Fast Approximate Spectral Clustering for Dynamic Networks

Lionel Martin [1] **Andreas Loukas** [1] **Pierre Vandergheynst** [1]

Abstract

Spectral clustering is a widely studied problem, yet its complexity is prohibitive for dynamic graphs of even modest size. We claim that it is possible to reuse information of past cluster assignments to expedite computation. Our approach builds on a recent idea of sidestepping the main bottleneck of spectral clustering, i.e., computing the graph eigenvectors, by using fast Chebyshev graph filtering of random signals. We show that the proposed algorithm achieves clustering assignments with quality approximating that of spectral clustering and that it can yield significant complexity benefits when the graph dynamics are appropriately bounded.

1. Introduction

Spectral clustering (SC) is one of the most well-known methods for clustering multivariate data, with numerous applications in biology (e.g., protein-protein interactions, gene co-expression) and social sciences (e.g., call graphs, political study) among others (Von Luxburg, 2007; Fortunato, 2010). However, because of its inherent dependence on the spectrum of some large graph, SC is also notoriously slow. This has motivated a surge of research focusing in reducing its complexity, for example using matrix sketching methods (Fowlkes et al., 2004; Li et al., 2011; Gittens et al., 2013) and more recently compressive sensing techniques (Ramasamy & Madhow, 2015; Tremblay et al., 2016).

Yet, the clustering complexity is still problematic for dynamic graphs, where the edge set is a function of time. Temporal dynamics constitute an important aspect of many network datasets and should be taken into account in the algorithmic design and analysis. Unfortunately, SC is poorly suited to this setting as eigendecomposition –its main computational bottleneck– has to be recomputed from scratch whenever the graph is updated, or at least periodically (Ning et al., 2007). This is a missed opportunity since the clustering assignments of many real networks change slowly with time, suggesting that successive algorithmic runs wastefully repeat similar computations.

Motivated by this observation, this paper proposes an algorithm that reuses information of past cluster assignments to expedite computation. Different from previous work on dynamic clustering, our objective is *not* to improve the clustering quality, for example by enforcing a temporal-smoothness hypothesis (Chakrabarti et al., 2006; Chi et al., 2007) or by using tensor decompositions (Gauvin et al., 2014; Tu et al., 2016). On the contrary, we focus entirely on decreasing the computational overhead and aim to produce assignments that are provably close to those of SC.

Our work is inspired by the recent idea of sidestepping eigendecomposition by utilizing as features random signals that have been filtered over the graph (Tremblay et al., 2016). Our main argument is that, instead of computing the clustering assignment of a graph G_1 using d filtered signals as features, one may utilize a percentage of features of a different graph G_2 without significant loss in accuracy, as long as G_1 and G_2 are appropriately close. This leads to a natural clustering scheme for time-varying topologies: each new instance of the dynamic graph is clustered using pd signals computed previously and only $(1-p)d$ new filtered signals, where p is a percentage. Moreover, inspired by similar ideas we can also attain further complexity reductions with respect to the graph filter design, i.e., by identifying the k-th eigenvalue.

Concretely, we provide the following contributions:

1. In Section 3 we refine the analysis of compressive spectral clustering (CSC) presented in (Tremblay et al., 2016). Our goal is to move from assertions about distance preservation to guarantees about the quality of the solution of CSC itself. We prove that with probability at least $1 - \exp(-t^2/2)$, the quality of the clustering assignments of CSC and SC differ by less than $2\sqrt{k/d}\,(\sqrt{k}+t)$. Our analysis suggests that $d = \mathcal{O}(k^2)$ filtered signals are sufficient to guarantee a good approximation, while not making any restricting assumptions about the graph structure, e.g., assuming a stochastic block model (Pydi & Dukkipati, 2017).

2. In Section 4, we focus on dynamic graphs and propose dynamic CSC, an algorithm that reuses information of past cluster assignments to expedite computation. We discover

[1]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. Correspondence to: Lionel Martin <lionel.martin@epfl.ch>.

Technical report. Preliminary work only.

that the algorithm's ability to reuse features is inherently determined by a metric of spectral similarity ρ between consecutive graphs. Indeed, we prove that, when pd features are reused, the clustering assignment quality of dynamic CSC approximates with high probability that of CSC up to an additive term in the order of $p\rho$.

3. We complement our analysis with a numerical evaluation in Section 5. Our experiments illustrate that dynamic CSC yields in practice computational benefits when the graph dynamics are bounded, while producing assignments with quality closely approximating that of SC.

2. Background

We start by briefly summarizing the standard method for spectral clustering as well as the idea behind the more recent fast (compressive) methods.

2.1. Spectral clustering (SC)

To determine the best node-to-cluster assignment, spectral clustering entails solving a k-means problem, with the eigenvectors of the graph Laplacian $\mathbf{L}$ as features.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ be a simple symmetric undirected weighted graph with a fixed set of vertices $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ of cardinality n, and a set of m edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ where the edge between v_i and v_j has weight $\mathbf{W}_{i,j} > 0$ if it exists, and $\mathbf{W}_{i,j} = 0$ otherwise. Some versions of spectral clustering make use of the combinatorial Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and others of the normalized Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, see e.g., (Ng et al., 2002; Shi & Malik, 2000). Here, $\mathbf{D}$ is the diagonal matrix whose entries are the degree of the nodes in the graph. We denote the eigendecomposition of the Laplacian of choice by $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, with the diagonal entries of $\mathbf{\Lambda}$ sorted in non-decreasing order, such that $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

Spectral clustering consists of computing the first k eigenvectors of $\mathbf{L}$ arranged in a matrix called $\mathbf{U}_k$ and subsequently computing a k-means assignment of the n vectors of size k found in the rows of $\mathbf{U}_k$. Formally, if $\mathbf{\Phi} \in \mathbb{R}^{n \times d}$ is the feature matrix (here $\mathbf{\Phi} = \mathbf{U}_k$ and $d = k$), and k is a positive integer denoting the number of clusters, the k-means clustering problem finds the indicator matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$ which satisfies

$$\mathbf{X}_\Phi = \underset{\mathbf{X} \in \mathcal{X}}{\arg\min} \|\mathbf{\Phi} - \mathbf{X}\mathbf{X}^\top \mathbf{\Phi}\|_F, \tag{1}$$

with associated cost $C_\Phi = \|\mathbf{\Phi} - \mathbf{X}_\Phi \mathbf{X}_\Phi^T \mathbf{\Phi}\|_F$. Symbol $\mathcal{X}$ denotes the set of all $n \times k$ indicator matrices $\mathbf{X}$. These matrices indicates the cluster membership of each data point by setting

$$\mathbf{X}_{i,j} = \begin{cases} \frac{1}{\sqrt{s_j}} & \text{if data point } i \text{ belongs to cluster } j \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where s_j is the size of cluster j, also equals to the number of non-zero elements in column j. Note that the cost described in eq. (1) is the square root of the more traditional definition expressed with the distances to the cluster centers (Cohen et al., 2015, Sec 2.3). We refer the reader to the work by Boutsidis et al. (2015) and its references for more details.

2.2. Compressive spectral clustering (CSC)

To reduce the cost of spectral clustering, i.e. $\mathcal{O}(kn^2)$, Tremblay et al. (2016) and Boutsidis et al. (2015) independently proposed to fasten spectral clustering using approximated eigenvectors based on random signals. The former also introduced the benefits of compressive sensing techniques reducing the total cost down to $\mathcal{O}(k^2 \log^2(k) + mn(\log(n) + k)$, where m is the order of the polynomial approximation. Their argument consists of two steps:

Step 1. Approximate features. The costly to compute feature matrix $\mathbf{\Phi} = \mathbf{U}_k$ is approximated by the projection of a random matrix over the same subspace. In particular, let $\mathbf{R} \in \mathbb{R}^{N \times d}$ be a random (gaussian) matrix with centered i.i.d. entries, each having variance $\frac{1}{d}$. We can project $\mathbf{R}$ onto $span\{\mathbf{U}_k\}$ by filtering each one of its columns by a low-pass graph filter $g(\mathbf{L}) = \mathbf{H}$ defined as

$$\mathbf{H} = \mathbf{U} \begin{pmatrix} \mathbf{I}_k & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U}^\top. \tag{3}$$

It is then a simple consequence of the Johnshon-Lindenstrauss lemma that the rows $\boldsymbol{\psi}_i^\top$ of matrix $\mathbf{\Psi} = \mathbf{H}\mathbf{R}$ can act as a replacement of the features used in spectral clustering, i.e., the rows $\boldsymbol{\phi}_i^\top$ of $\mathbf{\Phi} = \mathbf{U}_k$.

Theorem 2.1 (adapted from (Tremblay et al., 2016))**.** *For every two nodes v_i and v_j the restricted isometry relation*

$$(1-\varepsilon)\|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\|_2 \leq \|\boldsymbol{\psi}_i - \boldsymbol{\psi}_j\|_2 \leq (1+\varepsilon)\|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j\|_2 \tag{4}$$

holds with probability larger than $1 - n^{-\beta}$, as long as the dimension is $d > \frac{4+2\beta}{\varepsilon^2/2 - \varepsilon^3/3} \log(n)$.

We note that, even though $\mathbf{H}\mathbf{R}$ is also expensive to compute, it can be approximated in $O(|\mathcal{E}|dm)$ number of operations using Chebychev polynomials (Shuman et al., 2011a; Hammond et al., 2011), resulting in a small additive error that decreases with the polynomial order.

Step 2. Compressive k-means. The complexity is reduced further by computing the k-means step for only a subset of the nodes. The remaining cluster assignments are then

inferred by solving a Tikhonov regularization problem involving k additional graph filtering operations, each with a cost linear in $m|\mathcal{E}|$.

To guarantee a good approximation, it is sufficient to select $\mathcal{O}(\nu_k^2 \log(k))$ nodes uniformly at random, where $\nu_k = \sqrt{n}\max_i \|\boldsymbol{\phi}_i\|_2$ is the global cumulative coherence. However as shown by Puy et al. (2016), it is always possible to sample $\mathcal{O}(k\log(k))$ nodes using a different distribution (variable density sampling).

In the following, we will present our theoretical results with respect to the non-compressed version of their algorithm.

3. The approximation quality of static CSC

Before delving to the dynamic setting, we refine the analysis of compressive spectral clustering. Our objective is to move from assertions about distance preservation currently known (see Thm. 2.1) to guarantees about the quality of the solution of CSC itself. Formally, let

$$\mathbf{X}_\Psi = \operatorname*{arg\,min}_{\mathbf{X}\in\mathcal{X}} \|\boldsymbol{\Psi} - \mathbf{X}\mathbf{X}^\top\boldsymbol{\Psi}\|_F. \tag{5}$$

be the clustering assignment obtained from using k-means with $\boldsymbol{\Psi}$ as features (CSC assignment), and define the *CSC cost* C_Ψ as

$$C_\Psi = \|\boldsymbol{\Phi} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top\boldsymbol{\Phi}\|_F. \tag{6}$$

The question we ask is: *how close is C_Ψ to the cost C_Φ of the same problem, where the assignment has been computed using $\boldsymbol{\Phi}$ as features, i.e., the SC cost corresponding to* (1)*?* Note that we choose to express the approximation quality in terms of the difference of clustering assignment costs and not of the distance between the assignments themselves. This has the benefit of not penalizing approximation algorithms that choose alternative assignments of the same quality.

This section is devoted to the analysis of the quality of the assignments outputted by CSC compared to those of SC for the same graph. Our central theorem, stated below, asserts that with high probability the two costs are close.

Theorem 3.1. *The SC cost C_Φ and the CSC cost C_Ψ are related by*

$$C_{\boldsymbol{\Phi}} \leq C_\Psi \leq C_{\boldsymbol{\Phi}} + 2\sqrt{\frac{k}{d}}(\sqrt{k}+t), \tag{7}$$

with probability at least $1-\exp(-t^2/2)$.

The result above emphasizes the importance of the number of filtered signals d and directly links it to the distance with the optimal assignment for the spectral features. Indeed, one can see that the difference between the two costs vanishes when d is sufficiently large. Importantly, setting $d = \Omega(k^2)$ guarantees a small error. Keeping in mind that the complexity of CSC is $\mathcal{O}(k^2\log^2(k) + mn(\log(n)+k)$, we see that our result implies that CSC is particularly suitable when the number of desired cluster is small, e.g., $k = O(1)$ or $k = O(\log n)$.

3.1. The approximation quality of CSC

The first step in proving Thm. 3.1 is to establish the relation between C_Φ and C_Ψ. The following lemma relates the two costs by an additive error term that depends on the feature's differences $\|\boldsymbol{\Psi} - \boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F$[1]. Since $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ have different sizes we introduced the multiplication by a unitary matrix $\mathbf{Q}$. We will first show that any unitary $\mathbf{Q}$ can be picked in Lem. 3.1 and then derive the optimal $\mathbf{Q}$, the one minimizing the additive term, in Thm. 3.2.

Lemma 3.1. *For any unitary matrix $\mathbf{Q} \in \mathbb{R}^{d\times d}$, the SC cost C_Φ and the CSC cost C_Ψ are related by*

$$C_\Phi \leq C_\Psi \leq C_\Phi + 2\|\boldsymbol{\Psi} - \boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F, \tag{8}$$

where, the matrix $\mathbf{I}_{\ell\times m}$ of size $\ell \times m$ above contains only ones on its diagonal and serves to resize matrices.

Being able to show that the additive term is small encompasses the result of Thm. 2.1, ensuring distance preservation. However, this statement is stronger than the previous one as our lemma is not necessarily true under distance preservation only.

Proof. Let $\mathbf{X}_\Phi$ and $\mathbf{X}_\Psi$ be respectively the SC and CSC clustering assignments. Moreover, we denote for compactness the additive error term by $\mathbf{E} = \boldsymbol{\Psi} - \boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}$. We have that

$$\begin{aligned}
C_\Psi &= \|\boldsymbol{\Phi} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top\boldsymbol{\Phi}\|_F \\
&= \|(\mathbf{I} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top)(\boldsymbol{\Psi} - \mathbf{E})\|_F \\
&\leq \|(\mathbf{I} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top)\boldsymbol{\Psi}\|_F + \|(\mathbf{I} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top)\mathbf{E}\|_F \\
&\leq \|(\mathbf{I} - \mathbf{X}_\Psi\mathbf{X}_\Psi^\top)\boldsymbol{\Psi}\|_F + \|\mathbf{E}\|_F \\
&\leq \|(\mathbf{I} - \mathbf{X}_\Phi\mathbf{X}_{\boldsymbol{\Phi}}^\top)\boldsymbol{\Psi}\|_F + \|\mathbf{E}\|_F \\
&= \|(\mathbf{I} - \mathbf{X}_\Phi\mathbf{X}_\Phi^\top)(\boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q} + \mathbf{E})\|_F + \|\mathbf{E}\|_F \\
&\leq \|(\mathbf{I} - \mathbf{X}_\Phi\mathbf{X}_\Phi^\top)\boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F + 2\,\|\mathbf{E}\|_F \\
&= C_\Phi + 2\,\|\boldsymbol{\Psi} - \boldsymbol{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F
\end{aligned} \tag{9}$$

The lower bound directly comes from the fact that $\mathbf{X}_\Phi$ in eq. (1) defines the argmin of our cost functions thus $C_\Phi \leq C_\Psi$. □

[1] We assume all along that $d \geq k$ but a similar result holds when $d < k$. In this case, we can consider the term $\|\boldsymbol{\Psi}\mathbf{I}_{d\times k}\mathbf{Q} - \boldsymbol{\Phi}\|_F$ and derive the optimal unitary $\mathbf{Q}$ in order to obtain the same result as Thm. 3.2. However there is little interest in practice since one cannot expect the recovery of k eigenvectors with less random filtered signals as shown in (Paratte & Martin, 2016).

The remaining of this section is devoted to bounding the Frobenius error $\|\mathbf{\Psi} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F$ between the features of SC and CSC. In order to prove this result, we will first express our Frobenius norm exclusively in terms of the singular values of the random matrix $\mathbf{R}$ and then in a second step we will study the distribution of these singular values.

Our next result, which surprisingly is an equality, reveals that the achieved error is exactly determined by how close a Gaussian matrix is to a unitary matrix.

Theorem 3.2. *There exists a $d \times d$ unitary matrix $\mathbf{Q}$, such that*

$$\|\mathbf{\Psi} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F = \|\mathbf{\Sigma} - \mathbf{I}_{k\times d}\|_F, \tag{10}$$

where $\mathbf{\Sigma}$ is the diagonal matrix holding the singular values of $\mathbf{R}' = \mathbf{I}_{k\times n}\mathbf{U}^\top\mathbf{R}$.

Before presenting the proof, let us observe that $\mathbf{R}'$ is an i.i.d. Gaussian random matrix of size $k \times d$ and its entries have zero mean and the same variance as that of $\mathbf{R}$. We use this fact in the following to control the error by appropriately selecting the number of random signals d.

Proof. Let us start by noting that, by the unitary invariance of the Frobenius norm, for any $k \times k$ matrix $\mathbf{M}$

$$\|\mathbf{\Phi}\mathbf{M}\|_F = \|\mathbf{U}\mathbf{I}_{n\times k}\mathbf{M}\|_F = \|\mathbf{I}_{n\times k}\mathbf{M}\|_F = \|\mathbf{M}\|_F. \tag{11}$$

We can thus rewrite the feature error as

$$\begin{aligned}\|\mathbf{\Psi} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F &= \|\mathbf{\Phi}\mathbf{\Phi}^\top\mathbf{R} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &= \|\mathbf{\Phi}^\top\mathbf{R} - \mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &= \|\mathbf{I}_{k\times n}\mathbf{U}^\top\mathbf{R} - \mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &= \|\mathbf{R}' - \mathbf{I}_{k\times d}\mathbf{Q}\|_F. \end{aligned} \tag{12}$$

We claim that there is a unitary matrix $\mathbf{Q}$ that satisfies eq. (10). We describe this matrix as follows. Let $\mathbf{R}' = \mathbf{Q}_L\mathbf{\Sigma}\mathbf{Q}_R^\top$ be the singular value decomposition of $\mathbf{R}'$ and set

$$\mathbf{Q} = \begin{pmatrix}\mathbf{Q}_L & 0\\ 0 & \mathbf{I}_{d-k}\end{pmatrix}\mathbf{Q}_R^\top. \tag{13}$$

Substituting this to the feature error, we have that

$$\begin{aligned}\|\mathbf{R}' - \mathbf{I}_{k\times d}\mathbf{Q}\|_F &= \|\mathbf{Q}_L\mathbf{\Sigma}\mathbf{Q}_R^\top - \mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &= \|\mathbf{\Sigma} - \mathbf{Q}_L^\top\mathbf{I}_{k\times d}\mathbf{Q}\mathbf{Q}_R\|_F \\ &= \|\mathbf{\Sigma} - \mathbf{Q}_L^\top\mathbf{I}_{k\times d}\begin{pmatrix}\mathbf{Q}_L & 0\\ 0 & \mathbf{I}_{d-k}\end{pmatrix}\mathbf{Q}_R^\top\mathbf{Q}_R\|_F \\ &= \|\mathbf{\Sigma} - \mathbf{Q}_L^\top\begin{pmatrix}\mathbf{Q}_L & 0\end{pmatrix}\|_F \\ &= \|\mathbf{\Sigma} - \mathbf{I}_{k\times d}\|_F, \end{aligned} \tag{14}$$

which is the claimed result. □

To bound the feature error further, we will use the following result by Vershynin, whose proof is not reproduced.

Corollary 3.1 (adapted from Cor. 5.35 (Vershynin, 2010))**.** *Let $\mathbf{N}$ be an $d \times k$ matrix whose entries are independent standard normal random variables. Then for every $t, i \geq 0$, with probability at least $1 - \exp(-t^2/2)$ one has*

$$\sigma_i(\mathbf{N}) - \sqrt{d} \leq \sqrt{k} + t, \tag{15}$$

where $\sigma_i(\mathbf{N})$ is the ith singular value of $\mathbf{N}$.

Exploiting this result, the following corollary of Thm. 3.2 reveals the relation of the feature error and the number of random signals d.

Corollary 3.2. *There exists a $d \times d$ unitary matrix $\mathbf{Q}$, such that, for every $t \geq 0$, one has*

$$\|\mathbf{\Psi} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F \leq \sqrt{\frac{k}{d}}\,(\sqrt{k} + t), \tag{16}$$

with probability at least $1 - \exp(-t^2/2)$.

Proof. To obtain the following extremal inequality for the singular values of $\mathbf{R}'$, we note that $\mathbf{R}'$ is composed of i.i.d. Gaussian random variables with zero mean and variance $1/d$, and thus use Cor. 3.1 setting $\mathbf{R}' = \mathbf{N}/d$ and thus for every i,

$$\begin{aligned}\sigma_i(\mathbf{R}') &= \sigma_i(\mathbf{N})/\sqrt{d} \\ &\leq \frac{\sqrt{d} + \sqrt{k} + t}{\sqrt{d}} = 1 + \frac{\sqrt{k} + t}{\sqrt{d}}. \end{aligned} \tag{17}$$

By simple algebraic manipulation, we then find that

$$\begin{aligned}\|\mathbf{\Sigma} - \mathbf{I}_{k\times d}\|_F^2 &= \sum_{i=1}^{k}\left(\sigma_i(\mathbf{R}') - 1\right)^2 \\ &\leq k\left(\frac{\sqrt{k} + t}{\sqrt{d}}\right)^2 = \frac{k}{d}(\sqrt{k} + t)^2, \end{aligned} \tag{18}$$

which, after taking a square root, matches the claim. □

Finally, Cor. 3.2 combined with Lem. 3.1 provide the direct proof of Thm. 3.1 that we introduced earlier.

Before proceeding, we would like to make some remarks about the tightness of the bound. First, guaranteeing that the feature error is small is a stronger condition than distance preservation (though necessary for a complete analysis of CSC). For this reason, the bound derived can be larger than that of Thm. 2.1. Nevertheless, we should stress it is tight: the only inequality in our analysis stems from bounding the k largest singular values of the random matrix by Vershynin's tight bound of the maximal singular value.

3.2. Practical aspects

The study presented above assumes the use of an ideal low-pass filter $\mathbf{H}$ of cut-off frequency λ_k. In practice however, we opt to use the computationally inexpensive Chebyshev graph filters (Shuman et al., 2011a), which approximate low-pass responses using polynomials. In this case, the used filter takes the form $\tilde{\mathbf{H}}_k = \mathbf{U}h(\mathbf{\Lambda})\mathbf{U}^\top$, where $h(\cdot)$ is a polynomial function acting on the diagonal entries of $\mathbf{\Lambda}$. Nevertheless, it is not difficult to see that, when the filter approximation is tight, the clustering quality is little affected.

In particular, letting $\tilde{\mathbf{\Psi}} = \tilde{\mathbf{H}}_k\mathbf{R}$ the feature error becomes

$$\|\tilde{\mathbf{\Psi}} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F \leq \|\tilde{\mathbf{\Psi}} - \mathbf{\Psi}\|_F + \|\mathbf{\Psi} - \mathbf{\Phi}\mathbf{I}_{k\times d}\mathbf{Q}\|_F. \quad (19)$$

We recognize the second term that is exactly the result of Cor. 3.2 and focus thus on the first term.

$$\begin{aligned}\|\tilde{\mathbf{\Psi}} - \mathbf{\Psi}\|_F &\leq \|\mathbf{U}(h(\mathbf{\Lambda}) - \mathbf{I}_{n\times k}\mathbf{I}_{k\times n})\mathbf{U}^\top\mathbf{R}\|_F \\ &= \|(h(\mathbf{\Lambda}) - \mathbf{I}_{n\times k}\mathbf{I}_{k\times n})\,\mathbf{R}\|_F \\ &\leq \|h(\mathbf{\Lambda}) - \mathbf{I}_{n\times k}\mathbf{I}_{k\times n}\|_2\,\|\mathbf{R}\|_F. \end{aligned} \quad (20)$$

An extension of Thm. 3.1, taking into account filter approximation, can thus be derived where eq. (20) would read with probability as least $1 - \exp(-dt^2/2)$:

$$\|\tilde{\mathbf{\Psi}} - \mathbf{\Psi}\|_F \leq \mathcal{O}\left(m^{-m}(\sqrt{n} + t)\right), \quad (21)$$

where m is the order of the polynomial, $\|h(\mathbf{\Lambda}) - \mathbf{I}_{n\times k}\mathbf{I}_{k\times n}\|_2$ reduces to the approximation error of a steep sigmoid that can be bounded using (Shuman et al., 2011b, Proposition 3) and $\|R\|_F$ is bounded in (Laurent & Massart, 2000, Lemma 1). The details are left out due to space constraints.

We notice that the cost of the approximation of ideal low-pass filter depends directly on the quality of the filter. Indeed, the overall error rises with the discrepancies with respect to the ideal filter as shown in eq. (20). Interestingly, the determination of λ_k is also very important because a correct approximation will reduce the number of non-zero eigenvalues and thus the effect of the approximated filter in the very last term of the same equation. Towards these goals, we refer the readers to (Di Napoli et al., 2016; Paratte & Martin, 2016) and their respective eigencount techniques that allow to approximate the filter in $\mathcal{O}(sm|\mathcal{E}|\log(n))$ operations where s is the number of required iterations and m the order of the polynomial.

4. Compressive clustering of dynamic graphs

In this section, we consider the problem of spectral clustering a sequence of graphs. We focus on graphs $\mathcal{G}_t$ where $t \in \{1, \ldots, \tau\}$, composed of a static vertex set $\mathcal{V}$ and evolving edge sets $\mathcal{E}_t$.

Identifying each assignment from scratch (using SC or CSC) is in this context a computationally demanding task, as the complexity increases linearly with the number of time-steps. In the following, we exploit two alternative metrics of similarity between graphs at consecutive time-steps in order to reduce the computational cost of clustering.

Definition 4.1 (Metrics of graph similarity)**.** *Two graphs $\mathcal{G}_{t-1}$ and $\mathcal{G}_t$ are:*

- *(ρ, k)-**spectrally similar** if the spaces spanned by their first k eigenvectors are almost aligned*

$$\|\mathbf{H}_t - \mathbf{H}_{t-1}\|_F \leq \rho. \quad (22)$$

- *ρ-**edge similar** if the edge-wise difference of their Laplacians is less than ρ*

$$\|\mathbf{L}_t - \mathbf{L}_{t-1}\|_F \leq \rho. \quad (23)$$

We argue that both metrics of similarity are relevant in the context of dynamic clustering. Two spectrally similar graphs might have very different connectivity in terms of their detailed structure, but possess similar clustering assignments. On the other hand, assuming that two graphs are edge similar is a stronger condition that postulates fine-grained similarities between them. It is however more intuitive and computationally inexpensive to ascertain.

4.1. Algorithm

We now present an accelerated method for the assignment of the nodes of an evolving graph. Without loss of generality, suppose that we need to compute the assignment for $\mathcal{G}_t$ while knowing already that of $\mathcal{G}_{t-1}$ and possessing the features that served to compute it. Our approach will be to provide an assignment for graph $\mathcal{G}_t$ that reuses (partially) the features $\mathbf{\Psi}_{t-1}$ computed at step $t-1$. Let p be a number between zero and one, and set $q = 1 - p$. Instead of recomputing $\mathbf{\Psi}_t$ from scratch running a new CSC routine, we propose to construct a feature matrix $\mathbf{\Theta}_t$ which consists of dq new features (corresponding to $\mathcal{G}_t$) and dp randomly selected features pertaining to graph $\mathcal{G}_{t-1}$:

$$\begin{aligned}\mathbf{\Theta}_t &= \begin{pmatrix}\mathbf{H}_{t-1}\mathbf{R}_{dp} & \mathbf{H}_t\mathbf{R}_{dq}\end{pmatrix} \\ &= \mathbf{\Psi}_{t-1}\mathbf{S}_{dp}^d + \mathbf{\Psi}_t\overline{\mathbf{S}_{dp}^d}\end{aligned} \quad (24)$$

where we used the sub-identity matrix $\mathbf{S}_{dp}^d = \mathbf{I}_{d\times dp}\mathbf{I}_{dp\times d}$ and its complement $\overline{\mathbf{S}_{dp}^d} = \mathbf{I}_{d\times d} - \mathbf{S}_{dp}^d$.

We noticed that an important part of the complexity of CSC is intrinsic to the determination of λ_k (step 1 of their algorithm). We propose to benefit from the dynamic setting to avoid recomputing it at each step. We propose to admit that

Algorithm 1 Dynamic Compressive Spectral Clustering

Input: $(\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_\tau), p, d$
Output: $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_\tau)$

1: Determine h_k^1 the filter approximation for $\mathcal{G}_1$
2: Find an assignment $\mathbf{X}_1$ for $\mathcal{G}_1$ using CSC and h_k^1
3: **for** t from 2 to τ **do**
4: Randomly pick dp filtered signals generated on $\mathcal{G}_{t-1}$
5: Generate dq feature vectors by filtering as many random signals on $\mathcal{G}_t$ with h_k^{t-1}
6: Compute the eigencount on the features of step 5
7: Refine h_k^t if the eigencount is wrong, else keep h_k^{t-1}
8: Combine these two sets of features to find an assignment $\mathbf{X}_t$ using CSC and h_k^t
9: **end for**

the previous value for λ_k is a good candidate for the filter at the next step, use it to filter the new random signals and validate whether it suits the new graph. Indeed, the eigencount method requires exactly the result of the step 5 of our algorithm to determine if λ_k was correctly determined. We thus compute the new filtered signals and proceed if the eigencount using the new signals is close enough to k. Otherwise, we suggest to use the knowledge of the previous result and perform a dichotomy with this additional knowledge following (Di Napoli et al., 2016). The final set of features generated in the eigencount now serves as $\mathbf{\Psi}_t$.

The method is sketched in Algo. 4.1. For simplicity, in the following we set $p \leq 0.5$ such that the reused features always correspond to $\mathcal{G}_{t-1}$ (and not to some previous timestep).

Complexity analysis. We describe now the complexity of our method and compare it to that of Compressive Spectral Clustering. For simplicity, we focus in a first step on the aspects that do not involve compression. Note that the first graph in the time-series is computed following exactly the procedure of CSC. However, starting from the second graph, there are two steps where the complexity is reduced with respect to CSC. First, the optimization proposed for the determination of λ_k avoids computing s steps of dichotomy for every graph. We claim that spectrally similar graphs must possess close spectrum, thus close values for λ_k. One could then expect to recompute λ_k from time to time only and that when doing so, benefit from a reduced number of iterations due to the proximity. We call S the total number of steps that we gain. Since one step costs $\mathcal{O}(m|\mathcal{E}|\log(n))$ the total gain is $\mathcal{O}(mS|\mathcal{E}|\log(n))$. Second, since we reuse random filtered signals from one graph to the next, the total number of computed random signals will necessarily be reduced compared to the use of τ independent CSC calls. The gain here is $\mathcal{O}(m|\mathcal{E}|dp)$ per timestep. Finally, all reductions applied through compression can also benefit to our dynamic method. Indeed, we theoretically showed that reusing features from the past can replace the creation of new random signals. Thus, sampling the combination of old and new signals can be applied exactly as defined in CSC. Then, the result of the subassignment can be interpolated also as defined in (Tremblay et al., 2016).

4.2. Analysis of dynamic CSC

Similarly to the static case, our objective is to provide probabilistic guarantees about the approximation quality of the proposed method. Let

$$\mathbf{X}_{\Theta_t} = \arg\min_{\mathbf{X} \in \mathcal{X}} \|\mathbf{\Theta}_t - \mathbf{X}\mathbf{X}^\top \mathbf{\Theta}_t\|_F. \tag{25}$$

be the clustering assignment obtained from using k-means with $\mathbf{\Theta}_t$ as features, and define the *dynamic CSC cost* C_{Θ_t} as

$$C_{\Theta_t} = \|\mathbf{\Phi} - \mathbf{X}_{\Theta_t}\mathbf{X}_{\Theta_t}^\top \mathbf{\Phi}\|_F. \tag{26}$$

As the following theorem claims, the temporal evolution of the graph introduces an additional error term that is a function of the graph similarity (spectral- or edge- wise).

Theorem 4.1. *At time t, the dynamic CSC cost C_{Θ_t} and the SC cost C_{Φ_t} are related by*

$$C_{\Phi_t} \leq C_{\Theta_t} \leq C_{\Phi_t} + 2\sqrt{\frac{k}{d}}(\sqrt{k} + c) + (1+\delta)p\,\gamma, \tag{27}$$

with probability at least

$$1 - \exp(-c^2/2) - \exp\left(2\log(n) - dp(\frac{\delta^2}{4} - \frac{\delta^3}{6})\right),$$

where $0 < \delta \leq 1$. Above, γ depends only on the similarity of the graphs in question. Moreover, if graphs $\mathcal{G}_{t-1}$ and $\mathcal{G}_t$ are

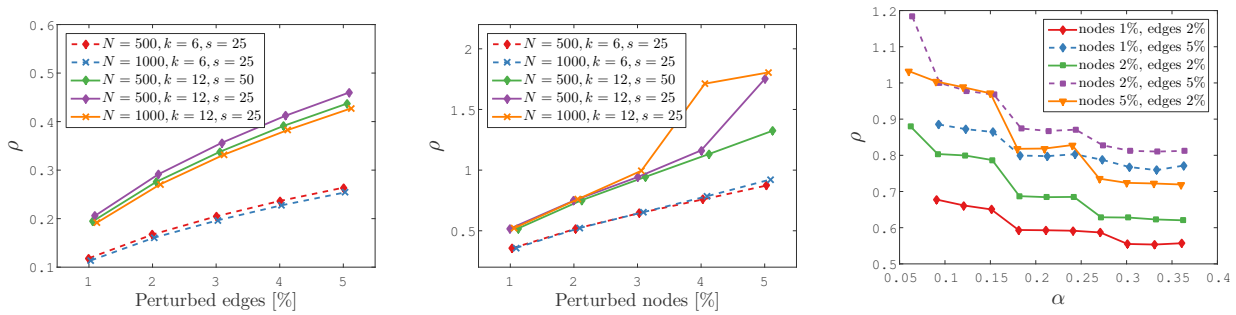
- *(ρ, k)-spectrally similar, then $\gamma = \rho$,*
- *ρ-edge similar, then $\gamma = (\sqrt{2}\,\rho)/\alpha$, where $\alpha = \min\{\lambda_k^t, \lambda_{k+1}^{(t-1)} - \lambda_k^t\}$ is the Laplacian eigen-gap.*

Proof. Let $\mathbf{X}_{\Phi_t}$ and $\mathbf{X}_{\Theta_t}$ be respectively the optimal SC and dynamic CSC clustering assignments at time t, and denote $\mathbf{E} = \mathbf{\Theta}_t - \mathbf{\Phi}_t \mathbf{I}_{k\times d}\mathbf{Q}$. We have that,

$$C_{\Theta_t} \leq C_{\mathbf{\Phi}} + 2\|\mathbf{\Theta}_t - \mathbf{\Phi}_t \mathbf{I}_{k\times d}\mathbf{Q}\|_F, \tag{28}$$

following the exact same steps as eq. (9).
By completing the matrices containing the filtering of both graphs, we can see that the error term can be rewritten as

$$\begin{aligned}\|\mathbf{E}\|_F &= \|\mathbf{\Psi}_{t-1}\mathbf{S}_{dp}^d + \mathbf{\Psi}_t\overline{\mathbf{S}_{dp}^d} - \mathbf{\Phi}_t\mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &= \|(\mathbf{\Psi}_{t-1} - \mathbf{\Psi}_t)\mathbf{S}_{dp}^d + \mathbf{\Psi}_t - \mathbf{\Phi}_t\mathbf{I}_{k\times d}\mathbf{Q}\|_F \\ &\leq \|(\mathbf{\Psi}_t - \mathbf{\Psi}_{t-1})\mathbf{S}_{dp}^d\|_F + \|\mathbf{\Psi}_t - \mathbf{\Phi}_t\mathbf{I}_{k\times d}\mathbf{Q}\|_F.\end{aligned} \tag{29}$$



(a) (b) (c)

Figure 1. Study of the different perturbation models and their impact on the graph spectral similarity. Graph possessing a large eigengap (highly clusterable) are less subject to perturbations. Proportionally, larger graphs are also less subject to perturbations allowing the number of edge modifications to be larger before a new clustering assignment is required for a given perturbation tolerance.

The rightmost term of eq. (29) corresponds to the effects of random filtering and has been studied in depth in Thm. 3.2 and Cor. 3.2. The rest of the proof is devoted to studying the leftmost term.

We apply the Johnson-Lindenstrauss lemma (Johnson & Lindenstrauss, 1984) on the term of interest. Setting $\mathbf{R}' = \frac{1}{\sqrt{p}}\mathbf{R}\mathbf{I}_{d\times dp}$, we have that

$$\begin{aligned}\|(\mathbf{\Psi}_t - \mathbf{\Psi}_{t-1})\mathbf{S}_{dp}^d\|_F^2 &= \|(\mathbf{H}_t - \mathbf{H}_{t-1})\mathbf{R}\mathbf{I}_{d\times dp}\|_F^2 \\ &= p\sum_{i=1}^{n}\|\mathbf{R}'^{\top}(\mathbf{H}_t - \mathbf{H}_{t-1})^{\top}\delta_i\|_2^2.\end{aligned}$$

Matrix $\mathbf{R}' = p^{-1/2}\mathbf{R}\mathbf{I}_{d\times dp}$ has $n \times dp$ Gaussian i.i.d. entries with zero-mean and variance $1/dp$. It follows from the Johnson-Lindenstrauss lemma that

$$\begin{aligned}\|(\mathbf{\Psi}_t - \mathbf{\Psi}_{t-1})\mathbf{S}_{dp}^d\|_F^2 &\leq p\,(1+\delta)\sum_{i=1}^{n}\|(\mathbf{H}_t - \mathbf{H}_{t-1})^{\top}\delta_i\|_2^2 \\ &\leq p\,(1+\delta)\|\mathbf{H}_t - \mathbf{H}_{t-1}\|_F^2,\end{aligned}$$

with probability at least $1 - n^{-\beta}$ and for $dp \geq \frac{4+2\beta}{\delta^2(\frac{1}{2}-\frac{\delta}{3})}\log(n)$. Coupling the two together we obtain a probability at least equal to $1-\exp(2\log(n)-\frac{dp\delta^2}{2}(\frac{1}{2}-\frac{\delta}{3}))$, where δ can be set between 0 and 1. A loose bound gives $2p\|\mathbf{H}^{(2)}-\mathbf{H}^{(1)}\|_F^2$ with probability $1-\exp(2\log(n)-\frac{dp}{12})$.

This concludes the part of the proof concerning spectrally similar graphs. The result for edge-wise similarity follows from Cor. 4.1. □

Corollary 4.1 (adapted from Cor. 4 (Hunter & Strohmer, 2010))**.** *Let $\mathbf{H}_{t-1}$ and $\mathbf{H}_t$ be the orthogonal projection on to the span of $[\mathbf{U}_k]_{t-1}(=\mathbf{\Phi}_{t-1})$ and $[\mathbf{U}_k]_t(=\mathbf{\Phi}_t)$. If there exists an $\alpha > 0$ such that $\alpha \leq \lambda_{k+1}^{(t-1)} - \lambda_k^t$ and $\alpha \leq \lambda_k^t$, then,*

$$\|\mathbf{H}_t - \mathbf{H}_{t-1}\|_F \leq \frac{\sqrt{2}}{\alpha}\|\mathbf{L}_t - \mathbf{L}_{t-1}\|_F. \qquad (30)$$

Note that the bounds on α are those described in their Thm. 3.

5. Experiments

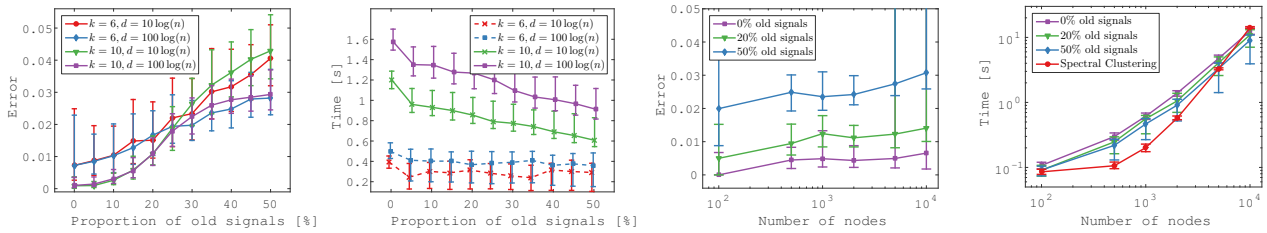
This section complements the theoretical results described in Section 4. First, we study the impact of graph modifications under different perturbation models to the ρ-spectral similarity. From there, we present the results of our dynamic clustering algorithm on graphs of different sizes and connectivity.

As is common practice (e.g., Görke et al., 2013; Tremblay et al., 2016) we apply our methods to Stochastic Block Models (SBM). This graph model simulates data clustered into k classes where the n nodes are connected at random with probability for each pair of nodes that depends if the two extremities are belonging to the same cluster (q_1) or not (q_2), with $q_1 \ll q_2$. In the following, we will qualify the SBM parameters in terms of the node's average degree s and the ratio $e = \frac{q_2}{q_1}$ that represents the graph clusterability.

All our experiments are designed using the GSPBox (Perraudin et al., 2014).

5.1. Spectral similarity

Our theoretical approach highlights the importance of the spectral similarity between two consecutive steps of the graph. We thus start this section by describing how much the graph can change between two assignments. Starting from a SBM, we perform two types of perturbations: edge redrawing and node reassignment. The former simply consists in removing some edges that existed at random and then adding the same number following the probabilities defined by the model. In the latter, one selects nodes instead, removes all edges that share at least one end with the



(a) (b) (c) (d)

Figure 2. Performances of our algorithm for dynamic graph clustering on synthetic data. Figures (a) and (b) presents the benefits of reusing previous features, while (c) and (d) focus on the scalability of the method with increasing number of nodes.

nodes previously picked, assigns those nodes to any other class at random and reconnects these nodes with new edges following the probabilities defined by the graph model.

Figure 1 shows the similarity of graphs under various perturbation models. Figures 1(a) and 1(b) illustrate the impact of the two aforementioned perturbation models separately on SBM of different sizes, whereas in Figure 1(c) the models are combined. We have three main observations.

First, the number of clusters k plays a major role in spectral similarilty ρ. This can be explained by the fact that ρ is bounded by $2\sqrt{k}$. This means that if a similarity threshold is set, one can afford more modifications in the graph when looking for fewer clusters. Second, we observe that graphs with a larger eigengap remains more similar to the original graph under a given perturbation models, confirming Cor. 4.1. Finally, it might be also interesting to notice that ρ increases with n. This suggests that the algorithm's ability to save computation by reusing information is enhanced for larger graphs.

5.2. Dynamic clustering of SBM

We proceed to study the efficiency of dynamic CSC. Based on our previous observations, we set the perturbation model as a combination of the two described in the previous subsection where 1% of the nodes are relabeled and 3% of the edges change. All the results presented here are statistics obtained from simulations replicated 200 times.

Figure 2 displays the results of our clustering for different proportions of previous signals reused in terms of two important metrics: time and accuracy. The error displayed on the figures is the multiplicative error of the k-means cost defined in Thm. 4.1, namely $\frac{C_{\Theta_t} - C_{\Phi_t}}{C_{\Phi_t}}$. Since this quantity requires the computation of SC, we are forced to consider problems where n stays in the order of thousands due to its important complexity. While 2(a) and 2(b) illustrate the benefits of reusing large parts of the previously computed features on graphs with $n = 1000, s = 25, e = \frac{1}{6}$, 2(c) and 2(d) sketch the intuition on large problems with varying values of n, setting $k = 2\log(n), d = 30\log(n)$ while keeping $s = 25, e = \frac{1}{6}$.

First, it is important to notice that as n increases, the time required to perform clustering using CSC methods (including dynamic CSC) outperform that of using SC. Second, as it could be expected, the error that we observe is slightly increasing as p grows, up to 3% of the SC cost when reusing 50% of the previously computed signals. This is very encouraging since in practice, such proportional error is not significant. Finally, we observe a computational benefit by looking at the time gained by more use of the previous features, as shown in Fig. 2(b). We emphasize that the improvement in terms of time can attain 25% of the total time in the most extreme cases depicted in this figure.

6. Conclusion and Future Work

The major contribution of this paper is the presentation of a fast clustering algorithm for dynamic graphs that achieves similar quality than Spectral Clustering. We proved theoretically how much the graph can change before losing information for a given computational budget.

We highlighted in this paper several open directions of research for the future. First, it appears clearly in the experiments that the majority of the remaining complexity lies in two steps: the partial k-means and the determination of λ_k. The former is the heart of Spectral Clustering and thus challenging to avoid but seems legitimate to address since there might be various ways to obtain a sub-assignment for some nodes in the graph. The latter, on the opposite, has been already researched in the past although the current methods remain approximated and hamper the results of the filterings.

References

Boutsidis, Christos, Gittens, Alex, and Kambadur, Prabhanjan. Spectral clustering via the power method-

provably. In *Proceedings of the 24th International Conference on Machine Learning (ICML)*, 2015.

Chakrabarti, Deepayan, Kumar, Ravi, and Tomkins, Andrew. Evolutionary clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 554–560. ACM, 2006.

Chi, Yun, Song, Xiaodan, Zhou, Dengyong, Hino, Koji, and Tseng, Belle L. Evolutionary spectral clustering by incorporating temporal smoothness. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 153–162. ACM, 2007.

Cohen, Michael B, Elder, Sam, Musco, Cameron, Musco, Christopher, and Persu, Madalina. Dimensionality reduction for k-means clustering and low rank approximation. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pp. 163–172. ACM, 2015.

Di Napoli, Edoardo, Polizzi, Eric, and Saad, Yousef. Efficient estimation of eigenvalue counts in an interval. *Numerical Linear Algebra with Applications*, 2016.

Fortunato, Santo. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

Fowlkes, Charless, Belongie, Serge, Chung, Fan, and Malik, Jitendra. Spectral grouping using the nystrom method. *IEEE transactions on pattern analysis and machine intelligence*, 26(2):214–225, 2004.

Gauvin, Laetitia, Panisson, André, and Cattuto, Ciro. Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. *PloS one*, 9(1):e86028, 2014.

Gittens, Alex, Kambadur, Prabhanjan, and Boutsidis, Christos. Approximate spectral clustering via randomized sketching. *Ebay/IBM Research Technical Report*, 2013.

Görke, Robert, Maillard, Pascal, Schumm, Andrea, Staudt, Christian, and Wagner, Dorothea. Dynamic graph clustering combining modularity and smoothness. *Journal of Experimental Algorithmics (JEA)*, 18:1–5, 2013.

Hammond, David K, Vandergheynst, Pierre, and Gribonval, Rémi. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2): 129–150, 2011.

Hunter, Blake and Strohmer, Thomas. Performance analysis of spectral clustering on compressed, incomplete and inaccurate measurements. *arXiv preprint arXiv:1011.0997*, 2010.

Johnson, William B and Lindenstrauss, Joram. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

Laurent, Beatrice and Massart, Pascal. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pp. 1302–1338, 2000.

Li, Mu, Lian, Xiao-Chen, Kwok, James T, and Lu, Bao-Liang. Time and space efficient spectral clustering via column sampling. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 2297–2304. IEEE, 2011.

Ng, Andrew Y, Jordan, Michael I, Weiss, Yair, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

Ning, Huazhong, Xu, Wei, Chi, Yun, Gong, Yihong, and Huang, Thomas. Incremental spectral clustering with application to monitoring of evolving blog communities. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pp. 261–272. SIAM, 2007.

Paratte, Johan and Martin, Lionel. Fast eigenspace approximation using random signals. *arXiv preprint arXiv:1611.00938*, 2016.

Perraudin, Nathanaël, Paratte, Johan, Shuman, David, Martin, Lionel, Kalofolias, Vassilis, Vandergheynst, Pierre, and Hammond, David K. GSPBOX: A toolbox for signal processing on graphs. *ArXiv e-prints*, August 2014.

Puy, Gilles, Tremblay, Nicolas, Gribonval, Rémi, and Vandergheynst, Pierre. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis*, 2016.

Pydi, Muni Sreenivas and Dukkipati, Ambedkar. Spectral clustering via graph filtering: Consistency on the high-dimensional stochastic block model. *arXiv preprint arXiv:1702.03522*, 2017.

Ramasamy, Dinesh and Madhow, Upamanyu. Compressive spectral embedding: sidestepping the svd. In *Advances in Neural Information Processing Systems*, pp. 550–558, 2015.

Shi, Jianbo and Malik, Jitendra. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

Shuman, David I, Vandergheynst, Pierre, and Frossard, Pascal. Chebyshev polynomial approximation for distributed signal processing. In *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, pp. 1–8. IEEE, 2011a.

Shuman, David I, Vandergheynst, Pierre, and Frossard, Pascal. Distributed signal processing via chebyshev polynomial approximation. *arXiv preprint arXiv:1111.5239*, 2011b.

Tremblay, Nicolas Tremblay, Puy, Gilles, Gribonval, Rémi, and Vandergheynst, Pierre. Compressive Spectral Clustering. In *33rd International Conference on Machine Learning*, New York, United States, June 2016.

Tu, Kun, Ribeiro, Bruno, Swami, Ananthram, and Towsley, Don. Detecting cluster with temporal information in sparse dynamic graph. *arXiv preprint arXiv:1605.08074*, 2016.

Vershynin, Roman. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Von Luxburg, Ulrike. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.