{P}}$ its additive reversibilization; and those for undirected (or symmetrized) graphs, in the set $\bar{\mathcal{C}} = \{\mathbf{A}_{sym}, \mathbf{P}_{sym}\}$ where \mathbf{A}_{sym} is the normalized version of the symmetrized adjacency matrix $(\mathbf{A} + \mathbf{A}^\top)/2$ and \mathbf{P}_{sym} the random walk operator based on \mathbf{A}_{sym} . The random variables $Y_j, j = 1, \dots, N'$ are distributed in one of the two following cases:

Case 1: Random variables $Y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(p)$ with p the proportion of known labels.

Case 2: Random variables $Y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(\alpha\pi_j)$, such that $\sum_j \mathbb{E}(\varepsilon_j) = pN'$ with p the proportion of known labels.

The proportions of correctly reconstructed labels are measured for various p values using 500 realizations of \mathbf{Y} .

Numerical simulations: Case 1

We evaluate the reconstruction performance of the graph signal f'_0 in figure 1. For all proportions p of known labels,

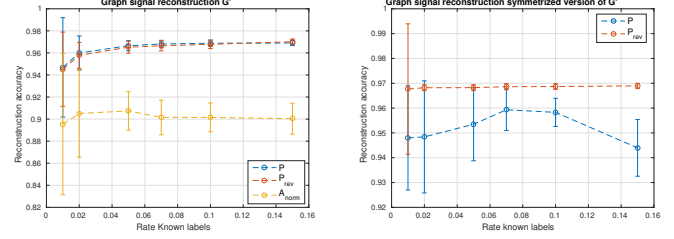


Fig. 2. (Case 2) Left: Reconstruction of the graph signal f'_0 on the subgraph \mathcal{G}' . Right: Reconstruction of the graph signal f'_0 on the symmetrized version of \mathcal{G}' .

the average rate of correctly reconstructed labels in the graph signal f'_0 from a learned polynomial graph filter of \mathbf{P} or $\bar{\mathbf{P}}$ is significantly better than the one based on a learned polynomial graph filter of \mathbf{A}_{norm} in the case where we consider either the subgraph \mathcal{G}' or its symmetrized version. Furthermore, the reconstruction performance using a filter based on \mathbf{P} is slightly better than using a filter based on $\bar{\mathbf{P}}$.

Numerical simulations: Case 2

The reconstruction performance is shown in figure 2. We observe a significantly better performance using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$ than based on \mathbf{A}_{norm} in the case where we consider either the subgraph \mathcal{G}' or its symmetrized version and a slightly better performance using \mathbf{P} rather than $\bar{\mathbf{P}}$. We notice the reconstruction performance based on \mathbf{A}_{norm} has a large variability and even decreases for larger p .

Case 1 vs. Case 2

The reconstruction performance using a filter based on \mathbf{P} or $\bar{\mathbf{P}}$ is slightly better when the samples in \mathbf{Y} are selected according to a distribution proportional to the stationary distribution π (case 2) than when the distribution is uniform (case 1). This is expected as vertices with larger π_k values correspond to better-connected blogs, which are likely to have an influence on a larger number of other blogs. Although the reconstruction performance using a filter based on \mathbf{A}_{norm} is better in case 2 than in case 1, the large variability and the poorer performance at large p in case 2 suggest that these filters provide generally poorer models of the signal.

4.2. Signal modeling on \mathcal{G}' with $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$

Here we consider the same problem as in the previous section but now compare the performance using reference operators $\bar{\mathbf{P}}_\alpha \in \bar{\mathcal{P}}$. The random variables $Y_j, j = 1 \dots, N'$ are distributed as $Y_j = \varepsilon_j f'_0(v_j)$ where $\varepsilon_j \sim \text{Ber}(\alpha\pi_j)$, such that $\sum_j \mathbb{E}(\varepsilon_j) = pN'$.

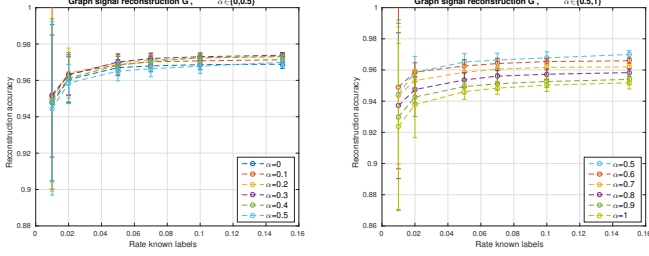


Fig. 3. Reconstruction of the graph signal f'_0 on the subgraph \mathcal{G}' .

Numerical simulations

We evaluate the reconstruction performance of f_0 as a function of α in figure 3. We notice that the reconstruction performances using a filter based on $\bar{\mathbf{P}}_\alpha$, $\alpha \simeq 0.3$ are the best except for the rates $p = \{0.01, 0.02\}$. We notice poorer performance overall when α increases from $\alpha = 0.5$ to $\alpha = 1$. This suggests that, for this signal, graph filters based on $\bar{\mathbf{P}}_\alpha$ perform better when the convex combination of \mathbf{P} and \mathbf{P}^* involves more the forward random walk, i.e. $\alpha \in (0, 0.5)$, than the backward one, i.e. $\alpha \in (0.5, 1)$.

4.3. Signal modeling on \mathcal{G}

We now consider the same problem on the whole graph \mathcal{G} . The random variables $Y_j, j = 1 \dots, N$ are distributed as $Y_j = \varepsilon_j f_0(v_j), \varepsilon_j \sim \text{Ber}(p)$.

As some vertices in \mathcal{G} have an out-degree equal to zero, we cannot create a transition matrix directly from the adjacency matrix. To overcome this problem, we propose two approaches that make the graph strongly connected and the associated random walk ergodic:

1. Rank-one perturbation: from the original adjacency matrix \mathbf{A} , we construct a new adjacency matrix \mathbf{A}_ϵ as follows

$$\mathbf{A}_\epsilon = \mathbf{A} + \epsilon \mathbf{J},$$

where $\mathbf{J} = \mathbf{1}\mathbf{1}^\top/N$ is a rank-one matrix and ϵ is small. The weak perturbation of the adjacency matrix \mathbf{A} by the matrix \mathbf{J} ensures that the random walk on \mathcal{G} is ergodic with stationary measure π_ϵ and its associated transition matrix \mathbf{P}_ϵ is well-defined. For our experiments, we choose $\epsilon = 10^{-4}$.

2. Construction of the Google matrix of \mathcal{G} [14]. The construction is in two steps. Firstly, we construct an adjacency matrix $\tilde{\mathbf{A}}$ from \mathbf{A} by adding one for all dangling nodes, that is nodes with no out-edges. From $\tilde{\mathbf{A}}$ we can construct the transition matrix \mathbf{S} . Secondly, we define the Google matrix \mathbf{P}_G as

$$\mathbf{P}_G = (1 - \gamma)\mathbf{S} + \gamma\mathbf{J}.$$

with $\gamma = 0.85$ [14] and stationary measure π_G .

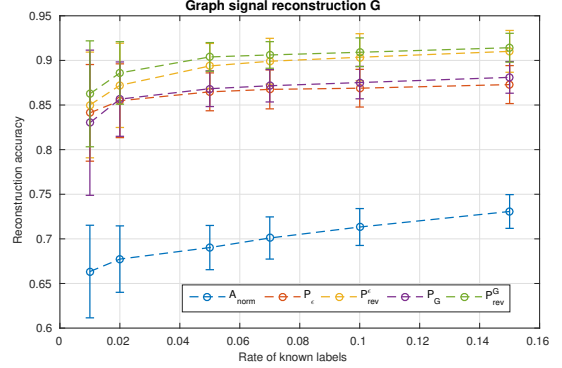


Fig. 4. Reconstruction of the graph signal f_0 on \mathcal{G} .

For each proportion of known labels p , the good reconstruction rates are derived by averaging 500 graph signals realizations of \mathbf{Y} . We consider the set of reference operators $\mathcal{F} = \{\mathbf{A}_{norm}, \mathbf{P}_\epsilon, \mathbf{P}_G, \bar{\mathbf{P}}_\epsilon, \bar{\mathbf{P}}_G\}$.

Numerical simulations

The reconstruction performance is shown in the figure 4. For all p values, the reconstruction performance using filters based on \mathbf{A}_{norm} is significantly worse than with all other reference operators. Compared to the other reference operators, \mathbf{A}_{norm} is the only one that can be used without modifying the graph to make it strongly connected. Still, its performance is always worse.

Among the other reference operators, we notice a clearly better performance generally when using the reversibilizations $\bar{\mathbf{P}}_\epsilon$ and $\bar{\mathbf{P}}_G$ compared to the non-reversible random walks \mathbf{P}_ϵ and \mathbf{P}_G . This differs from the results on \mathcal{G}' where both would perform similarly.

Finally, the reconstruction performance does not depend as much on the approach we use to make the graph strongly connected. It seems though that the Google approach slightly outperforms the rank-one approach, both for the non-reversible random walks and their reversibilizations.

5. CONCLUSION

We studied parametric modeling on directed graphs using graph filters. Our approach based on the random walk operator is suitable on directed as well as undirected graphs. We formulated the problem and derived a closed-form solution. Experiments on a dataset of political blogs showed that the signal modeling based on the random walks operators outperforms the one based on the adjacency matrix. In addition, the parameter α in the convex combination between the random walk operator and its adjoint is seen as a tunable parameter that can improve the modeling. The set $\bar{\mathcal{P}}$ might be an alluring set of reference operators for convolutional neural networks [15, 16] on directed graphs.

6. REFERENCES

- [1] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
- [2] Aliaksei Sandryhaila and José MF Moura, “Discrete signal processing on graphs,” *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [3] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, “Graph signal processing,” *arXiv preprint arXiv:1712.00468*, 2017.
- [4] Aliaksei Sandryhaila and Jose MF Moura, “Discrete signal processing on graphs: Frequency analysis.,” *IEEE Trans. Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.
- [5] Rasoul Shafipour, Ali Khodabakhsh, Gonzalo Mateos, and Evdokia Nikolova, “A digraph fourier transform with spread frequency components,” *arXiv preprint arXiv:1705.10821*, 2017.
- [6] Stefania Sardellitti, Sergio Barbarossa, and Paolo Di Lorenzo, “On the graph fourier transform for directed graphs,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 796–811, 2017.
- [7] Harry Sevi, Gabriel Rilling, and Pierre Borgnat, “Multiresolution analysis of functions on directed networks,” in *Wavelets and Sparsity XVII*. International Society for Optics and Photonics, 2017, vol. 10394, p. 103941Q.
- [8] Xiaojin Zhu, “Semi-supervised learning literature survey,” 2005.
- [9] Xiaojin Zhu, John Lafferty, and Ronald Rosenfeld, *Semi-supervised learning with graphs*, Ph.D. thesis, Carnegie Mellon University, language technologies institute, school of computer science, 2005.
- [10] Lada A Adamic and Natalie Glance, “The political blogosphere and the 2004 us election: divided they blog,” in *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005, pp. 36–43.
- [11] Pierre Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31, Springer Science & Business Media, 2013.
- [12] Ravi Montenegro, Prasad Tetali, et al., “Mathematical aspects of mixing times in markov chains,” *Foundations and Trends® in Theoretical Computer Science*, vol. 1, no. 3, pp. 237–354, 2006.
- [13] James Allen Fill, “Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process,” *The annals of applied probability*, pp. 62–87, 1991.
- [14] Amy N Langville and Carl D Meyer, *Google’s PageRank and beyond: The science of search engine rankings*, Princeton University Press, 2011.
- [15] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst, “Geometric deep learning: going beyond euclidean data,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
- [16] James Atwood and Don Towsley, “Diffusion-convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2016, pp. 1993–2001.