GRAPH LOCAL-SMOOTH DICTIONARY LEARNING

Quentin Laborde[1]*, Antoine Mazarguil*[1]*, Laurent Oudre*[1]

[1]Université Paris Saclay, Université Paris Cité, ENS Paris Saclay, CNRS, SSA, INSERM, Centre Borelli, F-91190, Gif-sur-Yvette, France

ABSTRACT

In this work, we leverage recent graph uncertainty principles to introduce a new dictionary learning method on graphs. Our method considers two distinct classes of atoms, spatially local atoms on the graph, and smooth atoms, together with well suited penalty functions. Notably, the consideration of the notion of localized atoms on graphs allows to model local and interpretable phenomena.

Index Terms— Graph-signal, Dictionary-learning, Signal decomposition

1. INTRODUCTION

Dictionary learning is a key tool for estimating a linear representation of data, i.e., learning a linear subspace defined by a family of basis vectors — commonly referred to as *atoms* — that form a dictionary. This representation has been successfully used in a wide range of fields including, obviously, dimensionality reduction [1], denoising [2], and image classification [3]. However, if linear learning methods are mainly restricted to vector data, studying signals that live on topologically complicated domains — such as social networks [4], cyber-physical systems [5] or even protein interaction networks [6] — requires the use of graph as a flexible data representation tool, suitable for modeling the underlying structure of such signals. The study of interpretable factorization models applicable to such *structured data* thus motivated the development of graph-based dictionary learning models.

In this work, we address the issue of learning such dictionaries from graph signals by considering two distinct classes of atoms: (i) spectrally smooth atoms — characterizing continuous phenomena of the underlying structure of the graph — and (ii) spatially local atoms. Such decomposition methods are an efficient tool to isolate different phenomena whose sum is only measured, *e.g.* true signal and noise, or to detect abnormal signals on networks. As we will show in Section 4.2, taking into account local atoms is also of great interest for the analysis of complex bio-mechanical time series, since it allows to decompose signals into a combination of localized and therefore more easily interpretable basic elements.

2. STATE OF THE ART

2.1. Graph Signal Processing

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ be a symmetric, weighted and undirected graph composed of a set of N nodes $\mathcal{V} = \{v_1, v_2, ..., v_N\}$, a set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, as well as a weight matrix $\mathbf{W} \in \mathbb{R}_+^{N \times N}$. The graph is moreover supposed to be connected. Then, provided a signal $\mathbf{x} \in \mathbb{R}^N$, it is possible to define a signal on the graph $\mathcal{G}$, as a mapping $x : \mathcal{V} \longrightarrow \mathbb{R}$ from the set of vertices to the set of real numbers.

We also introduce the Laplacian matrix of the graph $\mathcal{G}$, defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, with $\mathbf{D}$ the degree matrix of $\mathcal{G}$ and $\mathbf{W}$ its weight matrix. The Laplacian matrix classically allows to quantify the *smoothness* of a graph signal, defined as

$$S_{\mathcal{G}}(\mathbf{x}) := \sum_{(i,j)\in E} w_{i,j}(x_i - x_j)^2 \;=\; \mathbf{x}^T \mathbf{L}\, \mathbf{x}$$

In addition, two notions of signal localization can be introduced: localization in the spatial domain — to what extent signal intensities are concentrated around one or more nodes called *central nodes* — and in the frequency domain — to what extent active frequencies are concentrated. Given a central node $c \in \mathcal{V}$, and a graph distance between each pair of nodes $d^{\mathcal{G}} : |\mathcal{V}| \times |\mathcal{V}| \longrightarrow \mathbb{R}_+$, the local spread [7] of a signal $\mathbf{x}$ around node c is defined as

$$\Delta^2_{loc}(\mathbf{x}, c) \;=\; \frac{1}{||\mathbf{x}||_2^2} \sum_{j=1}^{N} d^{\mathcal{G}}(c, v_j)^2 \, x_j^2 \;=\; \frac{1}{||\mathbf{x}||_2^2} \, \mathbf{x}^T \, \mathbf{P}_c \, \mathbf{x},$$

with $\mathbf{P}_c \;=\; \mathrm{diag}\big(d^{\mathcal{G}}(c, v_1), \ldots, d^{\mathcal{G}}(c, v_N)\big)$. Similarly, the spectral spread of a signal on a graph can be quantified as

$$\Delta^2_{sm}(\mathbf{x}) \;=\; \frac{1}{||\mathbf{x}||_2^2} \, \mathbf{x}^T \, \mathbf{L}\, \mathbf{x}$$

Given the notions of spatial and frequency uncertainty defined above, the graph uncertainty metric introduced by Agaskar [7] restricts the set of feasible pairs $(\Delta^2_{loc}(\mathbf{x}, c), \Delta^2_{sm}(\mathbf{x}))$, which implies that a signal cannot be arbitrarily localized simultaneously in the spatial and spectral domains. Formally, given some $\alpha > 0$, the following inequality applies to any non-zero signal $\mathbf{x} \in \mathbb{R}^N$

$$\Delta^2_{loc}(\mathbf{x}, c) \;+\; \alpha \, \Delta^2_{sm}(\mathbf{x}) \;\geq\; q(\alpha)$$

with $q(\alpha) > 0$ the smallest eigenvalue of the matrix $\mathbf{P}_c + \alpha \mathbf{L}$.

2.2. Graph Dictionary Learning

Given a graph signal $\mathbf{x}$, a dictionary-based *decomposition* of $\mathbf{x}$ is a vector $\mathbf{a} \in \mathbb{R}^M$, such that

$$\mathbf{x} \approx \sum_{m=1}^{M} a_m \, \mathbf{d}^{(m)}.$$

The dictionary $\mathbf{D}$ is thus defined as the set $(\mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(M)})$ of M vectors of $\mathbb{R}^N$ — also called *atoms* — while the vector $\mathbf{a} = (a_1, \ldots, a_M)$ is denoted as the *activation vector* of signal $\mathbf{x}$ for the dictionary $\mathbf{D}$. Given $\mathbf{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(L)})$ a set of L signals of $\mathbb{R}^{N \times L}$, learning a dictionary of size M from the signal $\mathbf{X}$ can be reduced to solving the following optimization problem:

$$\mathbf{D}^*, \, \mathbf{A}^* = \underset{\mathbf{D},\mathbf{A}}{\mathbf{argmin}} ||\mathbf{X} - \mathbf{D}\mathbf{A}||_F^2 \; + \; f(\mathbf{D}) \; + \; g(\mathbf{A}), \quad (1)$$

where $||\cdot||_F$ denotes the Frobenius norm and $f(\mathbf{D})$, $g(\mathbf{A})$ are penalty terms imposed to the learned dictionnary and activation matrix, respectively.

The optimization problem (1) is often — if not always — non-convex, and the search for a local solution is performed through an alternated minimization procedure, described as follows:

- *Activation update* (AU) step: given a dictionary $\mathbf{D}$ at step $k-1$, learns the optimal activation matrix at step k, such that $\mathbf{A}^* = \underset{\mathbf{A}}{\mathbf{argmin}} ||\mathbf{X} - \mathbf{D}\mathbf{A}||_F^2 + g(\mathbf{A})$.
- *Dictionary update* (DU) step: given an activation matrix $\mathbf{A}$ at step k, learns the optimal dictionary at step k, such that $\mathbf{D}^* = \underset{\mathbf{D}}{\mathbf{argmin}} ||\mathbf{X} - \mathbf{D}\mathbf{A}||_F^2 + f(\mathbf{D})$.

2.3. Penalties and Constraints

Several penalty terms g and f have been introduced in the literature. We now briefly review the main ones.

2.3.1. AU Step Penalties

Sparsity is a desired property for activation vectors. However, approaches that strictly impose an L_0-norm are not convex, making the problem NP-hard. A common solution [8] is to replace the L_0-norm by the L_1-norm, which tends to produce sparse solutions while simplifying the search for a solution. The associated penalty function is defined as follows

$$g^{(L_1)}(\mathbf{A}) \; = \; \alpha_p \, ||\mathbf{A}||_1,$$

with $\alpha_p > 0$ a penalty parameter.

A second approach, introduced by Ramamurthy [9], consists in integrating an *a priori* knowledge of the graph structure when updating the activation matrix. Formally, it consists in considering a second weighted graph $\mathcal{G}^x = (\mathcal{V}^x, \mathcal{E}^x, \mathbf{W}^x)$ in which each node carries a signal $\mathbf{x}^{(l)}$ of the original set $\mathbf{X}$ of signals, while the weights $\mathbf{W}^x_{i,j}$ provide a measure of similarity between each pair of signals $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$. The penalty function — the so-called graph-penalized activation penalty — associated with this approach is formulated as

$$\begin{aligned} g^{(\mathcal{G}^x)}(\mathbf{A}) \; &= \; \alpha_s \sum_{(i,j) \in \mathcal{E}^x} \mathbf{W}^x_{i,j} ||\mathbf{a}^{(i)} - \mathbf{a}^{(j)}||_2^2 \\ &= \; \alpha_s \mathrm{Tr}\big(\mathbf{A}\mathbf{L}^x\mathbf{A}^T\big), \end{aligned}$$

with $\alpha_s > 0$ a penalty parameter and $\mathbf{L}^x$ the Laplacian matrix associated to graph $\mathcal{G}^x$.

2.3.2. DU Step Penalties

A first popular constraint that has been used in graph DL is the *smoothness* penalty on the atoms rather than on the signals themselves. Given a graph $\mathcal{G}$, this constraint, introduced by Yankelevsky [10] — denoted graph-smooth dictionary penalty — is defined as follows

$$f^{(sm)}(\mathbf{D}) \; = \; \beta_s \sum_{m=1}^{M} S_{\mathcal{G}}(\mathbf{d}^{(m)}) \; = \; \beta_s \, \mathrm{Tr}\big(\mathbf{D}^T \, \mathbf{L} \, \mathbf{D}\big), \quad (2)$$

with $\beta_s > 0$ a penalty parameter. The addition of this constraint to the optimization problem allows guiding the learning process toward atoms that are regular with regard to the underlying graph structure.

An alternative approach is to consider atoms of a specific family, imposing a particular structure on the dictionary and learning the parameters of that structure. The structure in question usually incorporates desirable properties of the dictionary such as translation invariance or minimal *consistency* [11]. In particular, let us mention the method introduced by Thanou [12] which proposes to structure the learned dictionary as a set of sub-dictionaries, each one being a polynomial of the graph's Laplacian matrix.

3. PROPOSED METHOD

We now present our variation of the dictionary learning problem motivated by the uncertainty principles discussed in Section 2.1.

3.1. Problem Statement

Activation and dictionary update problems are introduced separately in the two following sections.

3.1.1. AU Step

Similarly to Ramamurthy [9], our method uses the graph-penalized activation penalty $g^{(\mathcal{G}^x)}$, leading to the following

AU step

$$\mathbf{A}^* \;=\; \underset{\mathbf{A}}{\mathbf{argmin}}\; ||\mathbf{X}-\mathbf{D}\mathbf{A}||_F^2 \;+\; \alpha_s\,\mathrm{Tr}\big(\mathbf{A}\,\mathbf{L}^x\mathbf{A}^T\big) \quad (3)$$

This penalty has been retained because it favors the dissimilarity of the atoms and limits the linear dependencies between them. In order to generate a similarity graph $\mathcal{G}^x$ that carries the signal proximities, the weight matrix $\mathbf{W}^x$ is computed, for $i,j=1,\ldots,L$, such that

$$\mathbf{W}^x_{i,j} = \exp\left(-\frac{||\mathbf{x}^{(i)}-\mathbf{x}^{(j)}||^2}{\sigma^2}\right).$$

with σ^2 a parameter that was set to the mean of the pairwise squared distances. Then, the Laplacian matrix $\mathbf{L}^x$ associated to the graph $\mathcal{G}^x$ can be obtained from the weight matrix $\mathbf{W}^x$.

3.1.2. DU Step

Our DU step considers mixed dictionaries, *i.e.* composed of several types of atoms, each type presenting a carefully chosen property. In the continuity of the uncertainty principles introduced in Section 2.1, we propose here to learn dictionaries with two types of atoms: local atoms on the graph, and frequencially local — or *smooth* atoms. Formally, given M_l a number of locality atoms and M_s a number of smooth atoms, the following decomposition is performed

$$\mathbf{D} \;=\; \Big(\mathbf{D}^{(loc)},\mathbf{D}^{(sm)}\Big) \;=\; \Big(\mathbf{d}^{(1)},\ldots,\mathbf{d}^{(M_l+M_s)}\Big),$$

where $\mathbf{D}^{(loc)} \in N\times M_l$ denotes the locality sub-dictionary and $\mathbf{D}^{(sm)} \in N\times M_s$ the smooth sub-dictionary.

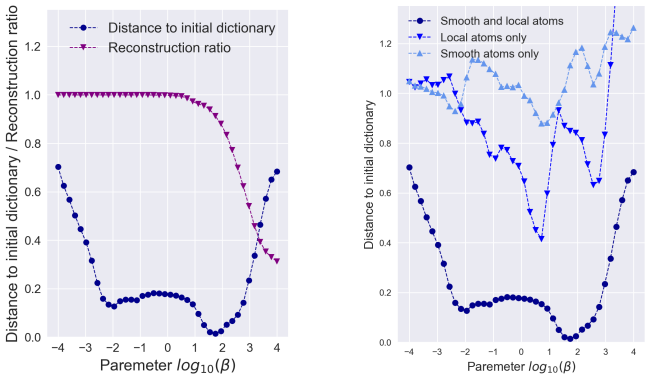
Then, it seems natural to penalize the smooth sub-dictionary $\mathbf{D}^{(sm)}$ learning using $f^{(sm)}$ the graph-smooth dictionary penalty — see Equation (2). However, penalizing the $\mathbf{D}^{(loc)}$ locality sub-dictionary learning requires the introduction of a suitable loss function, which is defined, given a graph $\mathcal{G}$, a graph distance $d^{\mathcal{G}}$ — *e.g.* the graph geodesic distance [13] — and $(c_1,\ldots,c_{M_l})$ a set of M_l centrality nodes of $\mathcal{G}$, as

$$f^{(loc)}(\mathbf{D}) \;=\; \beta_l \sum_{i=1}^{M} \mathrm{Loc}_{c_i}(\mathbf{d}^{(i)}),$$

with $\beta_l > 0$ a penalty parameter and such that $\mathrm{Loc}_{c_i}(\mathbf{x}) = \mathbf{x}^T\mathbf{P}_i\mathbf{x}$, where $\mathbf{P}_i \;=\; \mathrm{Diag}\big(d^{\mathcal{G}}(c_i,v_1)^2,\ldots,d^{\mathcal{G}}(c_i,v_N)^2\big)$. An important question for the penalty is the position of the centers. If prior information on these centers are available, it is possible to guide the dictionary learning by imposing them in a fixed way. Otherwise, centrality nodes can be randomly sampled.

Finally, given an activation matrix $\mathbf{A}\in\mathbb{R}^{(M_l+M_s)\times L}$, our *local-smooth* dictionary learning (LSDL) problem reads like

$$\mathbf{D}^* \;=\; \underset{\mathbf{D}^{(loc)},\mathbf{D}^{(sm)}}{\mathbf{argmin}}\; ||\mathbf{X}-[\mathbf{D}^{(sm)},\,\mathbf{D}^{(loc)}]\mathbf{A}||_2^2 \quad (4)$$
$$+\; f^{(loc)}(\mathbf{D}^{(loc)}) \;+\; f^{(sm)}(\mathbf{D}^{(sm)}).$$



(a) Reconstruction (b) Ablation

Fig. 1. Figure (a) shows reconstruction ratios and distances to the initial dictionary $\mathbf{D}^0$. Figure (b) shows the distances to the initial dictionary $\mathbf{D}^0$ of smooth (light blue), locality (blue) and mixed (dark-blue) learned dictionaries.

3.2. Resolution

Resolution approaches for the activation and dictionary update steps are exposed separately in the following sections.

3.2.1. AU Step

The problem (3) is convex in the matrix $\mathbf{A}$, but does not present a closed-form solution. Thus, gradient descent can be performed, such that the τ-th iteration is given by

$$\mathbf{A}^{(\tau)} = \mathbf{A}^{(\tau-1)} - \theta\nabla_{\mathbf{A}}\left(||\mathbf{X}-\mathbf{D}\mathbf{A}||_F^2 \;+\; g^{(\mathcal{G}^x)}(\mathbf{A})\right),$$

with $\theta>0$ the gradient step, and

$$\nabla_{\mathbf{A}}\left(||\mathbf{X}-\mathbf{D}\mathbf{A}||_F^2 \;+\; g^{(\mathcal{G}^x)}(\mathbf{A})\right) = 2\,\alpha_s\,\mathbf{A}\,\mathbf{L}_x$$
$$-2\,\mathbf{D}^T(\mathbf{X}-\mathbf{D}\mathbf{A}).$$

3.2.2. DU Step

In order to simplify the solution of the problem stated Equation (4), we introduce the vectorized expressions $\mathbf{v}(\mathbf{D}) \in \mathbb{R}^{N(M_l+M_s)}$ of $\mathbf{D}$, and $\mathbf{v}(\mathbf{X})\in\mathbb{R}^{NL}$ of $\mathbf{X}$. Then, the *local-smooth* dictionary update problem can be reformulated as

$$\mathbf{v}(\mathbf{D}^*) \;=\; \underset{\mathbf{d}\in\mathbb{R}^{NM}}{\mathbf{argmin}}\; ||\mathbf{v}(\mathbf{X})-\big(\mathbf{A}\otimes\mathbf{I}_N\big)^T\mathbf{d}||_2^2 \;+\; \mathbf{d}^T\mathbf{Q}\mathbf{d}$$

with $\mathbf{Q} \;=\; \mathrm{Diag}\Big(\beta_l\mathbf{P}_1,\ldots,\beta_l\mathbf{P}_{M_l},\beta_s\mathbf{L},\ldots,\beta_s\mathbf{L}\Big)$ a block-diagonal matrix, where the graph Laplacian matrix $\mathbf{L}$ is repeated M_s times. This formulation highlights the quadratic property of the LSDL problem and allows to derive a solution, as $\mathbf{v}(\mathbf{D}^*) = \Big((\mathbf{A}\mathbf{A}^T)\otimes\mathbf{I}_N+\mathbf{Q}\Big)^{-1}(\mathbf{A}\otimes\mathbf{I}_N)\mathbf{v}(\mathbf{X})$.

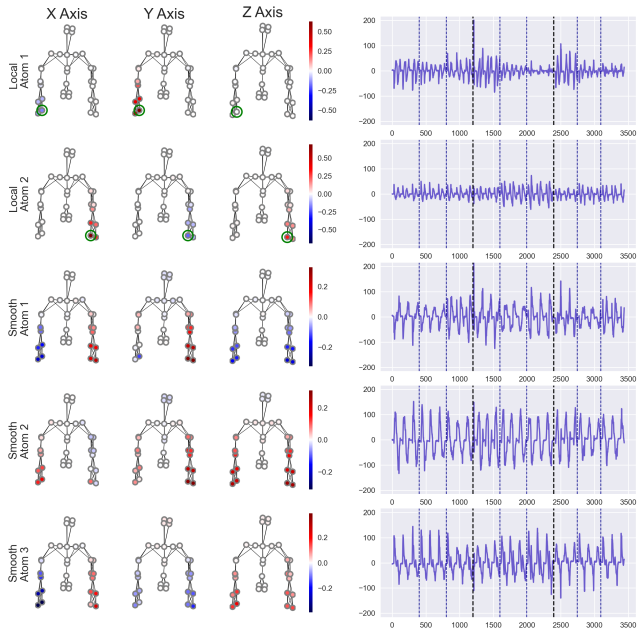


Fig. 2. Atoms learned from all movements are presented on the left. For local atoms, the center is indicated by a green circle. The intensity of an atom on a node is indicated by the color, from blue to red. The activations of each atom are presented as a concatenated time series on the right. Vertical blue lines indicate the separation between two types of movement, and vertical black lines separate different subjects.

4. EXPERIMENTS

4.1. Dictionary Recovery on a Synthetic Graph Signal

We first aim at reconstructing a synthetic dictionary that generated a set of signals. A random geometric graph $\mathcal{G}$ is generated by connecting a set of $N = 500$ points drawn uniformly on $[0, 1]^2$ through a 5-nearest neighbors procedure. The distance associated to $\mathcal{G}$ is the graph-geodesic distance [13].

A set of $M_s = 5$ smooth atoms is sampled from a Gaussian distribution with covariance matrix $\mathbf{L}^{-1}$. $M_l = 20$ nodes are randomly sampled and local signals at a geodesic distance < 2 from the centers are generated from a Gaussian distribution. Atoms are concatenated to form $\mathbf{D}^0$ and the activations $\mathbf{A}^0$ are drawn from a normal distribution, with $L = 1000$. Finally, the data matrix is computed as $\mathbf{X} = \mathbf{D}^0\mathbf{A}^0$.

The LSDL algorithm is applied for different values of the penalization parameters $\beta = \beta_s = \beta_l$ ranging from 10^{-4} to 10^4. Since center nodes are supposed to be unknown, 100 nodes are uniformly sampled and used as centers. Two metrics are computed from the learned dictionary $\mathbf{D}^*$ and activation matrix $\mathbf{A}^*$: (i) the reconstruction ratio $1 - ||\mathbf{X}^0 - \mathbf{D}^*\mathbf{A}^*||_2/||\mathbf{X}^0||_2$ quantifies how well the dictionary/activation pair represents the signals $\mathbf{X}^0$; and (ii) the distance $||\mathbf{D}^0 - \mathbf{D}^*||_2/(M_l + M_s)$ between the initial dictionary $\mathbf{D}^0$ and the learned dictionary $\mathbf{D}^*$ — to which only the $M_s = 20$ most active local atoms have been kept.

Three distinct behaviors appear on Figure 1.a. For β ranging from 10^{-4} to 10^{-2}, the parameters are not important enough to affect the learning algorithm. For parameters higher than 10^2, penalties enforce too smooth and local atoms to be learned. Finally, for parameters ranging from 10^{-2} to 10^2 the reconstruction ratio remains unchanged, while the distance to the initial dictionary decreases down to 3.10^{-2}.

Figure 1.b presents a brief ablation study: two partial dictionaries, one containing 15 smooth atoms, the other 15 locality atoms were learned from the data generated by the initial dictionary $\mathbf{D}^0$. The large distance to the initial dictionary for all the parameters explored highlights the complementary nature of the locality and smoothness penalties employed, forcing the learning of atoms with non-redundant behaviors.

4.2. Analysis of a Movement Dataset

The studied dataset includes 34 sensors' 3-dimensional speed time-series — sampled at 10 Hz — recorded from 3 participants while performing seated, bilateral, arm elevation in the sagital, scapular and frontal planes. Each movement was performed three times, with a rest time of ~ 3 sec.

The sensor graph $\mathcal{G}^{(sensor)}$ is generated through a 4-nearest-neighbors procedure with regards to the maximal pairwise sensors distance. As the process is three-dimensional, the signals must lie on a graph of size $3 \times N$ — one copy of the sensors graph for each dimension x, y and z. The resulting graph is not connected, and signal variations are evaluated on each spatial dimension independently.

Then, a set of $M_s = 3$ smooth atoms and $M_l = 2$ local atoms are learned around 2 different centers — displayed in Figure 2. The scaling parameters for the locality loss penalty $\beta_l = 2.10^2$ and for the smoothness penalty $\beta_s = 10^3$ were chosen through cross-validation. The atoms and activations learned are presented in Figure 2.

The choice was made to restrict the number of learned atoms in order to capture more generic phenomenons in a common small dictionary. Despite this limitation, the decomposition achieved a reconstruction ratio of 0.9. The choice of movements that involve arm elevation in the different planes has led to atom intensities that are evenly spread on the 3 dimensions, heavily correlated for most atoms — as the movements do not involve a displacement in a single dimension.

The proposed method provides an interpretable model that handles the high multiplicity of sensors through smoothness constraints and for which the compensatory components are assumed to be local and thus interpretable. For this exploratory experiment, intra/inter-subject differences appear clearly in well-chosen spaces, and the different atoms produced provide an accessible basis for a clinician.

5. REFERENCES

[1] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[2] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[3] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 689–696.

[4] Lars Backstrom and Jure Leskovec, "Supervised random walks: predicting and recommending links in social networks," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 635–644.

[5] Hamidreza Sadreazami, Arash Mohammadi, Amir Asif, and Konstantinos N Plataniotis, "Distributed-graph-based statistical approach for intrusion detection in cyber-physical systems," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 137–147, 2017.

[6] Monica Agrawal, Marinka Zitnik, and Jure Leskovec, "Large-scale analysis of disease pathways in the human interactome," in *PACIFIC SYMPOSIUM on BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*. World Scientific, 2018, pp. 111–122.

[7] Ameya Agaskar and Yue M Lu, "A spectral graph uncertainty principle," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4338–4356, 2013.

[8] Ingrid Daubechies, Michel Defrise, and Christine De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 57, no. 11, pp. 1413–1457, 2004.

[9] Karthikeyan Natesan Ramamurthy, Jayaraman J Thiagarajan, Prasanna Sattigeri, and Andreas Spanias, "Learning dictionaries with graph embedding constraints," in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2012, pp. 1974–1978.

[10] Yael Yankelevsky and Michael Elad, "Dual graph regularized dictionary learning," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 4, pp. 611–624, 2016.

[11] Mehrdad Yaghoobi, Laurent Daudet, and Mike E Davies, "Parametric dictionary design for sparse coding," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4800–4810, 2009.

[12] Dorina Thanou, David I Shuman, and Pascal Frossard, "Parametric dictionary learning for graph signals," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 487–490.

[13] Jérémie Bouttier, Philippe Di Francesco, and Emmanuel Guitter, "Geodesic distance in planar graphs," *Nuclear physics B*, vol. 663, no. 3, pp. 535–567, 2003.