Subgraph Detection using Graph Signals

Sundeep Prabhakar Chepuri and Geert Leus
Faculty of Electrical Engineering, Mathematics and Computer Science,
Delft University of Technology, The Netherlands.
Email: {s.p.chepuri;g.j.t.leus}@tudelft.nl.

***Abstract*—In this paper we develop statistical detection theory for graph signals. In particular, given two graphs, namely, a background graph that represents an usual activity and an alternative graph that represents some unusual activity, we are interested in answering the following question: To which of the two graphs does the observed graph signal fit the best? To begin with, we assume both the graphs are known, and derive an optimal Neyman-Pearson detector. Next, we derive a suboptimal detector for the case when the alternative graph is not known. The developed theory is illustrated with numerical experiments.**

***Index Terms*—Graph signal processing, subgraph detection, hypothesis testing, quadratic detector, locally most powerful test.**

I. Introduction

GRAPHS are mathematical objects that may be used to explain relationships among datasets. Vertices of the graph represent different elements of the dataset, whereas the edges of the graph explain the dependence between these elements. Some examples of such graph-structured data include transportation networks, gene networks, brain networks, social networks, to name a few. Processing signals supported on graphs is an emerging area of research [1], [2] that has recently received a lot of attention.

In this paper, we are interested in extending some of the classical tools used in statistical detection theory to graph signals. Detection theory for graph-structured data will be powerful for a wide range of applications within cyber security, traffic management, and network/data science applications, in general. For example, it is crucial for detecting and neutralizing malicious activity within a network.

Given two graphs, namely, a background graph that represents a typical activity and an alternative graph that represents some unusual activity, the focus in this paper is on answering the question: To which of the above two graphs does the observed graph data fit the best? To answer this question, we first model the graph signal as the output of a graph filter whose input is white noise. Thus, the intrinsic structure of the graph is incorporated in the observed signal. To begin with, we formulate a binary hypothesis testing problem assuming that both the graphs are perfectly known. The optimal Neyman-Pearson detector for this case is the well-known quadratic detector. Next, we derive a suboptimal detector for the case when the alternative graph is not known. The detector for this case is obtained using the locally most powerful test. The test statistic for both cases are simple in nature, and thus the hypothesis testing problem scales well for large datasets.

This work was supported by the KAUST-MIT-TUD consortium grant OSR-2015-Sensors-2700.

Given the graph signals, the problem of identifying the underlying topology [3] or learning a graph [4], [5] that explains the observed data is called *topology inference*. The proposed framework can be used to quantify the performance of two different topology inference algorithms that infer two different graphs.

Detection theory for graphs has been considered in the past [6], [7]. In [6], the focus is on detecting changes in random graphs, however, without taking the graph signals into account. That is, the graph itself forms the observations, and the discrimination is based on the properties of the adjacency matrices of the two graphs under test. The problem studied in [7] is similar to that of this paper. However, the detectors are developed under the assumption that the graph signal is smooth with respect to the supported graph. In contrast, the proposed framework doesn't impose any restrictions on the smoothness of the signal, however, we assume that the graph signal is stochastic in nature.

Notation: Upper (lower) bold face letters are used for matrices (column vectors). $(\cdot)^T$ denotes transposition. $\mathrm{diag}(\cdot)$ refers to a diagonal matrix with its argument on the main diagonal. $\mathbb{S}^N$ denotes the set of symmetric matrices of size $N \times N$. The (i,j)th entry of the matrix $\boldsymbol{A}$ is denoted as $[\boldsymbol{A}]_{i,j}$. The ℓ_0-(quasi) norm refers to the number of non-zero entries in $\boldsymbol{w}$, i.e., $\|\boldsymbol{w}\|_0 := |\{n : w_n \neq 0\}|$.

II. Preliminaries

Consider a dataset with N elements defined on the vertices of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Here, the vertex set $\mathcal{V} = \{v_1, \cdots, v_N\}$ denotes the set of nodes and the edge set $\mathcal{E}$ represents the relationship between the elements of the dataset. Therefore, such datasets are referred to as *signals on graphs*.

Let us introduce a symmetric matrix $\boldsymbol{S} \in \mathbb{S}^N$ that is related to the graph $\mathcal{G}$, where $[\boldsymbol{S}]_{i,j}$ can be nonzero only if $i = j$ or $(i,j) \in \mathcal{E}$. The sparsity pattern of $\boldsymbol{S}$ captures the local structure of the graph, and such a matrix is often referred to as the *graph-shift* operator [1], [8]. Possible candidates for $\boldsymbol{S}$ are the graph Laplacian $\boldsymbol{L}$, the adjacency matrix $\boldsymbol{A}$, or their variants.

The graph-shift operator $\boldsymbol{S}$ can be used to define *graph filters* of the form [1], [8]

$$\boldsymbol{H} = \sum_{l=0}^{L-1} h_l \boldsymbol{S}^l = \boldsymbol{U} \left(\sum_{l=0}^{L-1} h_l \boldsymbol{\Lambda}^l \right) \boldsymbol{U}^H, \tag{1}$$

where the filter $\boldsymbol{H}$ is of degree $L-1$ with filter coefficients $\boldsymbol{h} = [h_0, h_1, \ldots, h_{L-1}]^T \in \mathbb{R}^L$ and the diagonal matrix $\sum_{l=0}^{L-1} h_l \boldsymbol{\Lambda}^l = \mathrm{diag}(\boldsymbol{V}_L \boldsymbol{h})$ can be viewed as the frequency response of the graph filter. Here, $\boldsymbol{\Lambda}$ collects the graph frequencies on its diagonal, i.e., $\boldsymbol{\Lambda} = \mathrm{diag}([\lambda_1, \lambda_2, \cdots, \lambda_N]^T)$, and $\boldsymbol{V}_L$ is an $N \times L$ Vandermonde matrix with entries $[\boldsymbol{V}]_{i,j} = \lambda_i^{j-1}$. So when a graph signal, $\boldsymbol{w} \in \mathbb{R}^N$, is filtered using the graph filter $\boldsymbol{H}$, then the frequency content of the filter output, $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{w}$, is modified according to the frequency response of the graph filter.

III. Problem Statement

In this paper, the focus is on detecting topological changes in a graph using concepts from statistical detection theory. In particular, given the graph signal, the aim is to detect any changes to the known graph, e.g., the changes could be due to newly added edges between a subset of nodes. We pose this subgraph detection as a binary hypothesis testing problem. To do so, we first provide the following definitions.

Let us denote the *background graph* as $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0)$ with the graph-shift operator $\boldsymbol{S}_0 \in \mathbb{S}^N$. Let us denote an *alternative graph* as $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$. That is, the graphs $\mathcal{G}_0$ and $\mathcal{G}_1$ share the same vertex set, while the edge set $\mathcal{E}_1$ differs from $\mathcal{E}_0$. The graph $\mathcal{G}_1$ may be regarded as an *atypical graph*, with some edges added to the set $\mathcal{E}_0$ or removed from the set $\mathcal{E}_0$, representing some malicious network activity, for example. Therefore, we can express the graph-shift operator related to the graph $\mathcal{G}_1$ as $\boldsymbol{S}_0 + \boldsymbol{S}_1 \in \mathbb{S}^N$, where the (i,j)th entry of the matrix $\boldsymbol{S}_1$ denoted as $[\boldsymbol{S}_1]_{i,j}$ contains a non-zero entry only if $(i,j) \in \mathcal{E}_1$ and $(i,j) \notin \mathcal{E}_0$ or vice versa (i.e., at locations corresponding to the newly added or deleted edges).

Let us denote the graph signal as $\boldsymbol{x} \in \mathbb{R}^N$. We model the graph signal $\boldsymbol{x}$ as the output of a graph filter with zero-mean unit-variance white Gaussian noise as input, which is denoted as $\boldsymbol{w} \in \mathbb{R}^N$. In other words, in this work, we assume that the graph signal is stochastic in nature. The observations are related to the state of nature $\mathcal{H}$, where the random variable $\mathcal{H}$ is drawn from a binary alphabet set $\{\mathcal{H}_0, \mathcal{H}_1\}$. That is to say, the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the background and alternative graphs, respectively. We are interested in estimating the state of nature, and we denote the estimate of $\mathcal{H}$ as $\widehat{\mathcal{H}}$.

Suppose the graph signal follows the model

$$\mathcal{H}_0: \quad \boldsymbol{x} = \sum_{l=0}^{L-1} h_{0,l} \boldsymbol{S}_0^l \boldsymbol{w} = \boldsymbol{H}_0 \boldsymbol{w} \tag{2a}$$

$$\mathcal{H}_1: \quad \boldsymbol{x} = \sum_{l=0}^{L-1} h_{1,l} [\boldsymbol{S}_0 + \boldsymbol{S}_1]^l \boldsymbol{w} = (\boldsymbol{H}_0 + \boldsymbol{H}_1)\boldsymbol{w}, \tag{2b}$$

where

$$\boldsymbol{H}_1 = \sum_{l=0}^{L-1} h_{1,l} [\boldsymbol{S}_0 + \boldsymbol{S}_1]^l - \sum_{l=0}^{L-1} h_{0,l} \boldsymbol{S}_0^l.$$

Here, the graph filter coefficients $\{h_{i,l}\}_{l=0}^{L-1}$ for $i = 0, 1$ as well as the length of the filter L are assumed to be known or estimated. It might be also reasonable to assume that the graph filter coefficients are the same for both hypotheses (i.e., $h_{0,l} = h_{1,l}$, for $l = 1, 2, \ldots, L$), particularly, when the subgraph $\boldsymbol{S}_1$ is sparse. The graph filter is then estimated only for the background graph, which typically occurs.

Having described the observation model, we can now formally state the problem. In this paper, the questions of interest are, (i) given $\boldsymbol{S}_0$ and $\boldsymbol{S}_1$, i.e., given both graphs, decide on either $\mathcal{H}_0$ or $\mathcal{H}_1$ based on the observations $\boldsymbol{x}$ (this is the subject of Section IV); (ii) is it possible to solve the hypothesis testing problem if the alternative graph $\boldsymbol{S}_1$ is not known (this is the subject of Section V).

IV. Both the Graphs are Known

In this section, we will derive the detector that solves the hypothesis testing problem (2) when the graphs under both hypotheses are perfectly known. In other words, the aim is to determine: *to which of the two graphs does the graph signal fit the best?*

Suppose the graphs under both hypotheses are known perfectly. That is to say, the matrices $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$ are known. For this case, the binary hypothesis testing problem (2) can be expressed as

$$\mathcal{H}_0: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_0); \tag{3a}$$

$$\mathcal{H}_1: \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_1), \tag{3b}$$

where the $N \times N$ covariances matrices are given by

$$\boldsymbol{R}_0 = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = \boldsymbol{H}_0 \boldsymbol{H}_0^T$$

and

$$\boldsymbol{R}_1 = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = (\boldsymbol{H}_0 + \boldsymbol{H}_1)(\boldsymbol{H}_0 + \boldsymbol{H}_1)^T.$$

This is due to the assumption is that the seed signal $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

In what follows, we provide the Neyman-Pearson detector for the above problem, in which the probability of false alarm, denoted by

$$P_f = \Pr(\widehat{\mathcal{H}} = \mathcal{H}_1 | \mathcal{H}_0),$$

is fixed while the probability of detection, denoted by

$$P_d = \Pr(\widehat{\mathcal{H}} = \mathcal{H}_1 | \mathcal{H}_1),$$

is maximized. More specifically, the decision is based upon the log-likelihood ratio test

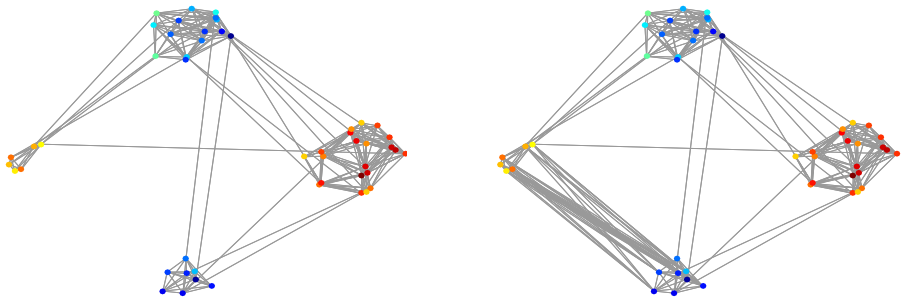
$$l(\boldsymbol{x}) = \log \frac{p(\boldsymbol{x} |\, \mathcal{H}_1)}{p(\boldsymbol{x} |\, \mathcal{H}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \lambda, \tag{4}$$

where the probability density function of $\boldsymbol{x}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ is denoted by $p(\boldsymbol{x} |\, \mathcal{H}_0)$ and $p(\boldsymbol{x} |\, \mathcal{H}_1)$, respectively. Here, λ is the threshold obtained by fixing the probability of false alarm.

The Neyman-Pearson detector for testing two known covariance matrices is the well-known quadratic detector with the log likelihood ratio test statistic [9]:

$$l(\boldsymbol{x}) = \boldsymbol{x}^T (\boldsymbol{R}_0^{-1} - \boldsymbol{R}_1^{-1}) \boldsymbol{x}, \tag{5}$$

where we decide on hypothesis $\mathcal{H}_1$ if $l(\boldsymbol{x}) > \lambda$, and λ is chosen such that a fixed false alarm rate is achieved. The threshold, however, has to be computed numerically, by inverting the χ^2 probability density function of the test statistic; see [9] for more details.



(a) Background graph. (b) Alternative graph.

Fig. 1: *Subgraph detection problem.* Community graph with four dominant communities and $N = 50$ vertices. (a) Background graph $\mathcal{G}_0$ does not have direct connections between all the communities. (b) Alternative graph $\mathcal{G}_1$ has direct connections between all the communities.

The quadratic detector is very simple to implement, and scales well for large datasets. A further reduction in the processing costs may be achieved by observing only a subset of graph vertices as discussed in the following remark.

Remark 1 (Graph sampling)**.** *In case of large-scale graphs, it is advantageous to observe only a subset of graph nodes that results in a desired detection performance. Suppose we model the graph sampling operation through a binary vector $\boldsymbol{z} \in \{0,1\}^N$, where the nth entry of $\boldsymbol{z}$ is denoted as z_n and $z_n = 0$ indicates that the nth graph node is not observed, and it is observed when $z_n = 1$.*

For the hypothesis testing problem (3)*, the sampling vector may be designed by optimizing the so-called J-divergence measure given by [10]*

$$\mathcal{D}(\boldsymbol{z}) = \frac{1}{2}\mathrm{tr}\{\boldsymbol{R}_0^{-1}(\boldsymbol{z})\boldsymbol{R}_1(\boldsymbol{z})\} + \frac{1}{2}\mathrm{tr}\{\boldsymbol{R}_1^{-1}(\boldsymbol{z})\boldsymbol{R}_0(\boldsymbol{z})\},$$

where for $i = 0, 1$ the notation $\boldsymbol{R}_i(\boldsymbol{z})$ denotes the submatrix of $\boldsymbol{R}_i$ that includes only the entries corresponding to the selected measurements. Maximizing $\mathcal{D}(\boldsymbol{z})$ subject to a cardinality constraint on $\boldsymbol{z} \in \{0,1\}^N$ as

$$\arg \max_{\boldsymbol{z}\in\{0,1\}^N} \quad \mathcal{D}(\boldsymbol{z}) \quad \text{s.to.} \quad \|\boldsymbol{z}\|_0 = K,$$

then results in the desired graph sampling scheme that subsamples K graph nodes. This optimization problem can be relaxed and solved using convex optimization techniques; see [10] for details.

Although the quadratic detector (5) is optimal in the Neyman-Pearson sense, we need to know the alternative graph $\mathcal{G}_1$. So if both graphs are known, then one should solve the quadratic detector. In what follows, we will seek suboptimal detectors when the alternative graph is not known.

V. Graph $\mathcal{G}_1$ is Not Known

Suppose the alternative graph is not known, then we have to solve a composite hypothesis testing problem, where the unknowns are replaced by their estimates to obtain the test statistic. Such tests are referred to as generalized likelihood ratio tests (GLRTs). In this case, to solve a GLRT we need to estimate $\boldsymbol{S}_1$ from the observations under $\mathcal{H}_1$. The shift-operator $\boldsymbol{S}_1$ can be estimated, e.g., using techniques proposed in [3]–[5], but this requires training data. Next, we will derive a one-sided parameter test for the subgraph detection problem that allows us to approximate the test statistic such that it depends only on the graph $\mathcal{G}_0$.

By defining

$$\boldsymbol{x} = (\boldsymbol{H}_0 + \mu\boldsymbol{H}_1)\boldsymbol{w},$$

the binary hypothesis testing problem (2) can be recast as a *scalar* one-sided parameter test as

$$\begin{aligned} \mathcal{H}_0 : &\quad \mu = 0; \\ \mathcal{H}_1 : &\quad \mu > 0. \end{aligned} \tag{6}$$

For such one-sided scalar parameter tests, the so-called *locally most powerful* (LMP) detector can be used. The LMP test constrains the P_f and minimizes the P_m for all μ close to zero with $\mu > 0$. For cases when μ is not around 0, the GLRT should also be tried. Nevertheless, $\boldsymbol{H}_1$ can always be constructed such that μ is around zero.

The probability density function under $\mathcal{H}_i$ for $i = 0, 1$ is parameterized by μ and is given by $p(\boldsymbol{x}, \mu)$. The test statistic for the LMP test is given by [11]

$$T_{\mathrm{LMP}}(\boldsymbol{x}) = \left.\frac{\partial \ln p(\boldsymbol{x};\mu)}{\partial \mu}\right|_{\mu=0} \tag{7}$$

That is, we decide on hypothesis $\mathcal{H}_1$ if $T_{\mathrm{LMP}}(\boldsymbol{x}) > \lambda'$, for some threshold λ' that leads to a fixed false alarm rate.

To derive the LMP test, we express the probability density function of the observations as

$$p(\boldsymbol{x};\mu) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}(\mu)|^{1/2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}(\mu)\boldsymbol{x}\right\},$$

where the covariance matrix depends on μ and is given by

$$\boldsymbol{C}(\mu) = \boldsymbol{H}_0\boldsymbol{H}_0^T + \mu^2\boldsymbol{H}_1\boldsymbol{H}_1^T + \mu\boldsymbol{H}_0\boldsymbol{H}_1^T + \mu\boldsymbol{H}_1\boldsymbol{H}_0^T.$$

It can be shown that the derivate of the log-likelihood function for $\mu = 0$ that depends on the data (and hence the test statistic) will be

$$T_{\mathrm{LMP}}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{R}_0^{-1}[\boldsymbol{H}_0\boldsymbol{H}_1^T + \boldsymbol{H}_1\boldsymbol{H}_0^T]\boldsymbol{R}_0^{-1}\boldsymbol{x}. \tag{8}$$

The test statistic is a weighted energy detector, and it has a similar quadratic form as the Neyman-Pearson detector derived in Section IV. As before, the threshold λ' has to be computed numerically, by inverting the χ^2 probability density function of the test statistic $T_{\mathrm{LMP}}(\boldsymbol{x})$.

The test statistic $T_{\mathrm{LMP}}(\boldsymbol{x})$ in (8) still depends on $\boldsymbol{H}_1$, but as a weight term $\boldsymbol{H}_0\boldsymbol{H}_1^T + \boldsymbol{H}_1\boldsymbol{H}_0^T$. Suppose we drop this weight term. Then we arrive at a suboptimal detector with the test statistic given by

$$T(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{R}_0^{-2}\boldsymbol{x}. \tag{9}$$

Letting $\boldsymbol{y} = \boldsymbol{R}_0^{-1}\boldsymbol{x}$, i.e., by pre-whitening $\boldsymbol{x}$, we obtain the simple quadratic test statistic

$$T(\boldsymbol{y}) = \boldsymbol{y}^T\boldsymbol{y},$$

which essentially tests how smooth the signal $\boldsymbol{y}$ is with respect to the background graph. However, there is no general answer to the question, how suboptimal is the detector (9) as compared to the LMP detector or quadratic detector, and this has to be verified through numerical experiments.

VI. Numerical Experiments

In this section we will provide numerical results that demonstrate the performance of the developed detectors. For our experiments, we use a community graph with $N = 50$ vertices and four dominant communities as shown in Figure 1.

We use the Laplacian matrix as the graph-shift operator. We model the graph signal by filtering zero-mean white Gaussian noise with a graph filter that has $L = 4$ coefficients and we use $h_{0,l} = h_{1,l}$, for $l = 1, 2, \ldots, L$. For the background graph $\mathcal{G}_0$, i.e., the graph under hypothesis $\mathcal{H}_0$, we use a community graph that does not have direct connections between all the communities; see Figure 1a. This graph may represent the usual activity between the nodes, for instance. For the alternative graph $\mathcal{G}_1$, i.e., the graph under hypothesis $\mathcal{H}_1$, we use a community graph that has direct connections between all the dominant communities; see Figure 1b.

The performance of the proposed detectors is illustrated via a receiver operating characteristics curve (i.e., a curve of P_d vs. P_f). This is computed numerically by varying the decision thresholds. The experiments are averaged over 2000 independent Monte-Carlo trials.

In Figure 2, we show the performance of the three proposed detectors. Recall that the quadratic detector is optimal in the Neyman-Pearson sense, and this detector assumes that both the graphs are perfectly known. This detector is the uniformly most powerful test and it provides an *upper bound* for the performance of any suboptimal detector (including the detector based on the generalized likelihood ratio test). We also show the performance of the locally most powerful detector in Figure 2, where we simply assume $\mu = 1$ to illustrate the worst-case performance. Using a smaller value for μ (i.e., scaling the entries of $\boldsymbol{H}_1$ accordingly) may improve the performance of the LMP detector. Finally, we show the performance of the suboptimal detector in Figure 2. Recall that the suboptimal detector is also based on the LMP test, however, using an approximate test statistic that doesn't depend on the alternative graph $\mathcal{G}_1$. Not knowing the alternative graph comes at a loss in performance as can be seen in Figure 2.

Software and datasets to produce results of this paper can be downloaded from http://cas.et.tudelft.nl/~sundeep/sw/asilomar16Gdet.zip

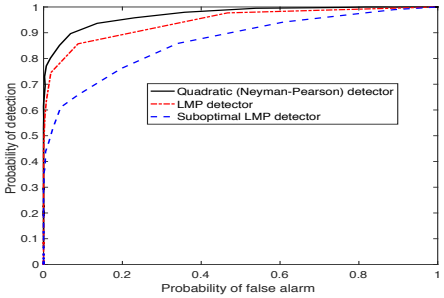


Fig. 2: Performance of the detectors.

VII. Conclusions

In this work have provided some initial results on statistical detection theory applied to graph-structured data. In particular, we have formulated a binary hypothesis testing problem to detect changes in the graph topology based on the observed graph signals. In other words, suppose we have two graphs, e.g., a background graph that represents some usual activity and an alternative graph that represents a malicious activity. Based on the observed graph signals, the developed detector decides whether the underlying graph was the background graph or the alternative graph. To this end, we have developed a Neyman-Pearson optimal detector. Further, we have also developed a suboptimal detector for the case when the alternative graph might not be known.

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, 2013.

[2] A. Sandryhaila and J. M. Moura, "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, 2014.

[3] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology identification from spectral templates," *arXiv preprint arXiv:1604.02610*, 2016.

[4] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *arXiv preprint arXiv:1406.7842*, 2014.

[5] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero III, "Learning sparse graphs under smoothness prior," *arXiv preprint arXiv:1609.03448*, 2016.

[6] B. A. Miller, N. T. Bliss, P. J. Wolfe, and M. S. Beard, "Detection theory for graphs," *Lincoln Laboratory Journal*, vol. 20, no. 1, pp. 10–30, 2013.

[7] C. Hu, J. Sepulcre, K. A. Johnson, G. E. Fakhri, Y. M. Lu, and Q. Li, "Matched signal detection on graphs: Theory and application to brain imaging data classification," *NeuroImage*, vol. 125, pp. 587–600, 2016.

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

[9] L. L. Scharf, *Statistical signal processing*. Reading, MA, USA: Addison-Wesley, 1991.

[10] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1446–1460, 2015.

[11] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*. Englewood Cliffs, NJ, USA: Prentice-Halll, 1998.