On Subgraph Signal Processing

Feng Ji, Giacomo Kahn, and Wee Peng Tay, Senior Member, IEEE

Abstract

Graph signal processing allows us to analyze a graph signal by transforming it into the frequency domain. In general, the procedure requires one to know the signal value at every node. In real situations, it is possible that only partial observations are made, called subgraph signals. In this paper, we introduce a subgraph signal processing framework. It allows us to define Fourier transform and the notion of frequency domain with signals available on a subset of nodes. As a consequence, we are able to give meaningful frequency interpretation of subgraph signals and perform standard signal processing tasks.

Index Terms

Subgraph signal processing, semi shift invariant filters, graph Fourier transform

I. INTRODUCTION

Since its emergence, the theory and applications of graph signal processing (GSP) have rapidly developed. GSP incorporates geometric properties of a graph in analyzing signals supported on it. The theory covers a wide range of topics with a full array of applications, including graph signal filtering, downsampling, clustering, detection and learning with graph neural networks [1]–[13].

The heart and soul of the theory is the notion of graph shift operator. Once an appropriate graph shift operator is chosen, there is a standard procedure to produce a signal processing framework (see for example [1] for more details). To ensure usefulness, such a shift operator should be associated with the topology of the graph, e.g., imitating a diffusion process on the graph; and in addition, enjoy a few prescribed algebraic properties, e.g., admitting a real eigenbasis. Common examples include the adjacency matrix and graph Laplacian. The decomposition of a signal w.r.t. an eigenbasis of the graph shift operator gives rise to the graph Fourier transform, and the coordinates w.r.t. this basis is the frequency domain, analogous to its counterpart in the classical theory of discrete Fourier transform (DFT). As a consequence, we are able to give a
different “interpretation” of a graph signal $x$ by looking at its new appearance $\hat{x}$ in the frequency domain.

However, in order to compute the frequency spectrum $\hat{x}$, one need to make use of every entry of $x$, i.e., the signal value at every node of the graph. However, there can be scenarios where full access of $x$ is impossible. For example, in a sensor network, it is possible that readings from some sensors are missing due to reasons such as processing delay [14]. As another example, if we view an image as a lattice graph and pixel values as signals, some of them can be missing due to corruption. As partial observation of $x$ lives on a subset of nodes, it is called a subgraph signal. In this paper, we want to propose a signal processing framework that allows one to make meaningful interpretation of subgraph signals, by finding the “right frequency domain”.

**Example 1.** Figure 1 shows a network of weather stations in the United States. The colored nodes (left) indicate the actual daily temperature readings. However, the readings for a large subset of stations are missing (as shown by the black nodes). With such kind of partial information, we still want to perform signal processing tasks. For example, suppose we introduce anomaly by raising the temperature at a single station $v$ by $30^\circ F$, where $v$ is highlighted by the red square (right). To spot this, it might be insufficient to look at the neighborhood of $v$, which contains only 2 other nodes in the subgraph with proper temperature readings. However, the subgraph signal processing framework proposed in the paper allows us to draw a more accurate conclusion based on a larger picture, including all the readings from stations in the northern part of US.

Further discussion and more details shall be provided in Section $V$.

![Fig. 1. Actual (left) and perturbed (right) temperature signals in a network of weather stations in the United States.](image-url)
The objective might seem to be similar to graph sampling [15]–[17]. However, graph sampling usually assumes prior knowledge on a full graph signal $x$ such as band-limitedness. With such knowledge, one selects an optimal sample of nodes of small size such that the restriction of $x$ at these samples captures most of even full information of $x$. On the other hand, in this paper, the subset of nodes is fixed and there is no specific prior information on $x$. The restriction of $x$ to the given subset does not carry full information of $x$, which prevents us from using GSP on the entire graph.

The rest of the paper is organized as follows. In Section II, we formally introduce the problem of subgraph signal processing on a given subset of nodes in a graph. As we recalled briefly above, the essential step is to find an operator $f$ as a shift on the given subset. We summarize some of the signal processing tasks one can perform once such an operator is found. To obtain $f$, we need to search within a family of operators, called semi shift invariant filters, as a generalization of several familiar filter families. The main theoretical properties are discussed in Section III. In Section IV we discuss how to cast the problem of finding $f$ as an optimization problem. We perform simulations in Section V and conclude in Section VI.

II. WHAT IS SUBGRAPH SIGNAL PROCESSING?

In this section, we describe a general subgraph signal processing framework cum problem formulation.

Let $G = (V, E)$ be a graph of size $|V| = n$ and $V_0 \subset V$ be a subset of nodes of size $m$. Graph signals on the entire graph $G$ can be canonically identified with $\mathbb{R}^n$, denoted by $S(V)$; while the signals on $V_0$ identified with $\mathbb{R}^m$, denoted by $S(V_0)$. There is an obvious restriction map $P_{V_0} : S(V) \rightarrow S(V_0)$ of a signal on $V$ to $V_0$, which is nothing but the projection of the coordinates.

**Definition 1.** Suppose $X_V$ and $X_{V_0}$ are fixed sets of linear transformations on $S(V)$ and $S(V_0)$ respectively. Furthermore, assume that for each $f \in X_{V_0}$, there is an orthonormal basis of $S(V_0)$ consisting of eigenvectors of $f$. A transformation for subgraph signal processing w.r.t. $X_V$ and $X_{V_0}$ is a pair $(f_1, f_2)$ with $f_1 \in X_V$ and $f_2 \in X_{V_0}$ such that $P_{V_0} \circ f_1 = f_2 \circ P_{V_0}$ (as illustrated in Figure 2).

Intuitively, we want to let $f_2$ play the role of a shift operator on $S(V_0)$. On the other hand, we want it to be related to operators on $S(V)$, and this is achieved via the relation $P_{V_0} \circ f_1 = f_2 \circ P_{V_0}$. 
The definition is vacuous without specifying $X_V$ and $X_{V_0}$. This will take a large bulk of the work and the choice of $X_V$ and $X_{V_0}$ shall be described in details in the next two sections. For the rest of this section, we shall describe how to use such a pair $(f_1, f_2)$ to perform signal processing tasks. In addition, we shall also give some heuristic reasoning on how one should choose $X_V$ and $X_{V_0}$.

By our assumption on $X_{V_0}$, there is an orthonormal basis of $\{x_1, \ldots, x_m\}$ of eigenvectors of $f_2$. We may further assume that they are ordered (increasingly) according to the magnitude of the associated eigenvalues. We may use $\{x_1, \ldots, x_m\}$ as a “Fourier basis” to deal with signals on $V_0$. More specifically, we may perform standard signal processing tasks similar to GSP on $G$, as we briefly recall some of them below:

1. **Fourier transform**: for a signal $x \in S(V_0)$, its *Fourier transform* is given by
   \[
   \hat{x}(i) = \langle x, x_i \rangle, \ 1 \leq i \leq m.
   \]

   The inverse transformation is given by:
   \[
   x = \sum_{1 \leq i \leq m} \hat{x}(i)x_i.
   \]

2. **Bandlimit and bandpass filters**: suppose $B$ is a subset of $\{1, \ldots, m\}$. As we assume that $\{x_1, \ldots, x_m\}$ is ordered, a signal $x$ has *bandlimit* $B$ if $\hat{x}(i) = \langle x, x_i \rangle = 0$ for $i \notin B$. The bandpass filter associated with $B$ is given by $x \mapsto \sum_{i \in B} \hat{x}(i)x_i$. For denoising and data-compression, one may consider bandpass filters associated with $B$ consisting of small indices; while for anomaly detection, one may instead choose $B$ containing large indices.
(3) Downsampling: if a signal $x$ on $V_0$ is bandlimited with $B$ of small size, we can always have full knowledge of $x$ by looking at the signal values at a subset $V_1 \subset V_0$ of size $|B|$. This is called **downsampling**.

(4) Convolution: convolution is a generalization of bandpass filters. A convolution kernel is a signal $z \in S(V_0)$. The associated convolution filter $x \mapsto z * x$ is defined by requiring $\hat{z} * \hat{x} = \hat{z} \hat{x}$.

(5) Shift invariant filters: a filter $F$ is **shift invariant** w.r.t. $f_2$ if $F \circ f_2 = f_2 \circ F$. Suppose there is an underlying shift operator $f$ of an actual graph such that $f_2 \circ f = f \circ f_2$ and $f_2$ does not have repeated eigenvalues. Then any shift invariant filter w.r.t. $f$ is also shift invariant w.r.t. $f_2$. Though it is more desirable to know $f$, it is usually sufficient to know $f_2$ to learn a shift invariant filter. However, if $F$ is a polynomial in $f$ of degree $d$, it is in general not true that $F$ is polynomial in $f_2$ of the same degree.

We have an explicit condition on $X_{V_0}$ in Definition 1. Most naturally, it is satisfied if $X_{V_0}$ contains only symmetric matrices. If in addition we want $X_{V_0}$ to contain actual graph Laplacians, we may require a matrix in $X_{V_0}$ has any constant vector as a 0-eigenvector and non-positive off-diagonal entries.

The conditions we impose on $X_{V_0}$ is mainly algebraic to ensure we can perform signal processing tasks. The resulting space is of high dimension. However, $X_V$ should be geometrically defined by taking account the topology of the embedding of $V_0$ in $G$. Defined with more concrete information, we want $X_V$ to be of much smaller dimension.

More specifically, we want the transformation in $X_V$ to model shifts on $G$. However, it will be over simplified to consider just the usual shift operators such as adjacency or Laplacian matrices, as we now outline why this is so.

For each signal $x$ on $V_0$, a signal $y$ on $V$ is an extension of $x$ if $P_{V_0}(y) = x$. For each $x$, all its extensions can be identified with $\mathbb{R}^{n-m}$. Requiring $P_{V_0} \circ f_1 = f_2 \circ P_{V_0}$ amounts to $P_{V_0} \circ f_1(y_1) = P_{V_0} \circ f_1(y_2)$ for any two extensions of the same $x$. Therefore, if we write $f_1$ in the matrix form, the bottom left $(n-m) \times m$-block must be 0. Therefore, it will be too restrictive to require $f_1$ being a shift operator in GSP theory, or even symmetric.

Now we are in the position to give a preview of our approach to tackle the above issues. First of all, in Section 3 we shall introduce new families of filters, called semi shift invariant filters. These families are intermediates between the space of shift invariant filters and the space of all linear filters. Moreover, a semi shift invariant filter is defined by piecing together shifts
on different parts of the graph. Such a filter serves as a suitable candidate to be included in $X_V$, because $V_0$ may be in-homogeneously embedded in $V$.

In actual learning, it is usually too restrictive to require $P_{V_0} \circ f_1 = f_2 \circ P_{V_0}$ once we fix parametrized families $X_V$ and $X_{V_0}$. For example, if $X_V$ is parametrized by a small parameter family, it is unlikely that we can find a non-trivial pair $(f_1, f_2)$ with symmetric $f_2$ so that the identity holds exactly. Therefore instead, we shall seek $(f_1, f_2)$ that minimizes $|P_{V_0} \circ f_1 - f_2 \circ P_{V_0}|$.

At this point, a slightly different point of view might be worth taking into consideration, which allows transformation of the problem to a least square optimization. We may assume that there is an unknown underlying distribution $\Omega$ on pairs of signals $(x, y)$ on $V_0$ and $V$ respectively such that $P_{V_0}(y) = x$, i.e. $y$ is an extension of $x$. For example, if $x$ is a constant vector, we expect $y$ to have a high chance to be constant for a sensible $\Omega$. We would like to have $P_{V_0} \circ f_1(y) \approx f_2(x)$ on average.

**Example 2.** Let $G_n = (V, E)$ be the directed cycle graph with $n$ nodes, and $A_n$ be the adjacency matrix of $G_n$, whose eigenbasis $\{y_i, 1 \leq i \leq n\}$ gives the basis of the discrete Fourier transform. More specifically, $y_i = (e^{2\pi(i-1)/n})_{0 \leq j \leq n-1}$.

Suppose $m \neq 1$ is a proper divisor of $n$. Let $V_0$ be a subset of $m$ equally spaced nodes. The projection $\{x_i, 1 \leq i \leq n\}$ of $\{y_i, 1 \leq i \leq n\}$ has size $m$ and forms the basis of the discrete Fourier transform on the cycle graph of $m$ nodes. With these pairs $(x_i, y_i)$, we are able to find $f_1 = A_n^{n/m}$ that makes $P_{V_0} \circ f_1(y_i) = f_2(x_i)$ for each $1 \leq i \leq n$, where $f_2$ is a multiple of $A_m$. In fact, the identity of operators $P_{V_0} \circ f_1 = f_2 \circ P_{V_0}$ holds without referring to any specific test vectors, which is special in this very homogeneous setting. We shall formalize the argument once the relevant concepts are introduced in Example 4.

In this example, $y_i$ in samples $(x_i, y_i)$ are taken from an orthonormal eigenbasis of a shift operator on $G$. They represent signals of various degrees of smoothness. To learn $f_1$ and $f_2$, we may perform a relaxed $L^2$ optimization on samples $(x_i, y_i)$. We shall further explore these ideas in Section III and Section IV.

### III. Semi shift invariant filters

In this section, we will describe explicitly our choice of $X_V$. For $G = (V, E)$ of size $n$, let $L$ be a fixed shift operator, e.g., the graph Laplacian is the primary example.
Definition 2. Let \( V_0 \subset V \) and \( d \leq n - 1 \). A semi shift invariant filter supported on \( V \) of degree \( d \) is any composition \( P_{V_0} \circ P_d(L) \), here by abuse of notation, \( P_{V_0} : S(V) \rightarrow S(V) \) is the projection onto the components of \( V_0 \) and set the signal to be 0 elsewhere and \( P_d \) is a polynomial of degree \( d \).

If \( V_0 \) is not too large, then a semi shift invariant filter can be effectively computed as follows.

Lemma 1. For \( V_0 \subset V \) and \( d \geq 0 \), the \( d \) neighborhood \( B_d(V_0) \) of \( V_0 \) is the union of nodes at most \( d \) hop away from some nodes of \( V_0 \). Suppose \( L \) is the Laplacian of \( G \) and \( L_{V_0,d} \) is the Laplacian of the induced subgraph on \( B_d(V_0) \), extended by 0 to \( V \setminus V_0 \). Then a semi shift invariant filter \( F = P_{V_0} \circ P(L) \) supported on \( V_0 \) of degree \( d \) is also given by \( F = P_{V_0} \circ P(L_{V_0,d}) \).

Proof: Let \( x \) be any graph signal. As \( P \) is of degree \( d \), for each node \( v \in V \), the value of \( P(L)(x) \) depends only on the signal values at nodes at most \( d \) hops away from \( v \). The filter \( P(L) \) and \( P(L_{V_0,d}) \) are the same on \( V_0 \). As the signal value of \( F(x) \) is 0 outside \( V_0 \), we have the desired identity \( F = P_{V_0} \circ P(L_{V_0,d}) \).

This observation allows a fast computation of a semi shift invariant filter if the size of \( B_d(V_0) \) is small. We now extend the definition to a collection of subsets.

Definition 3. Let \( C = \{V_1, \ldots, V_k\} \) be a collection of subsets of nodes and \( D = (d_1, \ldots, d_k) \) be a tuple of non-negative integers each smaller than \( n \). The space of semi shift invariant filters \( X_{C,D} \) on \( C \) of degree type \( D \) is the span of semi shift invariant filters supported on \( V_i \) of degree \( d_i \), \( 1 \leq i \leq k \).

For special choices of \( C \) and \( D \), we recover familiar families of filters.

Example 3. (1) If \( C = V \) and \( D = \{n - 1\} \), then \( X_{C,D} \) is the space of all shift invariant filters in the usual sense.

(2) Suppose \( L \) does not have repeated eigenvalues, and no eigenvector of \( L \) has 0 entry. For each \( v \in V \) indexed by \( j \) and \( 0 \leq d \leq n - 1 \), we have \( \dim X_{\{v\},d} = d + 1 \).

To see this, the space \( \dim X_{\{v\},d} \) is spanned by \( Pr_{\{v\}} \circ L^i, 0 \leq i \leq d \). We have \( \dim X_{\{v\},d+1} - \dim X_{\{v\},d} \leq 1 \). Therefore, it suffices to prove the case for \( d = n - 1 \). Suppose

\[
\sum_{0 \leq i \leq n - 1} a_i Pr_{\{v\}} \circ L_i = 0.
\] (1)
We write the orthonormal decomposition of \( L = O\Lambda O^{-1} \). The diagonal entries of \( \Lambda \) are the distinct eigenvalues \( \lambda_1, \ldots, \lambda_n \). Unwrapping \((1)\) as a transformation in the frequency domain, we have

\[
\sum_{0 \leq i \leq n-1} a_i(O_j \lambda_1^i, \ldots, O_j \lambda_n^i) = 0.
\]

As the row vectors \((\lambda_1^i, \ldots, \lambda_n^i), 0 \leq i \leq n - 1\) are independent of each other and \(O_j, \ldots, O_{jn}\) are non-zero. The vectors \((O_j \lambda_1^i, \ldots, O_j \lambda_n^i), 0 \leq i \leq n - 1\) are independent of each other. Consequently, \(a_i = 0\) for \(0 \leq i \leq n - 1\).

(3) Suppose \( L \) does not have repeated eigenvalues, and no eigenvector of \( L \) has 0 entry. If \( C = \{\{v_i\}, v_i \in V\} \) and \( D = \{n-1, \ldots, n-1\} \), then \( X_{C,D} \) is the space of all linear filters on \( S(V) \). Indeed, the dimension of the space of all linear filters is \( n^2 \). For any polynomials \( P_1 \) and \( P_2 \), \( Pr_{\{v_1\}} \circ P_1(L) \) and \( Pr_{\{v_2\}} \circ P_2(L) \) are independent as they are supported on different nodes. It suffices to show that for each \( v \in V \), \( \dim X_{\{v\},n-1} = n \), but this follows from [2].

More generally, if \( D = \{l, \ldots, l\} \) for some \( l \leq n-1 \), then \( X_{C,D} \) is the space of node-variant graph filters up to degree \( l \) described in [18].

For our applications in the paper, we are also interested in other intermediate cases.

**Definition 4.** \( C = \{V_1, \ldots, V_k\} \) is called essential if \( V_i \setminus \cup_{1 \leq j \neq i \leq k} V_j \neq \emptyset \) for \( 1 \leq i \leq k \).

A collection \( C' = \{V'_1, \ldots, V'_l\} \) is said to be a refinement of \( C \) if the following holds:

1. \( \cup_{1 \leq i \leq k} V_i = \cup_{1 \leq j \leq l} V'_j \).
2. Each \( V'_j \) is contained in some \( V_i \).
3. If distinct \( V'_{j_1}, V'_{j_2} \) are in the same \( V_i \) for some \( i \), then \( V'_{j_1} \cap V'_{j_2} = \emptyset \).

We now summarize the main structural result on \( X_{C,D} \) for various collection of nodes \( C \) and tuples of degrees \( D \).

**Theorem 1.** (1) If \( V_1 \subset V_2 \subset V \) and \( d_1 \leq d_2 \leq n-1 \), then \( \dim X_{V_1,d_1} \leq \dim X_{V_2,d_2} \).

(2) Suppose \( L \) does not have repeated eigenvalues, and no eigenvector of \( L \) has 0 entry. If \( C \) is essential and \( D = (d_1, \ldots, d_k) \), then \( \dim X_{C,D} = \sum_{1 \leq i \leq k} d_i + k \).

(3) Suppose \( C' = \{V'_1, \ldots, V'_l\} \) is a refinement of \( C = \{V_1, \ldots, V_k\} \). If \( D = (d_1, \ldots, d_k) \) and \( D' = (d'_1, \ldots, d'_l) \) satisfy \( d_i \leq d'_j \) whenever \( V'_j \subset V_i \), then \( X_{C,D} \subset X_{C',D'} \).
Fig. 3. In (a), \( C = \{V_1, V_2\} \) (blue) and \( C' = \{V'_1, V'_2, V'_3\} \) (red). \( C' \) is not a refinement of \( C \) because \( V'_2 \) is contained in neither \( V_1 \) nor \( V_2 \). In (b), \( C = \{V_1, V_2\} \) (blue) and \( C' = \{V'_1, V'_2, V'_3, V'_4\} \) (red). Though Definition 1(2) is satisfied, \( C' \) is not a refinement of \( C \) because \( V'_1, V'_3 \) are both in \( V_1 \) but \( V'_1 \cap V'_3 \neq \emptyset \).

Conversely, suppose \( L \) does not have repeated eigenvalues, and no eigenvector of \( L \) has 0 entry. Assume further that \( C' \) is essential and \( \cup_{1 \leq j \leq k} V_j = V \). If \( D, D' \) are constant tuples for the same \( 0 \leq d < n \) and \( X_{C,D} \subset X_{C',D'} \), then \( C' \) is a refinement of \( C \).

Proof: (1) As \( V_1 \subset V_2 \) and \( d_1 \leq d_2 \), any \( F \in X_{V_1,d_1} \) is the projection \( Pr_{V_1} \circ F' \) of some \( F' \in X_{V_2,d_2} \) to \( V_1 \). Furthermore, if \( F'_1, \ldots, F'_k \in X_{V_2,d_2} \) are filters such that \( Pr_{V_1} \circ F'_1, \ldots, Pr_{V_1} \circ F'_k \) are independent in \( X_{V_1,d_1} \), then \( F'_1, \ldots, F'_k \) are independent in \( X_{V_2,d_2} \). Hence, a basis of \( X_{V_1,d_1} \) are the projection of independent filters in \( X_{V_2,d_2} \), whence \( \dim X_{V_1,d_1} \leq \dim X_{V_2,d_2} \).

(2) As \( C \) is essential, each \( V_i \) contains a node \( v_i \) outside every \( V_j, j \neq i \). By Example 3(2) and using Part (1) we have \( d_i + 1 \leq \dim X_{\{v_i\},d_i} \leq \dim X_{V_i,d_i} \leq d_i + 1 \). Therefore, we must have \( \dim X_{\{v_i\},d_i} = \dim X_{V_i,d_i} \), and the (surjective) projection \( Pr_{\{v_i\}} : X_{V_i,d_i} \to X_{\{v_i\},d_i} \) is an isomorphism.

For each \( i \), let \( C_i \) be obtained from \( C \) by removing \( V_i \) and \( D_i \) be obtained from \( D \) by removing the \( i \)-th component \( d_i \). If \( F \neq 0 \in X_{\{v_i\},d_i} \) and \( F' \neq 0 \in X_{C_i,D_i} \), then they are independent of each other. Indeed, if \( aF + bF' = 0 \), then applying \( Pr_{\{v_i\}} \), we have \( aPr_{\{v_i\}} \circ F = 0 \in X_{\{v_i\},d_i} \). As \( Pr_{\{v_i\}} \circ F \) is non-zero, we must have \( a = 0 \) and hence \( b = 0 \). Consequently, filters in \( X_{V_i,d_i} \) are all independent and \( \dim X_{C,D} = \sum_{1 \leq i \leq k} X_{V_i,d_i} = \sum_{1 \leq i \leq k} d_i + k \).

(3) If \( C' \) is a refinement of \( C \), then each \( V_i \) in \( C \) is a disjoint union of \( V_i = \bigcup_{j=1}^{k'} V'_{ij} \) with
$V_{ij} \in C'$, by the definition of refinement. Let $F = Pr_{V_i} \circ (\sum_{0 \leq j \leq d} a_j L^j) \in X_{V_i,d}$, then

$$F = \sum_{1 \leq j \leq k'} Pr_{V'_{ij}} \circ (\sum_{0 \leq j \leq d} a_j L^j) \in X_{C', d'}.$$  

For the converse, we verify each of the conditions in Definition 4.

Suppose $\cup_{1 \leq i \leq k} V_i$ is not contained in $\cup_{1 \leq j \leq l} V'_j$. Let $v \in (\cup_{1 \leq i \leq k} V_i) \setminus (\cup_{1 \leq j \leq l} V'_j)$, and $F \in X_{C,D}$ be a filter such that $Pr_{\{v\}} \circ F$ is non-trivial. Such an $F$ always exists as $\dim X_{\{v\}, d} \geq 1$ by Example 3(2). However, $F \notin X_{C', d'}$ as the projection of any filter of $X_{C', d'}$ to $v$ is trivial. This gives rise to a contradiction.

For the second condition, suppose without loss of generality that $V'_i$ is not contained in any single $V_i$. As we assume that $C'$ is essential, there is an $v$ contained only in $V'_i$. Let $V_i \in C$ contain $v$ and $F = Pr_{V_i} \circ (\sum_{0 \leq i \leq d} a_i L^i) \in X_{V_i,d}$ be a non-zero filter. If $F \in X_{C', d'}$, then by considering the projection to $v$, $F$ must have a summand $F_1 = Pr_{V'_j} \circ (\sum_{0 \leq i \leq d} a_i L^i)$. However, the projection of $F_1$ to $V'_1 \setminus V_i$ is non-zero. There must be some other $V'_j$ such that

(a) $V'_1 \cap V'_j \neq \emptyset$.

(b) $F$ has a non-zero summand $F_j \in X_{V'_j,d}$.

For such a $V'_{j}$, there is a $v_j$ contained exclusively in $V'_{j}$. However, $Pr_{\{v_j\}} \circ F_j \neq 0$ and hence $v_j \in V_i$. This implies that $F_j = Pr_{V'_j} \circ (\sum_{0 \leq i \leq d} a_i L^i)$ to ensure $F_j$ and $F$ have the same projection to $v_j$. In conclusion, for any $v' \in V'_1 \setminus V_i$, there is a positive integer $m$ such that $0 = Pr_{v'} \circ F = m Pr_{\{v\}} \circ (\sum_{0 \leq i \leq d} a_i L^i) \neq 0$, which is a contradiction.

For the last condition, suppose $v \in V'_{j_1} \cap V'_{j_2}$. Choose any non-zero filter

$$F \in X_{V_i,d} \subset X_{C,D} \subset X_{C', d'}.$$  

For any $V'_{j} \subset V_i$, there is a $v_j$ contained exclusively in $V'_{j}$. Therefore, $F$ has a summand $Pr_{V'_j} \circ F$, and $Pr_{\{v\}} \circ F$ is the same as $m Pr_{\{v\}} \circ F$, where $m$ is the number of $V'_j$ contained in $V_i$ that contains $v$. Hence, $m = 1$. On the other hand, by the choice of $v$, $m$ is at least 2, and we obtain a contradiction. We conclude that $C'$ is a refinement of $C$.

$X_{C,D}$ performs different degrees of shifts on different parts of the graph. This is exactly what we want when $V_0$ is in-homogeneously embedded in $G$. For suitably chosen $C$ and $D$ based on the geometry of $V$ and $V_0$, $X_{C,D}$ shall be our choice of $X_V$ in Definition 1.

To end this section, we briefly describe how to express a semi shift invariant filter in the frequency domain of the shift operator $L$. Let $x_1, \ldots, x_n$ be an orthonormal basis consisting of
The eigenvectors of $L$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. For $V_0 \subset V$ and $d$, suppose $F \in \mathcal{X}_{V_0,d}$ is a semi shift invariant filter associated with polynomial $P$. Then for a signal $f = \sum_{1 \leq i \leq n} a_i x_i$,

$$F(f) = \begin{cases} \sum_{1 \leq i \leq n} P(\lambda_i) a_i f_j, & \text{the } j\text{-th node belongs to } V_0 \\ 0, & \text{otherwise.} \end{cases}$$

(2)

IV. LEARNING THE TRANSFORMATION FOR SUBGRAPH SIGNAL PROCESSING

As earlier, suppose $G = (V, E)$ is a connected undirected graph of size $|V| = n$ and $V_0 \subset V$ be a subset of nodes of size $m$. Let $L$ be the Laplacian of $G$. As a motivation, we first describe a straightforward approach for subgraph signal processing, discuss its insufficiency and explain how we may proceed using the tools introduced in Section III.

Let $H_0 = (V_0, E_0)$ be the induced subgraph of $V_0$, i.e., $V_0$ is the vertex set of $H_0$ and $u, v \in V_0$ are connected by an edge in $E_0$ if and only if they are connected by an edge $E$. Let $L_{H_0}$ be its Laplacian. It is natural to consider $L_{H_0}$ and its polynomials for $\mathcal{X}_{V}$ as in Definition I.

To motivates the need of an improvement over this choice of $\mathcal{X}_{V}$, consider the simple random model of placing a node in $V_0$ independently with probability $p$. The expected size of $V_0$ is thus $pn$. An edge $e$ of $G$ remains in $H_0$ if and only if both ends are in $V_0$, and thus the probability of such an event is $q = p^2$. We want to argue heuristically that $H_0$ cannot have large component and hence highly disconnected if $p$ is small. For this, we compare the random vertex selection model with the Erdös-Renyi random model $G(q)$ that keeps an edge of $G$ with probability $q$, which is independent over all the edges. The resulting graph $G(q)$ thus has the same expected amount of edges as $H_0$. However, the edges in $H_0$ tend to cluster together as only edges not sharing a vertex are retained independent of each other. Therefore, if with high probability $G(q)$ only has small components, then so does $H_0$. On the other hand, it is known (for example, [19], [20]) for a lot of cases that if $q$ is small enough, then $G(q)$ and hence $H_0$ tends to have small components, and becomes highly disconnected (an example is shown in Figure 4). We shall formalize parts of the heuristic in Appendix A.

Suppose a graph signal $y$ on $G$ is only observed at $V_0$ as $x = P_{V_0}(y)$. If the induced subgraph $H_0$ does not capture enough topological properties of $G$, e.g., $H_0$ is highly disconnected with small components, then it could be erroneous to handle $x$ with a shift operator of $H_0$ as the signals tend to stuck in a small part of the graph. The framework introduced in Section III is used exactly to overcome this difficulty.
Fig. 4. Suppose \( G \) is the 5 × 5 lattice in (a). If we choose \( p = 1/2 \), an example of \( H_0 \) is shown in (b) with \( |V_0| = 13 \), highlighted in red. We see that the largest component of \( H_0 \) has 6 nodes and 6 connected components.

A. Choice of \( X_{V_0} \)

To complete the picture envisioned in Definition 1, we need to describe \( X_{V_0} \) and \( X_V \). In this subsection, we outline a few candidates of \( X_{V_0} \) with reasoning. For preparation, we define the following.

**Definition 5.** Let \( V_0 \subset V \) be a subset of nodes and \( H_0 = (V_0, E_0) \) be the induced subgraph of \( V_0 \) in \( G \). An extension \( H \) of \( H_0 \) is a weighted graph with weight function \( w(\cdot, \cdot) \) such that: for any pairs \( u, v \in V_0 \) connected by an edge in \( H_0 \) (and hence in \( G \)), \( w(u,v) \) is the same as its weight in \( H_0 \). We denote the set of all such \( H \) by \( E_{H_0} \).

Immediately from the definition, the set \( E_{H_0} \) of all \( H \) extending \( H_0 \) is parametrized by non-negative real numbers each associated with a pair \( u, v \in V_0 \) that are not connected by an edge in \( H_0 \). As a real manifold, its dimension is \( |V_0|(|V_0| - 1)/2 - E_0 \).

Now, we are able to list a few candidates of \( X_{V_0} \), ordered by set inclusion.

1. \( X_{V_0} = \{ \text{the Laplacian of a graph in } E_{H_0} \} \). Any \( f \in X_{V_0} \) captures partial geometric features of the embedding of \( V_0 \) in \( G \). However, the downside is that this space can be less flexible due to small degree of freedom if \( H_0 \) is already well connected.

2. \( X_{V_0} = \{ \text{the Laplacian of any graph of size } |V_0| \} \). There is no particular restriction \( f \in X_{V_0} \) other than requiring it to be geometric. Both the eigenvector and eigenvalues can be used fully for signal processing tasks.

3. \( X_{V_0} = \{ \text{symmetric matrices whose rows sum to 0} \} \). This space is large enough as it includes not only graph Laplacians but also their shift invariant families. Because of this, for each \( f \in X_{V_0} \), its eigenvalues might not be accurate enough for certain signal processing
tasks. We require the rows of such $f$ sum to 0, so that constant vectors belongs to the 0-eigenspace, i.e., constant vectors are the smoothest.

**B. Choice of $X_V$**

For $X_V$, we want to make full use of the geometry of the embedding of $V_0$ in $G$. For this, $X_V$ will be a semi shift invariant family with appropriate chosen $C_{V_0}$ and $D_{V_0}$.

Consider two distinct nodes $v_1, v_2 \in V_0$ being $d$-hops away from each other in $G$. In order to exchange signals between $v_1$ and $v_2$ in $G$, the $d$-th power of the Laplacian $L_G$ is needed. Therefore, locally at each $v \in V_0$, the filter $f_1$ in Figure 2 should take the form of a polynomial in $L_G$, whose degree depends on how far it is away from other nodes in $V_0$. In view of this, we introduce the following and an illustration is given in Figure 5.

**Definition 6.** Fix a small integer $\epsilon \geq 0$. Let $D$ be the diameter of $G$. Given a subset of nodes $V_0 \in V$, let $C_{V_0} = \{V_i, 1 \leq i \leq D\}$ be a collection of subsets of $V_0$ such that

1. $\bigcup_{1 \leq i \leq D} V_i = V_0$.

2. If $v \in V_0$ has $i$-hop (calculated in $G$) neighbors but does not have any $i-1$-hop neighbors in $V_0$, then $v$ and all of its $i$-hop neighbors are in $V_i$.

$V_i$ can be empty, and if it is non-empty, we associate it with degree $i + \epsilon$ in forming $D_{V_0}$.

![Fig. 5](image)

Fig. 5. Suppose $G$ is the $5 \times 9$ lattice graph and $V_0$ consists of the red nodes. The collection $C_{V_0} = V_1 \cup V_2$, where $V_1$ consists of the 13 nodes, including $v_1, v_2, v_3$ and all the red nodes on their left. $V_2$ has 12 nodes, including $v_1, v_2, v_3$ and all the red nodes on their right. If $\epsilon = 1$, then $D_{V_0} = (2, 3)$.

In view of the above discussion, $X_{C_{V_0}, D_{V_0}}$ will be our primary candidate for $X_V$. By Theorem 1, the dimension of $X_{C_{V_0}, D_{V_0}}$ is $O(D^2)$, independent of the choice of $V_0$. 

As a tip, in actual applications, we may make small modifications such as specifying certain polynomial coefficients (c.f. Definition 2) or take refinement (c.f. Definition 4) of $C_0$.

C. Transformation for subgraph signal processing

As we discussed in Section II, it is possible that there is no non-trivial solution to the identify $|f_2 \circ P_{V_0} - P_{V_0} \circ f_1|$ for $f_1 \in X_V, f_2 \in X_{V_0}$. We turn to an optimization instead, whose basis form is given as follows.

**Problem 1.**

\[
(f_1, f_2) = \arg\min_{f_1 \in X_V, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|,
\]

(3)

where the norm is the operator norm.

Additional regularization or conditions might be imposed to prevent any trivial solution. As an example, we will fix a coefficient of $f_1$ in Section V.

**Lemma 2.** (1) Suppose $C' = \{V'_1, \ldots, V'_i\}$ is a refinement of $C$. If $D = \{d_1, \ldots, d_k\}$ and $D' = \{d'_1, \ldots, d'_l\}$ satisfy $d_i \leq d'_j$ provided $V'_j \subset V_i$, then

\[
\min_{f_1 \in X_{C', D'}, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1| \leq \min_{f_1 \in X_{C, D'}, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|.
\]

(2) For $C = \{V_0\}$ and $D = d \leq n - 1$, let $P$ be the polynomial associated with $f_1$ in solving Problem 7. Denote the eigenvalues and eigenvectors of $L_G$ by $\{\lambda_i\}, \{x_i\}, 1 \leq i \leq n$ and the eigenvalues of $f_2$ by $\{\mu_j\}, 1 \leq j \leq m$. Then

\[
\max_{1 \leq i \leq n} \min_{1 \leq j \leq m} |\mu_j - P(\lambda_i)||P_{V_0}(x_i)| \leq \min_{f_1 \in X_{C', D'}, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|.
\]

**Proof:** (1) Under the given conditions, we have $X_{C', D'} \subset X_{C, D}$ by Theorem 1. Consequently, we have the inequality

\[
\min_{f_1 \in X_{C', D'}, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1| \leq \min_{f_1 \in X_{C, D'}, f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|.
\]

(2) Let $\{y_j\}, 1 \leq j \leq m$ be the eigenvectors of $f_2$. For each $x_i$, write $P_{V_0}(x_i) = \sum_{1 \leq i \leq m} a_{ij}y_j$ for some coefficients $a_{ij}$. Therefore, $f_2 \circ P_{V_0}(x_i) = \sum_{1 \leq i \leq m} a_{ij}\mu_jy_j$. On the other hand,

\[
P_{V_0} \circ f_1(x_i) = P_{V_0}(P(\lambda_i)x_i) = \sum_{1 \leq j \leq m} a_{ij}P(\lambda_i)y_j.
\]
Hence,

\[ |(f_2 \circ P_{V_0} - P_{V_0} \circ f_1)(x_i)|^2 = \sum_{1 \leq j \leq m} a_{ij}(\mu_j - P(\lambda_i))y_j|^2 \]

\[ = \sum_{1 \leq j \leq m} (a_{ij}(\mu_j - P(\lambda_i)))^2 \]

\[ \geq (\min_{1 \leq j \leq m} (\mu_j - P(\lambda_i)))^2 \sum_{1 \leq j \leq m} a_{ij}^2 \]

\[ = (\min_{1 \leq j \leq m} (\mu_j - P(\lambda_i))|P_{V_0}(x_i)|)^2. \]

The results follows from

\[ \max_{1 \leq i \leq n} |(f_2 \circ P_{V_0} - P_{V_0} \circ f_1)(x_i)|^2 \leq \min_{f_1 \in X_{C^1}, \ f_2 \in X_{V_0}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|^2. \]  

Proof: The proof and related discussions will be given in Appendix B.

Part [2] of the above lemma describes how close the two sets of values \(\mu_j, 1 \leq j \leq m\) and \(P(\lambda_i), 1 \leq i \leq n\), in terms of the norm of the operator \(f_2 \circ P_{V_0} - P_{V_0} \circ f_1\). We shall illustrate with examples later on in Section IV.

The parameter space of \(X_V\) is usually small, as they are restricted space of polynomial coefficients. However, that of \(X_{V_0}\) can be large. We assume that \(X_{V_0}\) is one of the candidate spaces in Section IV-A, and the number of parameters is usually of order \(O(|V_0|^2)\). At a small cost of error difference, it is possible to reduce the size of parameter space.

**Proposition 1.** Given \(V_0\) and \(H_0\), assume that \(\beta_{V_0}\) is the minimal operator norm by solving Problem 2 with \(f_1 \in X_V, f_2 \in X_{V_0}\). Let \(r_{\text{max}}\) be the largest \(L^1\)-norm of the rows of \(f_2\). For any \(k > 0\), \(X_{V_0,k}\) is the subspace of \(X_{V_0}\) with at most \(k\) parameters, associated to undermined edge weights, nonzero.

For any \(\epsilon > 0\), there is a number \(N\) of order \(O(|V_0|/(\epsilon^2))\) such that the value:

\[ \beta_{V_0,N} = \min_{f_1 \in X_V, f_2 \in X_{V_0,N}} |f_2 \circ P_{V_0} - P_{V_0} \circ f_1|, \]

satisfies \(\sqrt{\beta_{V_0}} \leq \sqrt{\beta_{V_0,N}} \leq \sqrt{\beta_{V_0}} + \epsilon r_{\text{max}}.\)

Proof: The proof and related discussions will be given in Appendix B.
D. The least square problem

Recall in Section II, we assume that there is an underlying distribution \( \Omega \) on pairs of signals \((x, y)\) such that \( P_{\Omega}(y) = x \). Instead of solving Problem 1, we want to consider

\[
(f_1, f_2) = \arg \min_{f_1 \in X, f_2 \in X_0} E_{(x, y) \in \Omega}(|f_2(x) - P_{\Omega} \circ f_1(y)|^2).
\]

However, the distribution \( \Omega \) is hypothetical and unknown to us. Therefore, we want to find a reasonable set \( \Omega_{V_0} \) of explicit pairs \((x, y)\), treated as samples drawn from \( \Omega \). For this, we perform the following two steps.

S1. Take all the pairs \( \{(x_i, y_i), 1 \leq i \leq n\} \) where \( \{y_i, 1 \leq i \leq n\} \) forms an eigenbasis of \( L_G \) and \( x_i = P_{\Omega}(y_i) \). The pairs are ordered according to the size of the eigenvalues associated to \( y_i \).

These \( y_i \) represents well signals on \( G \), as they are of different degree of smoothness.

S2. Let \( 0 \leq \delta \leq 1 \) be a fixed number. We start by including the pair of constant vectors \((x_1, y_1)\) in \( \Omega_{V_0} \). Inductively, for \((x_i, y_i)\), we include it in \( \Omega_{V_0} \) if the cosine distance between \( x_i \) and \( x \) for each \((x, y) \in \Omega_{V_0}\) is at most \( \delta \).

The consideration here is that for pairs \((x_i, y_i)\) and \((x_j, y_j)\), \( i \leq j \), if \( x_i \) and \( x_j \) are close enough in cosine distance, then we prefer excluding \((x_j, y_j)\) from \( \Omega_{V_0} \). This is because a sensible \( \Omega \) should favor \( x \) and \( y \) of similar level of smoothness, for example, if \( x \) is constant, it is more likely that \( y \) is also constant.

Once we have chosen \( \Omega_{V_0} \), we can solve the following least square problem.

**Problem 2.**

\[
(f_1, f_2) = \arg \min_{f_1 \in X, f_2 \in X_0} \sum_{(x, y) \in \Omega_{V_0}} |f_2(x) - P_{\Omega} \circ f_1(y)|^2,
\]

where \( |\cdot|^2 \) is the Euclidean norm.

The optimal least square difference is denoted by \( \alpha_{V_0} \).

**Lemma 3.** Problem 2 is convex if the polynomial coefficients in \( X_V \) is either unconstrained or fixed.

**Proof:** Recall that the parameters of \( X_{V_0} \) are matrix entries. Being symmetric, rows sum to 0 and being non-negative are all convex conditions. For \( X_V \), the parameters are unconstrained
polynomial coefficients or constant. Thus the constraints on parameters give rise to a convex feasible set.

By (2), \( f_2(x) - P_{V_0} \circ f_1(y) \) is linear on all the parameters. Therefore, Problem 2 is a minimization of a convex function with convex feasible set, and hence the problem is convex. 

\[ \text{Example 4.} \]

(1) Let \( V_0 \) be a subset of \( V \) such that the induced subgraph \( H_0 \) is fully connected. Let \( \partial V_0 \) be the boundary of \( V_0 \), i.e., \( v \in \partial V_0 \) if some neighbor of \( v \) is not in \( V_0 \). By Definition 6 \( C_{V_0} = \{ V_0 \} \) and \( D_{V_0} = 1 \). We claim that if Problem 2 is solved with \( X_V = X_{C_{V_0}, D_{V_0}} \) and the coefficient of \( L_G \) in \( X_{V_0} \) is fixed to be 1, then we have \( \alpha_{V_0} \leq \lambda_n^2 |\partial V_0|^2 \), where \( \lambda_n \) is the largest eigenvalue of \( L_G \).

To see this, it suffices to consider the special choice \( f_1 = L_G, f_2 = L_{H_0} \), then

\[
\alpha_{V_0} \leq \sum_{1 \leq i \leq n} |f_2 \circ P_{V_0}(y_i) - P_{V_0} \circ f_1(y_i)|^2_2 \\
\leq \sum_{1 \leq i \leq n} \sum_{v \in \partial V_0} (L_G(y_i)|v|^2) = \sum_{v \in \partial V_0} \sum_{1 \leq i \leq n} (\lambda_i y_i|v|^2) \\
\leq \lambda_n^2 \sum_{v \in \partial V_0} \sum_{1 \leq i \leq n} y_i|v|^2 = \lambda_n^2 |\partial V_0|^2 ,
\]

where \( |v| \) means the signal value at \( v \).

The last equality follows from the fact that the row vectors \( (y_i|v)|1 \leq i \leq n \) are pairwise orthonormal. From this observation, we see that if \( |\partial V_0| \) is small, then using \( L_{H_0} \) already gives a reasonable choice of \( f_2 \).

(2) This is a continuation of Example 2. To fit with the classical DFT scenario, we change the setup slightly by working with adjacency matrix of directed graphs. We are now ready to fully formalize the rudimentary discussion made there. Following Section IV-B \( X_V = X_{C_{V_0}, D_{V_0}} \) with \( C_{V_0} = V_{n/m} = V_0 \) and \( D_{V_0} = n/m \). To avoid trivial solution, we require the leading coefficient of \( f_1 \in X_V \), as the projection of a polynomial of \( A_n \), to be 1. For \( X_{V_0} \), we let it consist of adjacency matrix of constant weight directed graphs. If we set \( \delta = 0 \), due to periodicity, we have exactly \( m \) pairs \((x_i, y_i), 1 \leq i \leq m \) for \( y_i \) with the lowest \( m \) frequencies. If we solve Problem 2 with such a setup, we recover exactly Example 2 with \( \alpha_{V_0} = 0 \).
V. SIMULATION RESULTS

In this section, we present four sets of experiments, including one investigatory experiment and three applications. For all the experiments, regardless of graph and problem types, we use the following choice for $X_V$ and $X_{V_0}$:

- $X_V$: First, we form $C_{V_0}$ and $D_{V_0}$ with $\epsilon = 2$ as in Definition 6. Take $C$ to be the refinement of $C_{V_0}$ by including $U_i = V_i \setminus V_{i+1}, i \geq 1$. The degree associated with $U_i$ is the same as that of $V_i$. Furthermore, we enforce the coefficient of $P_{V_i} \circ L_G$ is fixed, say to be 1.
- $X_{V_0}$: We use Choice (3) of Section IV-A.

The graphs being used in the experiments include synthetic Girvanâ–Newman (GM) community graphs with 3 and 4 components of size 120 [21], square lattice graph, a real Enron Email graph with 500 nodes\(^1\) a part of the Arizona power plant of size 47 [22] and a weather station network in US of size 197\(^2\). The graph for the power plant network is based on physical connections of the devices representing the nodes of the graph; and the graph for the weather station network is constructed using the $k$-nearest neighbor algorithm based on actual geometric locations of the stations.

A. Spectral distribution of subgraph shift

In this section, we study the spectral distributions of $f_1$ and $f_2$ after solving Problem 2 with the prescribed $X_V$ and $X_{V_0}$. As $f_1$ models an (asymmetric) shift on $G$, we want to investigate how $f_2$ is so by comparing the eigenvalues of $f_2$ with that of $L_G$ in the sense of Lemma 2. We want to know whether we resolve the problem of $H_0$ of being highly disconnected. Moreover, by our choice of $X_{V_0}$, $f_2$ is $a$ priori not positive semi-definite, and we also want to see whether most of the eigenvalues are positive or not.

There is a subtlety here. In general, $f_1$ is not symmetric. Instead of considering the eigenvalues of $f_1$, which are complex numbers, we look at the behavior of $f_1$ at each $v \in V_0$. More specifically, as $f_1$ is semi shift invariant, locally its coefficients are $\lambda_{i,v}$, which is a polynomial evaluated at $\lambda_i$ (c.f. Section IV-A). We collect $\Lambda = \{\lambda_{i,v}, 1 \leq i \leq n\}$ common for the largest subset of $v \in V_0$, and call it the main spectral set of $f_1$, and the set of $v \in V_0$ such $\{\lambda_{i,v}, 1 \leq i \leq n\}$ the main component of $f_1$. Similarly, we can define the secondary spectral set of $f_1$, and so on.

\(^1\)https://snap.stanford.edu/data/email-Enron.html

\(^2\)http://www.ncdc.noaa.gov/data-access/land-based-station-data/station-metadata
For the simulations, we consider two types of GM community graphs, square lattice and the Email graph. The former two graphs have clear community structures, while the latter two do not. Moreover, the email graph is a small world graph, in contrast to the lattice graph. The graph statistics are summarized in Table I. In the table, “deg.” is the average degree of the graph $G$, "no. comp." is the average number of connected components of $H_0$, “size. main. comp.” is average size of the main component of $f_1$ measured as a percentage of $|V_0|$, and “size. $V_0$” is the average size of $V_0$.

For these graphs, we first randomly choose $V_0$, and solve Problem 2 for $f_1$ and $f_2$. We plot the eigenvalues of $f_2$ with the main spectral set of $f_1$ (Figure 6) and the secondary spectral set of $f_1$ (Figure 7).

From Figure 6, we notice that the eigenvalues of $f_2$ fit largely with the main spectral set of $f_1$, in particular, when for large indices. The main spectral set of $f_1$ are non-negative, and most of the eigenvalues of $f_2$ are non-negative as well. This indicates that it still enjoys some important features of a Laplacian. On the other hand, a small part of the eigenvalues of $f_2$ are negative. As a comparison, the secondary spectral set of $f_1$ are mostly negative. The interval containing the negative eigenvalues of $f_2$ is fully contained in the interval containing the secondary spectral set of $f_1$.

| Graphs $G$       | deg. | $p$ | no. comp. | size. main comp. | size. $V_0$ |
|------------------|------|-----|-----------|------------------|------------|
| 3-GM graph       | 5.6  | 0.4 | 7.7       | 89.5%            | 47.8       |
| 4-GM graph       | 5.2  | 0.4 | 8.3       | 88.6%            | 47.5       |
| Square lattice   | 3.7  | 0.4 | 18.9      | 82.9%            | 57.3       |
| Email graph      | 12.6 | 0.2 | 33.2      | 69.6%            | 98.7       |

TABLE I
GRAPH STATISTICS

B. Signal compression

In this experiment, we study signal compression with the following setting. For a graph $G = (V, E)$, we generate a smooth signal $y$ which is bandlimited w.r.t. the Laplacian $L_G$. The exact bandlimit is unknown and the nonzero Fourier coefficients $\hat{y}(i)$ are chosen uniformly random in the interval $[0, 1]$. Given $V_0 \subset V$ of size $m$, suppose we only observe $y$ at $V_0$, as $x$. We
first decompose $x$ w.r.t. an (ordered) eigenbasis $\{x_i, 1 \leq i \leq n\}$ of $f_2$ as $x = \sum_{1 \leq i \leq n} \hat{x}(i)x_i$ (c.f. Section II), and perform compression of $x$ by retaining only the first $r\%$ of Fourier coefficients $\hat{x}_i, 1 \leq i \leq r\%m$. Let the resulting signal be $x_c$. We want to investigate the performance of the compression by computing the compression error $|x - x_c|/|x|$.

For comparison, instead of using $f_2$, we consider the exact same compression scheme applied to $L_{H_0}$ and $K$, where $L_{H_0}$ is the Laplacian of the induced graph and $K$ is the Laplacian matrix of the Kron reduction procedure [23]. They will be used in subsequent experiments as well.

We perform simulations on GM community graphs (with 3 and 4 communities) and square lattice graph, with $|V_0| \approx 40\%|V|$ and $r\% = 40\%$. The average compression errors for $f_2, L_{H_0}$

Fig. 6. The plot of the eigenvalues of $f_2$ (blue) and the main spectral set of $f_1$ (orange).
and $K$ are summarized in Table II. From the results, we see that for each graph, the average compression error of $f_2$ is always the smallest. The results agree with our speculation that the smoothness of the eigenvectors of $f_2$ aligns with that of $L_G$.

C. Anomaly detection

We now consider the task of anomaly detection with the following setting. Suppose $y$ is a normal signal on $G$, which is smooth w.r.t. the topology of $G$. Again, we only observe the signals at $V_0$, denoted by $x$. We introduce anomaly to $x$ by randomly perturbing the value of $x$ at a single node, and the resulting signal is denoted by $x_a$. 

![Graph](image-url)
A classical signal processing approach for anomaly detection is to look at the high frequency components of the spectrum of $x_a$, decomposed with a suitably chosen graph shift operator.

More specifically, for $f_2$, let $\{x_i, 1 \leq i \leq m\}$ be an (ordered) eigenbasis. As in Section II, the Fourier coefficients are $\{\hat{x}(i), 1 \leq i \leq m\}$ with $\hat{x}(i) = \langle x, x_i \rangle$. We choose $0 < r < 1$ and let $m(x) = \max_{r_m \leq i \leq m}(|\hat{x}(i)|)$. With a fixed threshold $t > 1$, we declare that $x_a$ is abnormal if $m(x_a)/m(x) > t$. A sample is shown in Figure 8 on the problem described in Example I. The exact same procedure can be applied for $L_{H_0}$ or $K$, and we shall compare their performance for different signal perturbations.

![Graph](image)

**Fig. 8.** For the US weather station network, the blue curve is the (partial) spectral plot of the normal temperature reading on $V_0$, while the red curve is the plot for the perturbed reading. The large fluctuations in the high frequency regime help us identify the abnormal signal.
We run experiments on the power plant network $G_1$ and the US weather station network $G_2$. The signals on $G_1$ are random bandlimited signals with small frequencies, simulating sensor recordings of the devices such as temperature. On the other hand, the signals on $G_2$ are real daily temperature recorded over the year 2013\textsuperscript{3}. The experiment parameters are set as $r = 0.35$, $t = 1.1$, $|V_0| \approx 0.5|V|$ for $G_1$; and $r = 0.15$, $t = 1.02$, $|V_0| \approx 0.2|V|$ for $G_2$. The percentage of successful anomaly detections of various methods and signal perturbations are shown in Figure [9].

From the results, we see that the same scheme performs significantly better if we use $f_2$. However, the percentage of successful detection is low if the perturbation is small, e.g., 0.2 for $G_1$ and 10 for $G_2$. We investigate by plotting the $f_2$ Fourier coefficients of $x$ and $x_a$ (a typical example is shown in Figure [10]). We see that when the perturbation is small, the spectrum of $x$ and $x_a$ look alike. In this case, without knowing how $x_a$ is constructed, it is hardly possible to say it is abnormal.

D. Denoising

For the last task, we consider denoising. In contrast to anomaly detection in Section V-C, we add random Gaussian noise to the signal $x$ for every nodes of $V_0$ to form $x_b$. For denoising, we observe $x_b$ only and want to recover $x$ as far as possible.

\textsuperscript{3}ftp://ftp.ncdc.noaa.gov/pub/data/gsod
As briefly described in Section II, we use the approach by applying a convolution filter to $x_b$ in the frequency domain that scales down high frequency components. More specifically, we first choose $0 < r < 1$ and a scaling factor $0 \leq s < 1$. The recovered signal $x_c$ on $V_0$ is the unique signal whose $i$-th Fourier coefficient is $s\hat{x}_b(i)$ for $rm \leq i \leq m$ and $\hat{x}_b(i)$ for $i < rm$.

To evaluate the performance, we compute ratio between the errors of the recovered signal and noisy signal $err(x) = |x - x_c|/|x - x_b|$. The same procedure can be applied to $L_{H_0}$ and $K$.

We perform experiments on the Enron Email graph $G_3$ and the weather station network $G_2$. For $G_3$, we consider the synthetic signals simulating timestamps of information propagation on the graph under the SI model with a random source [24], [25]. For $G_2$, we still use daily temperature readings of the year 2013. The experiment parameters are set as $r = 0.2, s = 0.3, |V_0| \approx 0.2|V|$ for $G_3$ and $r = 0.4, s = 0.5, |V_0| \approx 0.2|V|$ for $G_2$. In Figure [11] we plot the average error ratios $err(x)$ against the average noise added to the signal measured by $nr(x) = |x - x_c|/|x|$.

From the plot, we see that using $f_2$ permits the best performance. However, we notice that when a small noise is introduced, the error ratio $err(x)$ can be larger than 1 even in the best case using $f_2$, meaning the proposed recovery increases the error. As in Section V-C we investigate by plotting the $f_2$ Fourier coefficients of $x$ and $x_b$ (typical examples are shown in Figure [12]). We see that with small noise, the entire spectrum of $x$ and $x_b$ have similar amplitude, therefore scaling down high frequency components may introduce additional error. However, with more noise, $x_c$ clearly has a large portion of spectrum with higher amplitude, which makes proposed
Fig. 11. Performance of denoising with $f_2, L_{H_0}, K$ as the graph shift operators.

method effective.

VI. CONCLUSION

In this paper, we propose a framework for subgraph signal processing of signals on a subset of nodes of a graph. The essential idea is to find an appropriate shift operator on the given subset. We test the performance of our approach on a few signal processing tasks with both synthetic and real graph and datasets. The results demonstrate the effectiveness of the proposed method. For future work, we want to explore further applications of our method and develop machine learning models analogous to graph neural network.

APPENDIX A

RANDOM SUBGRAPH MODELS COMPARED

Suppose $G = (V, E)$ is a graph of size $n$. In this appendix, we shall compare the random vertex selection model and the Erdős-Renyi (ER) random edge model, as initiated at the beginning Section IV when we give motivation to the framework of the paper. We want to discuss when the ER model tends to produce larger components than the random vertex model, as the former is well-studied [19], [20].

Definition 7. Let $H_0$ be an induced subgraph on a subset of nodes $|V_0|$ and $k$ be a positive integer. Introduce $\delta(G, H_0, k)$ to be the smallest size of an edge set $E'$ such that $G \setminus E'$ does
not contain any connected subgraph of size at least \( k \), other than \( H_0 \). Taking maximum over all connected \( H_0 \) of size \( k \), we define

\[
\delta(G, k) = \max_{H_0: \text{connected of size } k} \delta(G, H_0, k).
\]

**Example 5.** Let \( G \) be the \( n \times n \) grid. Then for each \( k = O(n^\alpha) \), \( \alpha < 2 \), we claim that

\[
\delta(G, k) = O(\max(n^{3-\alpha}, n^\alpha)).
\]

To see this, let \( H_0 \) be a connected subgraph of order \( O(n^\alpha) \). Its boundary \( \partial H_0 \) (edges connected to the rest of the graph) is of size \( O(n^\alpha) \), as each node has degree at most 4. On the other hand, we can cut \( G \) into \( O(n^{2-\alpha}) \) vertical pieces such the size of each piece is smaller than \( n^\alpha \).
Moreover, the boundary between two adjacent vertical subgraphs has size \( n \). Hence the union \( E' \) of \( \partial H_0 \) and all the vertical boundaries has size \( O(\max(n^{3-\alpha}, n^\alpha)) \). From the construction, it is easy to see that the size each component of \( G \setminus E' \) is smaller than \( n^\alpha \) except for \( H_0 \), and the claim follows. An illustration is shown in Figure 13.

![Figure 13](image-url)

Fig. 13. In the illustration, the graph \( H_0 \) occupies the central red region and its boundary is depicted by the red curve. The boundaries of the vertical subgraphs in the argument are depicted by vertical blue lines.

**Definition 8.** Given \( V_0 \) such that the induced subgraph \( H_0 \) is connected, let \( \theta(G, H_0) \) be the number of distinct spanning trees of \( H_0 \). For any positive integer \( k \), define

\[
\theta(G, k) = \min_{|H_0|=k} \theta(G, H_0).
\]

The number \( \theta(G, k) \) can be thought as a complexity measure of induced subgraphs of \( G \) containing \( k \) vertices. As two extreme cases: for every \( k \), if \( G \) is a tree, then \( \theta(G, k) = 1 \); while on the other hand, if \( G \) is the complete graph, then \( \theta(G, k) = k^{k-2} \) by Cayley’s formula.

For \( 0 < q < 1 \), denote by \( G_q \) the model on induced subgraph of \( G \) that includes each node independently with probability \( q \). As in Section IV, let \( G(q) \) be the ER model that edges are preserved independently with probability \( q \). For a subgraph \( H \) of \( G \), let \( C(H) \) be the size of the largest component of \( H \). Now, we are ready to state and prove the main result of this appendix.

**Theorem 2.** For \( 0 < q_1, q_2 < 1 \) and positive integer \( k_1 \leq k_2 \), if

\[
(1 - (1 - q_1^{k_1-1})\theta(G, k_1))q_1^{\delta(G, k_1)} \geq q_2^{k_2},
\]
\[
\mathbb{P}(C(G(q_1)) = k_1) \geq \mathbb{P}(C(G_{q_2}) = k_2).
\]

**Proof:** Let \(A_1\) be the event that \(C(G(q_1)) = k_1\) and \(A_2\) be the event that \(C(G_{q_2}) = k_2\). To show \(\mathbb{P}(A_1) \geq \mathbb{P}(A_2)\), we introduce an “intermediate” event \(A_3\):

(a) the largest component \(C\) of \(G(q_1)\) has \(k_1\) nodes, and

(b) \(\delta(G, k)\) edges are not in \(G(q_1)\) such that the size of each other component in \(G(q_1)\) other than \(C\) is smaller than \(k_1\).

It is clear that \(\mathbb{P}(A_1) \geq \mathbb{P}(A_3)\), and we want to show \(\mathbb{P}(A_3) \geq \mathbb{P}(A_2)\).

For any \(k\), let \(C_k\) be the sets of \(k\) nodes each inducing a connected subgraph of \(G\). Therefore, we can decompose \(A_2\) as a union of events \(A_2 \subseteq \bigcup_{C \in C_k} \{C \subset G_{q_2}\}\). Hence by the union bound, we have the estimation:

\[
\mathbb{P}(A_2) \leq \mathbb{P}(\bigcup_{C \in C_k} \{C \subset G_{q_2}\})
\leq \sum_{C \in C_k} \mathbb{P}(C \subset G_{q_2})
= \sum_{C \in C_k} q_{k_2}^2.
\]

On the other hand, the event \(A_3\) contains the union of events:

\[
\bigcup_{C \in C_{k_1}} A_{3,C} := \bigcup_{C \in C_{k_1}} A_3 \cap \{C \subset G(q_1)\text{ and spans connected subgraph}\}.
\]

We observe the union is disjoint this is because for any event in \(A_3\), there can not be two connected components of size \(k_1\). In \(G\), there are at least \(\theta(G, k_1)\) distinct trees, which span each \(C \in C_{k_1}\). Therefore, the probability of each single instance in \(A_{3,C}\) is at least

\[
(1 - (1 - q_1^{k_1-1})^{\theta(G,k_1)} )q_1^{\delta(G,k_1)}.
\]

Moreover, as \(k_1 \leq k_2\), we have \(|C_{k_1}| \geq |C_{k_2}|\).

We now package everything together, if \((1 - (1 - q_1^{k_1-1})^{\theta(G,k_1)} )q_1^{\delta(G,k_1)} \geq q_2^{k_2}\), then

\[
\mathbb{P}(A_3) \geq \mathbb{P}(\bigcup_{C \in C_{k_1}} A_{3,C}) = \sum_{C \in C_{k_1}} \mathbb{P}(A_{3,C})
\geq \sum_{C \in C_{k_1}} (1 - (1 - q_1^{k_1-1})^{\theta(G,k_1)} )q_1^{\delta(G,k_1)}
\geq \sum_{C \in C_{k_2}} q_{k_2}^2 \geq \mathbb{P}(A_2).
\]
Example 6. As a continuation of Example 5, if \( k_i = O(n^{\alpha_i}) \) for \( \alpha_2 > \alpha_1 > 1.5 \), then \( \delta(G, k_1) = O(\max\{n^{3-\alpha_1}, n^{\alpha_1}\}) = O(n^{\alpha_1}) \) by Example 5. Suppose we use the trivial estimation \( \theta(G, k) \geq 1 \). Then for any \( 0 < q_1, q_2 < 1 \), the condition of Theorem 2 becomes \( q_1^{k_1-1}q_1^{\delta(G, k_1)} \geq q_2^{k_2} \), which is true as \( n \to \infty \). Therefore, asymptotically, we always have that \( G(q_1) \) has a larger chance to contain a component of size \( k_1 \) than that of \( G(q_2) \) to contain a component of size \( k_2 \).

Appendix B

Sparsification of \( X_{V_0} \)

We first give the proof of Proposition 1.

Proof: The left half of the inequality is clear, as in Proposition 1, \( f_2 \) is chosen from a smaller space.

For the right half, we want to make use results from the theory of spectral graph sparsification [26]. We shall prove the result for Choice (1) of \( X_{V_0} \) in Section [IV-A]. Decompose \( f_2 = L_{\tilde{H}} = L_{H_1} + L_{H_0} \), where \( L_{H_1} \) is the Laplacian of \( H_1 \), whose edges are those not in \( H_0 \). We may apply spectral matrix sparsification to \( H_1 \) yielding a graph \( H_2 \) with \( N = O(|V_0|/(\epsilon^2)) \) edges whose Laplacian \( L_{H_2} \) is an \( \epsilon \)-approximation of \( L_{H_1} \). This means \( (1-\epsilon)L_{H_1} \preceq L_{H_2} \preceq (1+\epsilon)L_{H_1} \), where \( A \preceq B \) means \( B-A \) is positive semi-definite. Let \( H_3 \) be the union of \( H_1 \cup H_0 \). In summary, (a) \( L_{H_3} \in X_{V_0, N} \), and (b) the sum \( L_{H_3} = L_{H_2} + L_{H_0} \) is an \( \epsilon \)-approximation of \( f_2 \), as \( L_{H_0} \) is clearly an \( \epsilon \)-approximation of itself.

We can now perform the following estimation:

\[
\sqrt{\beta_{V_0, N}} - \sqrt{\beta_{V_0}} \leq |L_{H_3} - L_{\tilde{H}}| \leq \epsilon|L_{\tilde{H}}| = \epsilon \mu_m,
\]

here \( \mu_m \) is largest eigenvalue of \( L_{\tilde{H}} \).

Let \( d_{\text{max}} \) be the maximum degree of \( \tilde{H} \). It suffices to show the inequality \( \mu_m \leq r_{\text{max}} = 2d_{\text{max}} \), which should be well-known. For completeness, we give a short self-contained proof. Consider the eigenvector \( y_m \) associated to \( \mu_m \). Let \( y_{m,k} \) be the entry with the largest absolutely value, associated to the \( k \)-th node. Then \( \mu_m y_{m,k} \) is the \( k \)-th entry of \( L_{\tilde{H}}y_m \). If \( d_k \) is the degree of node \( k \). The \( k \)-th entry of \( L_{\tilde{H}}y_m \) is bounded by \( d_k(2|y_{m,k}|) \) due to the maximality of \( |y_{m,k}| \). Therefore, \( \mu_m \leq 2d_k \leq 2d_{\text{max}} \), and the theorem follows.
For the error tolerance \( \epsilon_r^{\max} \), the proposition allows us to perform searching in a possibly smaller subset \( X_{V_0,N} \subset X_{V_0} \). However, it is intractable to search over \( X_{V_0,N} \) as there are exponential amount of ways to choose \( N \) nonzero parameters in \( X_{V_0} \).

A reasonable approach is to first determine the non-zero parameter set. Unfortunately, from the proof of Proposition 1, one needs to find \( f_2 \) first, after which, matrix sparsification is applied. To achieve sparsification with linear order, the latter step relies on an intricate analysis of \( \tilde{L}_f \) with the aid of barrier functions [26].

However, if we want to first apply matrix sparsification before solving the optimization problem, then a randomized edge selection heuristic is more suitable [27]. As a payoff, we can only hope to find \( f_2 \in X_{V_0} \) with \( O(|V_0| \ln |V_0|/\epsilon^2) \) parameters. For the rest of this section, we assume that \( X_{V_0} \) is either Choice (1) or (2) and each matrix from Choice (3) is the difference of two matrices from Choice (2).

For randomized edge selection, we need the following notion.

**Definition 9.** For a graph \( G \) with Laplacian \( L \), the effective resistance \( R_{u,v} \) between two nodes \( u, v \) on \( H \) is defined as:

\[
R_{u,v} = (\delta_u - \delta_v)'L^+(\delta_u - \delta_v),
\]

where \( \delta_u \) is the (column) vector taking value 1 at \( u \) and 0 elsewhere, and \( L^+ \) is the pseudo-inverse of \( L \) obtained by inverting non-zero eigenvalues of \( L \).

Intuitively, effective resistance \( R_{u,v} \) between two nodes \( u, v \) gives a good measure on the connectivity between them, and hence we may use \( R_{u,v} \) to determine whether their should be an edge between \( u, v \).

We briefly recall the steps for randomized edge sparsification for a graph of size \( m \):

1. For a pair of nodes \( u, v \), compute the effective resistance \( R_{u,v} \).
2. Let \( w(u, v) \) be the edge weight between \( u \) and \( v \). Define

\[
p_{u,v} = \min\{1, 4(\log m)w(u, v)R_{u,v}\epsilon^{-2}\}.
\]
3. An edge \( (u, v) \) is included in the new sparse graph with probability \( p_{u,v} \).

To solve the problem in Proposition 1 we are required to perform (a) edge selection and (b) optimization simultaneously, causing exponential complexity. We notice that the above randomized edge selection only requires input from the graph. Therefore, we may alternate between edge selection and graph estimation. The main steps can be described as follows:
(1) Set $G_0 = G$ and let $m = |V_0|$. Choose an integer parameter $N_1 = O(m \log m)$, whose exact size depends on error $\epsilon$. Let $N_2$ be a small fraction of $rN_1$, with $0 < r < 1$.

(2) Assume $L_{G_i} \in X_{V_0}$ is determined for a graph $G_i$. Add small random edge weights to every pair $u, v \in V_0$ not connected by an edge in $H_0$ to obtain a graph $G'_i$.

(3) For $(u, v) \notin H_0$, compare the effective resistance $R_{u,v}$ in $G'_i$. Rank such pairs in descending order according to $w(u, v)R_{u,v}$, denoted by $Q$.

(4) Select the top $N_1$ pairs in $Q$ and randomly choose $N_2$ pairs from the remaining pairs in $Q$. The union of these $N_1 + N_2$ pairs is denoted by $Q_i$.

(5) Construct the subspace $X_{V_0, Q_i} \subset X_{V_0}$ parametrized by nonzero variables associated with pairs in $Q_i$.

(6) $G_{i+1}$ is obtained by solving the optimization problem in Proposition 1 with $f_2 \in X_{V_0, Q_i}$.

The steps are repeated for a fixed amount of iterations or until converge. It is worth mentioning that the purpose of Step (2) and (4) is to ensure in each iteration, the nonzero parameter family $Q_i$ does not stuck in initial iterations.

REFERENCES

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

[2] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs,” IEEE Trans. Signal Process., vol. 61, no. 7, pp. 1644–1656, April 2013.

[3] ——, “Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure,” IEEE Signal Process. Mag., vol. 31, no. 5, pp. 80–90, Sept 2014.

[4] A. Gadde, A. Anis, and A. Ortega, “Active semi-supervised learning using sampling theory for graph signals,” in Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, USA, 2014, pp. 492–501.

[5] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, “Learning Laplacian matrix in smooth graph signal representations,” IEEE Trans. Signal Process., vol. 64, no. 23, pp. 6160–6173, Dec 2016.

[6] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in Advances in Neural Inform. Process. Syst., USA, 2016, pp. 3844–3852.

[7] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.

[8] H. E. Egilmez, E. Pavez, and A. Ortega, “Graph learning from data under Laplacian and structural constraints,” IEEE J. Sel. Top. Signal Process., vol. 11, no. 6, pp. 825–841, Sept 2017.

[9] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, “Network topology inference from non-stationary graph signals,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, March 2017, pp. 5870–5874.

[10] F. Grassi, A. Loukas, N. Perraudin, and B. Ricaud, “A time-vertex signal processing framework: Scalable processing and meaningful representations for time-series on graphs,” IEEE Trans. Signal Process., vol. 66, no. 3, pp. 817–829, Feb 2018.
[11] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, “Graph signal processing: Overview, challenges, and applications,” Proc. IEEE, vol. 106, no. 5, pp. 808–828, May 2018.

[12] B. Girault, A. Ortega, and S. S. Narayanan, “Irregularity-aware graph fourier transforms,” IEEE Transactions on Signal Processing, vol. 66, no. 21, pp. 5746–5761, Nov 2018.

[13] F. Ji and W. P. Tay, “A Hilbert space theory of generalized graph signal processing,” IEEE Trans. Signal Process., 2019, accepted.

[14] A. A. Khan and H. Agrawal, “Optimization of delay of data delivery in wireless sensor network using genetic algorithm,” in 2016 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC), 2016.

[15] A. Agaskar and Y. M. Lu, “A spectral graph uncertainty principle,” IEEE Trans. Inf. Theory, vol. 59, no. 7, pp. 4338–4356, 2013.

[16] S. Chen, R. Varma, A. Sandryhaila, and J. Kováčević, “Discrete signal processing on graphs: Sampling theory,” IEEE Trans. Signal Process., vol. 63, no. 24, pp. 6510–6523, 2015.

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

[18] S. Segarra, A. G. Marques, and A. Ribeiro, “Optimal graph-filter design and applications to distributed linear network operators,” IEEE Trans. Signal Process., vol. 65, no. 15, pp. 4117–4131, 2017.

[19] A. Frieze, M. Krivelevich, and R. Martin, “The emergence of a giant component in random subgraphs of pseudo-random graphs,” Random Struct. Algorithms, vol. 24, pp. 42–50, 2004.

[20] F. Chung, P. Horn, and L. Lu, “The giant component in a random subgraph of a given graph,” in Algorithms and Models for the Web-Graph, 2009.

[21] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.

[22] United States. Federal Energy Regulatory Commission North American Electric Reliability Corporation, “Arizona-southern california outages on september 8, 2011: Causes and recommendations,” 2012.

[23] F. Dorfler and F. Bullo, “Kron reduction of graphs with applications to electrical networks,” IEEE Trans. Circ. Syst. I, vol. 60, no. 1, pp. 150–163, 2013.

[24] D. Shah and T. Zaman, “Rumors in a network: Who’s the culprit?” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5163–5181, 2011.

[25] W. Luo, W. P. Tay, and M. Leng, “Identifying infection sources and regions in large networks,” IEEE Trans. Signal Process., vol. 61, no. 11, pp. 2850–2865, 2013.

[26] Y. T. Lee and H. Sun, “Constructing linear-sized spectral sparsification in almost-linear time,” in 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, 2015.

[27] D. A. Spielman and S.-H. Teng, “Spectral sparsification of graphs,” SIAM Journal on Computing, vol. 40, no. 4, pp. 981–1025, 2011.