Identification of Outliers in Graph Signals*

Karthik Gopalakrishnan, Max Z. Li, and Hamsa Balakrishnan

***Abstract*— Outlier detection, or the identification of observations that differ significantly from the norm, is an important aspect of data mining. Conventional outlier detection tools have limited applicability to networks, in which there are interdependencies between the variables. In this paper, we consider the problem of identifying unusual spatial distributions of nodal signals on a graph. Leveraging tools from graph signal processing and statistical analysis, we propose a methodology to identify outliers in graph signals in a computationally efficient manner. Specifically, we examine a projection of the graph signal into a lower dimensional representation that enables easier outlier identification. Additionally, we derive analytical expressions for the outlier bounds. We apply our technique by identifying off-nominal days in the context of the US airport network using aviation delay data.**

I. Introduction

Outlier or anomaly detection – the identification of data points that differ in a significant manner from the majority of the observations – is an important problem in data analysis, for a number of reasons. When such data points are included while training models, the resulting models can be unrepresentative of the real system. Furthermore, several commonly-used algorithms (for example, linear or logistic regression, and AdaBoost) are particularly sensitive to outliers. Anomaly detection also plays an important role in system health monitoring and diagnosis. Outliers could correspond to valid observations resulting from unusual, off-nominal, or unexpected events in the system. In such cases, outliers represent interesting observations that merit further investigation. These situations motivate the need to identify outliers, and to provide interpretable rationales for them being classified as such. When the data corresponds to observations of a networked system, an observation may be an outlier not only because of its absolute value (for example, too large or too small), but also because it corresponds to an unusual distribution of values across the nodes of the network. The growing ubiquity of networked systems in science and engineering further motivates the development of outlier identification methods for signals on networks.

In this paper, we consider the problem of identifying outliers in data obtained from networked systems. We abstract the network interactions in the form of a graph with N vertices, and consider the data observations as signals supported on the vertices. Thus, each element of the observation vector (also referred to as a data point or a graph signal vector) is a scalar value associated with a graph vertex. The strength of the interdependencies between the signal values at pairs of nodes determines the edge weights of the graph.

Outlier detection in graph signals poses a different set of challenges when compared to anomaly detection in other types of data. The underlying relationships between the signals at different vertices of the graph result in a narrower class of "nominal" data points compared to the situation when the signal at each vertex is independent of that at other vertices. Therefore, conventional outlier detection techniques for multi-dimensional data sets that test whether the norm of the data vector is outside a certain range of values would be overly conservative for graph signals. Simply using the magnitude of the signal at a vertex is not sufficient to detect if the observation is an outlier or not. A particular observation of the graph signal may be considered an outlier because the *spatial distribution* of the signal magnitudes across the vertices significantly differs from other observations within the data set. In our work, we formalize these different notions of outlier characteristics, and derive criteria for detecting outliers in graph signals.

A. Prior work

Several techniques exist to detect outliers in multidimensional data sets. One approach is to identify clusters, and any observation that falls far away from a cluster is defined as an outlier [1]. Alternatively, an underlying multivariate statistical distribution of the data is assumed, and tests are performed to see if an observation lies at the extremes of the distribution [2], [3]. Techniques from signal processing such as filtered wavelet transforms for multidimensional data sets have also been considered [4]. In our work, we propose and analyze an outlier detection method that extends the above-mentioned signal processing techniques to analyze the spatial distribution of data in the graph domain.

Previous literature related to graph-based data have used information theory to identify structural features [5] and spatial outliers [6]. One particular characteristic, *Total Variation (TV)*, a measure of the smoothness of a signal supported on a graph, has been used as a feature vector for classification [7]. While the outlier detection problem has been addressed in several contexts (for example, time-series data [8]), it has received little attention in the context of graph signals. Our work attempts to fill this gap by using TV as a metric to identify graph signals with off-nominal or unexpected spatial distributions across the vertices. For a comprehensive overview of general graph signal processing techniques, we refer the reader to [9].

This work was partially supported by NSF under CPS Award No. 1739505 and an NSF Graduate Research Fellowship (Max Li). The authors would like to acknowledge useful discussions with Kristyn Pantoja from Texas A&M University.

Karthik Gopalakrishnan, Max Z. Li and Hamsa Balakrishnan are with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. {karthikg, maxli, hamsa}@mit.edu

B. Contributions

We define a new notion of an outlier in a high-dimensional dataset originating from a networked system. We extend prior works in outlier detection on graph signals by considering the spatial distribution of the signals rather than only using the signal magnitude. Specifically, we make a distinction between two kinds of outliers – *outliers in scale* and *outliers in spatial distribution* – to aid in the interpretability of our approach. We develop analytical bounds on the Total Variation (TV) of the graph signal in order to identify outliers. Furthermore, we derive analytic expressions for the mean and variance of the TV for a multivariate Gaussian signal on a graph. The proposed methodology can identify multiple outliers simultaneously, provide insights as to why a particular observation was identified as an outlier, and is computationally inexpensive. The availability of such analytic expressions allows us to perform outlier detection even in the case of limited datasets, when there is insufficient data for empirical techniques. Finally, we extend our results to a "partial information" setting, where the exact strength of the interaction between adjacent graph vertices is unknown. Through simulations, we provide further intuition for the bounds and relationships that we characterize. Finally, we demonstrate our methodology using real-world air traffic delay data.

II. Setup and Notations

Consider a data set with M observations $\mathscr{O}_M = \{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(k)},\ldots,\mathbf{x}^{(M)}\}$, with each observation $\mathbf{x}^{(i)} \in \mathbb{R}^{N\times 1}$ and $\mathbf{x}^{(i)} = \left(x_1^{(i)},\ldots,x_N^{(i)}\right)^\intercal$. We drop the superscript when talking about a generic observation or data point, and refer to it as $\mathbf{x}$. The N elements (or features) of $\mathbf{x}$ are not independent, and their pairwise interactions are captured using an undirected graph $G=(V,E)$ with $|V|=N$ vertices and adjacency matrix A describing the edges E. We assume that the interaction between any two elements is symmetric, hence $A=A^\intercal$. The observation $\mathbf{x}$ can equivalently be considered as signals x_i supported on vertices $i \in V$ on the graph with connectivities and weights given by A.

The empirical mean of the signal at vertex i is $\hat{\mu}_i = \frac{1}{M}\sum_{k=1}^{M} x_i^{(k)}$. For each pair of unique vertices (i,j) and given the set of observations $\mathscr{O}_M$, we can compute the sample Pearson correlation coefficient, denoted as $r_{ij|\mathscr{O}_M}$:

$$r_{ij|\mathscr{O}_M} = \frac{\sum_{k=1}^{M}\left(x_i^{(k)}-\hat{\mu}_i\right)\left(x_j^{(k)}-\hat{\mu}_j\right)}{\sqrt{\sum_{k=1}^{M}\left(x_i^{(k)}-\hat{\mu}_i\right)^2}\sqrt{\sum_{k=1}^{M}\left(x_j^{(k)}-\hat{\mu}_j\right)^2}}. \quad (1)$$

In terms of notation, bold-face fonts indicate vectors, random variables are given in upper case, and a "hat" represents an empirically-derived quantity. The graph signal vector $\mathbf{x}$ are assumed to be specific realizations of a random variable $\mathbf{X} = (X_1,\ldots,X_N)^\intercal \in \mathbb{R}^{N\times 1}$, where $\mathbf{X}$ is a multivariate Gaussian random variable $\mathbf{X} \overset{iid}{\sim} \mathscr{N}(\boldsymbol{\mu},\Sigma)$. $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ is the vector of means and $\Sigma \in \mathbb{S}^{N\times N}$ is the positive semi-definite covariance matrix. The correlation coefficient between the signals on two adjacent vertices i and j is given by

$$\rho_{ij} = \frac{\mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)]}{\sqrt{\mathbb{E}[(X_i-\mu_i)^2]\mathbb{E}[(X_j-\mu_j)^2]}}. \quad (2)$$

The sample correlation coefficient, $r_{ij|\mathscr{O}_M}$, is a consistent estimator of ρ_{ij}, *i.e.*, $\lim_{M\to\infty}\left(r_{ij|\mathscr{O}_M}\right) = \rho_{ij}$. The combinatorial graph Laplacian corresponding to an adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{N\times N}$ is given by $\mathfrak{L} = D - A$, where $D = [d_{ij}] \in \mathbb{R}^{N\times N}$ is a diagonal matrix with $d_{ii} = \sum_j a_{ij}$.

Definition 1. *The Total Variation (TV) of a signal* $\mathbf{x}$ *supported on the vertices of a graph with adjacency matrix* A *and graph Laplacian* $\mathfrak{L}$ *is defined as:*

$$TV(\mathfrak{L},\mathbf{x}) = \frac{1}{2}\sum_{i\neq j} a_{ij}(x_i-x_j)^2 = \mathbf{x}^\intercal\mathfrak{L}\mathbf{x}. \quad (3)$$

For notational brevity, since the TV is always defined with respect to some graph Laplacian, we will write $TV(\mathbf{x})$ *when examining the observation of a particular signal's TV, and* $TV(\mathbf{X})$ *for the TV as a derived random variable.*

The TV can be interpreted as a measure of the smoothness of a graph signal. A higher value of the TV means that there is more variation of the signal across edges. Specifically, edges with a higher weight would contribute more to the TV when the signal values differ at the vertices joined by those edges. For the results in this paper, we set the edge weights (a_{ij}) of the graph to be the correlation between signals on pairs of vertices (either ρ_{ij} or its estimate).

III. TV as a Metric for Outlier Detection

We now present two definitions of outliers, and argue that these outliers can be identified through examining the smoothness of a graph signal via the TV metric. Recall that all the observations of our graph signal vector $\mathbf{x}$ are assumed to be realizations of the random variable $\mathbf{X} \overset{iid}{\sim} \mathscr{N}(\boldsymbol{\mu},\Sigma)$. If the magnitude, or scale, of each measured vertex signal differs significantly from historical observations, *i.e.*, if $\|\mathbf{x}\|$ differs by a significant amount relative to $\mathbb{E}[\|\mathbf{X}\|]$, then we call this observation an *outlier in scale* or a *scale outlier*. This is the most intuitive definition of an outlier for multidimensional data: Observations that differ significantly in magnitude compared to what is expected.

Definition 2. *An observation is considered to be an* outlier in scale *or a* scale outlier *of level* k *if*

$$\|\mathbf{x}\| \notin \left[\mathbb{E}[\|\mathbf{X}\|] - k\sqrt{Var[\|\mathbf{X}\|]},\ \mathbb{E}[\|\mathbf{X}\|] + k\sqrt{Var[\|\mathbf{X}\|]}\right],$$

for some $k \geq 0$*. In other words, an observation is considered an outlier in scale if the norm of* $\mathbf{x}$ *does not lie within* k *standard deviations of its expected value.*

Note that $\|\bullet\| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ is any valid norm map given a properly-equipped metric space. On the other hand, if the spatial distribution exhibited in a graph signal $\mathbf{x}$ is unexpected given historical observations, then the observation of that signal is called an *outlier in distribution* or *distribution*

outlier. Furthermore, we formalize the following notions of *strong* and *weak outliers in distribution*:

Definition 3. *An observation* $\mathbf{x}$ *is considered a* strong distribution outlier *or a strong outlier in distribution of level* k *if* $TV(\mathbf{x}) \notin [A,B]$*, where:*

$$A = \mathbb{E}[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|] - k\sqrt{Var[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]}$$
$$B = \mathbb{E}[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|] + k\sqrt{Var[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]}.$$

In other words, we say that an observation is a strong *outlier in distribution if the value of its TV does not lie within* k *standard deviations of its expected value, where both the expectation and standard deviation are conditioned on the magnitude or norm of* $\mathbf{x}$*.*

Definition 4. *An observation* $\mathbf{x}$ *is considered a* weak distribution outlier *or a weak outlier in distribution of level* k *if* $TV(\mathbf{x}) \notin [A',B']$*, where:*

$$A' = \mathbb{E}[TV(\mathbf{X})] - k\sqrt{Var[TV(\mathbf{X})]}$$
$$B' = \mathbb{E}[TV(\mathbf{X})] + k\sqrt{Var[TV(\mathbf{X})]}$$

The definition of a weak *outlier in distribution is similar to the strong equivalent, except that it corresponds to the unconditioned probability distributions.*

We now motivate the intuition behind using TV as an outlier detection metric. The edge weights of the graph are given by the correlations of pairwise adjacent vertex signals. If the correlation is low, then any observed difference in the signal magnitudes is in some sense expected, and the contribution of that pair of vertex signals to the TV given by $\rho_{ij}(x_i - x_j)^2$ is small since ρ_{ij} is small. However, if the realized TV is large, then it means that $(x_i - x_j)^2$ was significantly larger than usual, indicating an unusual, or rare distribution of signals across two vertices. Thus, an observation that can be considered an outlier can be identified by deviations in its TV, summed across all unique pairs of vertices $(i,j) \in V \times V$. As we will see later, this idea can be extended to other values of ρ_{ij} as well, and forms the basis for identifying weak outliers in distribution.

While TV by itself is a useful metric to classify graph signals with an unexpected spatial distribution across vertices, a more faithful metric that considers *only* the spatial distribution and not the influence of signal magnitude is desirable. This requires conditioning the TV on $\|\mathbf{X}\| = \|\mathbf{x}\|$, leading to the definition of strong outliers in distribution. For classifying strong outliers in distribution, the bounds now fluctuate dependent on the realized $\|\mathbf{x}\|$.

In order to perform a tractable and interpretable analysis using the definitions, we make some assumptions regarding the observations $\mathbf{x}$. We consider only the 1-norm, *i.e.* $\|\mathbf{x}\| = \sum_i |x_i|$, and assume that all vertex signals are non-negative. This assumption is generally acceptable, as it is true that for many physical systems, the vertex signals is always a non-negative quantity (*e.g.*, delays at an airport, number of cars at an intersection, etc.). These two assumptions allow us to express $\|\mathbf{x}\|$ as $\sum_i x_i$.

A notional representation of the various bounds for strong and weak outliers in distribution and scale is shown in Figure 1. While deriving explicit analytical bounds for strong outliers in distribution remain an open challenge, we have successfully obtained analytical expressions for the bounds on weak outliers in distribution (Section IV). Via simulation, we evaluate empirically-derived bounds for strong outliers in distribution, and show that the gap between the strong and weak outlier bounds depends on $\boldsymbol{\mu}$ and Σ. To summarize, examining $\mathrm{TV}(\mathbf{x})$ as a function of $\|\mathbf{x}\|$ provides a low-dimensional projection of a complex, networked data set with pairwise interdependencies, and enables the detection of outliers in distribution and scale.

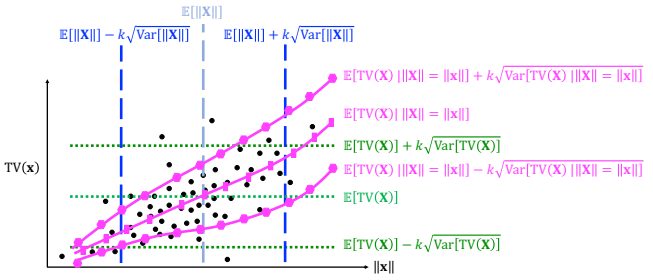

Fig. 1. Notional representation of bounds that we will derive analytically (outliers in scale and weak outliers in distribution), and empirically (strong outliers in distribution).

While the values for the mean and variance of the TV can be obtained from the set of observations $\mathscr{O}_M$, our contribution lies in deriving analytical closed-form expressions for such bounds. This is useful for two reasons: (1) Even though we have accurate information about the means and variances of signals at each node, there may be insufficient data points to reliably estimate the expectation and variance of TV; (2) We can obtain intuition regarding how the properties of the signals (data set), *i.e.* $\boldsymbol{\mu}$ and Σ, determine the bounds for outlier detection.

IV. Bounds for Outlier Detection

In this section, we derive analytic expressions for the expectation and variance of the TV for the case with "complete information" (that is, ρ_{ij} is known) as well as the "partial information" case (where only bounds on ρ_{ij} are known). Finally, we also present the bounds for outliers in scale.

A. Expectation and variance of TV given complete information

We first compute the expectation of the TV for a random graph signal $\mathbf{X}$. Note that computing the expectation does not require the distribution to be a multivariate Gaussian, only that it has a finite mean and variance.

$$\begin{aligned}\mathbb{E}[\mathrm{TV}(\mathbf{X})] &= \mathbb{E}\left[\frac{1}{2}\sum_{i\neq j}\{\rho_{ij}(X_i - X_j)^2\}\right] \\ &= \frac{1}{2}\sum_{i\neq j}\{\rho_{ij}\left(\mathbb{E}\left[X_i^2\right] + \mathbb{E}\left[X_j^2\right] - 2\mathbb{E}\left[X_i X_j\right]\right)\}\end{aligned} \tag{4}$$

Substituting $\mathbb{E}[X_iX_j] = \mu_i\mu_j + \rho_{ij}\sigma_i\sigma_j$ from (2) and $\mathbb{E}\left[X_i^2\right] = \mu_i^2 + \sigma_i^2$ in (4), we get

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left[(\mu_i-\mu_j)^2 + (\sigma_i^2+\sigma_j^2-2\rho_{ij}\sigma_i\sigma_j)\right]\right\}. \tag{5}$$

Equation (5) gives the contribution of each edge to the TV. The contribution of each edge – alternatively, of each unique pair of vertices – depends on the difference in signal means, variances, and correlation. We examine some specific cases that impose certain conditions on the means, variances, and correlations:

1) If there is no correlation between the signals at any two vertices, *i.e.* $\rho_{ij} = 0$ for all vertices $i \neq j$, then $\mathbb{E}[\mathrm{TV}(\mathbf{x})] = 0$.
2) If there is perfect correlation between the signals at any two vertices, *i.e.*, $\rho_{ij} = 1$ for all vertices $i \neq j$, then $\mathbb{E}[\mathrm{TV}(\mathbf{x})] = \frac{1}{2}\sum_{i\neq j}\left\{(\mu_i-\mu_j)^2 + (\sigma_i-\sigma_j)^2\right\}$.
3) If the means are the same for all signals, *i.e.*, $\mu_i = \mu_j$ for all vertices $i \neq j$, then $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left[\sigma_i^2+\sigma_j^2-2\rho_{ij}\sigma_i\sigma_j\right]\right\}$.
4) If all pairwise vertex signals have the same means ($\mu_i = \mu_j$), variances ($\sigma_i = \sigma_j = \sigma$), and correlation coefficient ($\rho_{ij} = \rho$) for all vertices $i \neq j$, then

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \sum_{i\neq j}\left\{\rho\sigma^2(1-\rho)\right\} = N(N-1)\rho\sigma^2(1-\rho). \tag{6}$$

The computation of the variance of the TV is more involved. While we can derive a closed-form analytic equation, it cannot be expressed in a simple form. The outline of the derivation is as follows:

$$\mathrm{Var}[\mathrm{TV}(\mathbf{X})] = \mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] - \mathbb{E}[\mathrm{TV}(\mathbf{X})]^2. \tag{7}$$

The second term, $\mathbb{E}[\mathrm{TV}(\mathbf{X})]^2$, is known from (5). The first term can be expanded as

$$\mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] = \frac{1}{4}\mathbb{E}\left[\left(\sum_{i\neq j}\left\{\rho_{ij}(X_i-X_j)^2\right\}\right)^2\right]. \tag{8}$$

Expanding (8) and using linearity of the expectation operator gives us an expression in terms of $\mathbb{E}\left[X_i^4\right]$, $\mathbb{E}\left[X_i^3X_j\right]$, and $\mathbb{E}\left[X_i^2X_j^2\right]$. We know $\mathbb{E}\left[X_i^4\right] = \mu_i^4 + 6\mu_i^2\sigma_i^2 + 3\sigma_i^4$. The analytical expression for higher-order moments of the product of dependent Gaussian random variables was derived in [10], with a more implementable form that also accounts for non-zero means given in [11].

Proposition 1 (Isserlis (1918) and Kan (2008))**.** *Suppose* $\mathbf{X} = (X_1,\ldots,X_N)^\mathsf{T} \sim \mathcal{N}(\boldsymbol{\mu},\Sigma)$*, where* Σ *is an* $N\times N$ *positive semi-definite matrix. For non-negative integers* s_1 *to* s_N*, we have*

$$\mathbb{E}\left[\prod_{i=1}^N X_i^{s_i}\right] = \sum_{\nu_1=0}^{s_1}\cdots\sum_{\nu_N=0}^{s_N}\sum_{r=0}^{[s/2]}\binom{s_1}{\nu_1}\cdots\binom{s_N}{\nu_N} \times\left\{\frac{\left(\frac{\boldsymbol{h}^\mathsf{T}\Sigma\boldsymbol{h}}{2}\right)^r(\boldsymbol{h}^\mathsf{T}\boldsymbol{\mu})^{s-2r}}{r!(s-2r)!}\right\},$$

where $s = s_1+\cdots+s_N$ *and* $\boldsymbol{h} = \left(\frac{s_1}{2}-\nu_1,\ldots,\frac{s_N}{2}-\nu_N\right)^\mathsf{T}$.

Proof. See [10] and [11]. □

Substituting the results from Proposition 1 in the expansion of (8), we can obtain an analytical expression for $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as a function of $\boldsymbol{\mu}$ and Σ. While the variance of the TV involves a significant number of terms and cannot be written in a simple algebraic form, it can be easily evaluated symbolically and numerically using a computer. We give the following case when we can evaluate the variance of the TV easily:

Proposition 2. *If* $\mathbb{E}[TV(\mathbf{X})] = 0$ *and* $\rho_{ij} \geq 0$, $\forall i,j$ *(or* $\rho_{ij} \leq 0$, $\forall i,j$*), then* $Var[TV(\mathbf{X})] = 0$.

Proof. Since $\mathrm{TV}(\mathbf{X}) \geq 0$ (or $\mathrm{TV}(\mathbf{X}) \leq 0$), $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = 0 \implies \mathrm{TV}(\mathbf{X}) = 0$. Hence, $\mathrm{Var}[\mathrm{TV}(\mathbf{X})] = 0$. □

However, note that $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = 0$ is not a necessary condition for $\mathrm{Var}[\mathrm{TV}(\mathbf{X})] = 0$. If $\rho_{ij} = 1$, $\mu_i \neq \mu_j$, and $\sigma_i^2 = \sigma_j^2$, then $\mathbb{E}[\mathrm{TV}(\mathbf{X})] \neq 0$ and $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ depends on $(\mu_i-\mu_j)^2$. However, the random variable $X_i - X_j$ will always be a constant, and thus $\mathrm{Var}[\mathrm{TV}(\mathbf{X})] = 0$ still holds.

The analytic expressions we derived for the expectation and variance of the TV allow us to compute the bounds given in Definition 4, thereby delineating weak outliers in distribution.

B. Expectation and variance of TV given partial information

We consider the setting where the mean and variance of the signal at each vertex is known, *i.e.*, $\boldsymbol{\mu}$ and σ_i^2 is known for all vertices $i \in V$, but the correlation between any pair of vertex signals, ρ_{ij}, is not known precisely. This can happen in real systems if a data-reporting agent at each vertex can only obtain local vertex information and does not share information with other agents. In such cases, we can only obtain marginal information regarding individual means and variances at every vertex. While the exact interdependency between two pairwise vertices may not be available, we develop a theory regarding detecting outliers in graph signals using only an approximate bound on the nature of the interaction. Specifically, we assume that the observations are drawn from a multivariate Gaussian distribution with a fixed $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ and $\Sigma \in \mathbb{S}^{N\times N}$, but the precise value of ρ_{ij} is unknown.

For the propositions we construct and prove in this section, we require all the correlation coefficients to have the same sign, *i.e.* all $\rho_{ij} \geq 0$ or all $\rho_{ij} \leq 0$, $\forall i,j \in V$. We consider the former, and introduce the following projections of the correlation coefficients into the non-negative half-plane:

$\rho_{ij}^+ = \max\{0, \rho_{ij}\}$ and $r_{ij|\mathscr{O}_M}^+ = \max\{0, r_{ij|\mathscr{O}_M}\}$. A similar projection can be defined for non-positive correlations, and all results follow analogously.

We derive tight bounds on $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ when we are only given bounds on each correlation coefficient $0 \leq \nu_{ij} < \rho_{ij} < \varepsilon_{ij} \leq 1$. We first present the setup leading to the two propositions that quantify the corresponding bounds on $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$.

Every edge between unique pairs of vertices i and j in a graph obtained from the set of observations $\mathscr{O}_M$ is assigned a weight $r_{ij|\mathscr{O}_M}^+$. Note that $r_{ij|\mathscr{O}_M}^+$ is a consistent estimator of ρ_{ij}^+, since the projection into the non-negative half-plane can be alternatively defined by piecewise affine transformations [12]. However, the estimator is biased, and we have from [12] that

$$\mathbb{E}[r_{ij|\mathscr{O}_M}] = \rho_{ij}\left(1 - \frac{1-\rho_{ij}^2}{2M} + O\left(\frac{1}{M^2}\right)\right). \tag{9}$$

Additionally, $r_{ij|\mathscr{O}_M}$ is a random variable with a valid probability density function $f(r_{ij|\mathscr{O}_M}|\rho_{ij})$ that has an explicit form expressible in terms of the Euler gamma function $\Gamma(x)$ [13], [14]. It is important to note that the probability density function is dependent only on the number of observations M, and independent of any new realizations of the random variables X_i and X_j. We will make use of this fact in the proofs for our propositions.

With the above setup, we can now redefine TV for an *unobserved* graph signal vector $\mathbf{X} \sim \mathscr{N}(\boldsymbol{\mu}, \Sigma)$ with respect to the graph Laplacian $\mathfrak{L} \in \mathbb{S}^{N\times N}$ constructed using $\mathscr{O}_M$. Note that $\mathrm{TV}(\mathbf{X})$ is a derived random variable,

$$\mathrm{TV}(\mathbf{X}) = \frac{1}{2}\sum_{i\neq j}\left\{r_{ij|\mathscr{O}_M}^+ (X_i - X_j)^2\right\}. \tag{10}$$

In Proposition 3, we provide bounds on $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$, and in Proposition 4, we proceed to derive bounds on $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$.

Proposition 3. *Suppose that* $0 \leq \nu_{ij} < \rho_{ij}^+ < \varepsilon_{ij} \leq 1$ *for all unique pairs of vertices* $i, j \in V$. *Then, there exists scalars* δ_1 *and* δ_2, *with* $\delta_2 \geq 0$, *such that* $\max\{0, \delta_1\} \leq \mathbb{E}[TV(\mathbf{X})] < \delta_2$.

Proof. Since $r_{ij|\mathscr{O}_M}$ is a random variable dependent only on M previous observations, and X_i, X_j are currently unobserved random variables, the expectation operator factorizes over the expression for the TV (from (10)):

$$\begin{aligned}\mathbb{E}[\mathrm{TV}(\mathbf{X})] &= \frac{1}{2}\sum_{i\neq j}\left\{\mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right]\mathbb{E}\left[(X_i - X_j)^2\right]\right\} \\ &= \frac{1}{2}\sum_{i\neq j}\left\{\mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right]\left((\mu_i-\mu_j)^2 + \sigma_i^2 + \sigma_j^2 - 2\rho_{ij}^+\sigma_i\sigma_j\right)\right\}.\end{aligned} \tag{11}$$

Since the bias of $r_{ij|\mathscr{O}_M}^+$ is given in (9), for any M, there exists a $\gamma_{ij} > 0$ that is a function of M (and $\lim_{M\to\infty}\gamma_{ij} = 0$) such that

$$\left|\mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right] - \rho_{ij}^+\right| < \gamma_{ij} \tag{12}$$

$$\Leftrightarrow \rho_{ij}^+ - \gamma_{ij} < \mathbb{E}[r_{ij|\mathscr{O}_M}^+] < \rho_{ij}^+ + \gamma_{ij}. \tag{13}$$

We can use the fact that $\nu_{ij} < \rho_{ij}^+ < \varepsilon_{ij}$ in order to rewrite the bounds of (13),

$$\max\{0, \nu_{ij} - \gamma_{ij}\} \leq \mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right] < \varepsilon_{ij} + \gamma_{ij}. \tag{14}$$

The maximum operator is included since $\nu_{ij} - \gamma_{ij}$ can be negative, but we know that the expectation of a non-negative random variable $\mathrm{TV}(\mathbf{X})$ is bounded below by 0. We aim to use (14) in conjunction with the bounds on ρ_{ij}^+ to bound (11). We focus first on deriving the upper bound; such an upper bound is given by evaluating (11) for the largest-possible contributions from the positive terms, and the smallest-possible deductions from the negative term. This gives:

$$\begin{aligned}\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\Big\{&\underbrace{\mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right]}_{<\varepsilon_{ij}+\gamma_{ij}}\left((\mu_i-\mu_j)^2 + \sigma_i^2 + \sigma_j^2\right) \\ &-2\underbrace{\mathbb{E}\left[r_{ij|\mathscr{O}_M}^+\right]\rho_{ij}^+}_{<\nu_{ij}\max\{0,\nu_{ij}-\gamma_{ij}\}}\sigma_i\sigma_j\Big\}.\end{aligned} \tag{15}$$

Therefore, this gives the upper bound $\mathbb{E}[\mathrm{TV}(\mathbf{X})] < \delta_2$, where

$$\delta_2 = \frac{1}{2}\sum_{i\neq j}\left\{\tilde{\varepsilon}_{ij}\left((\mu_i-\mu_j)^2 + \sigma_i^2 + \sigma_j^2\right) - 2\tilde{\nu}_{ij}^+\nu_{ij}\sigma_i\sigma_j\right\} \tag{16}$$

along with rewriting $\tilde{\varepsilon}_{ij} = \varepsilon_{ij} + \gamma_{ij}$ and $\tilde{\nu}_{ij}^+ = \max\{0, \nu_{ij} - \gamma_{ij}\}$.

To get the lower bound, we evaluate (11) for the smallest-possible contribution from the positive terms and the largest-possible contribution in terms of magnitude from the negative terms. This gives $\mathbb{E}[\mathrm{TV}(\mathbf{X})] > \delta_1$, where

$$\delta_1 = \frac{1}{2}\sum_{i\neq j}\left\{\tilde{\nu}_{ij}^+\left((\mu_i-\mu_j)^2 + \sigma_i^2 + \sigma_j^2\right) - 2\,\tilde{\varepsilon}_{ij}\varepsilon_{ij}\sigma_i\sigma_j\right\} \tag{17}$$

Note that δ_1 and δ_2 are functions of the bounds on ρ_{ij}^+ and M, so we have that

$$\lim_{\substack{\nu_{ij}\to\rho_{ij}^+\\ \varepsilon_{ij}\to\rho_{ij}^+}}(\delta_1) = \lim_{\substack{\nu_{ij}\to\rho_{ij}^+\\ \varepsilon_{ij}\to\rho_{ij}^+}}(\delta_2) = \mathbb{E}[\mathrm{TV}(\mathbf{X})]. \tag{18}$$

□

Proposition 4. *Suppose* $0 \leq \nu_{ij} < \rho_{ij}^+ < \varepsilon_{ij} \leq 1$ *for all unique pairs of vertices* $i, j \in V$. *Then, there exists scalars* δ_3 *and* δ_4, *with* $\delta_4 \geq 0$, *such that* $\max\{0, \delta_3\} \leq Var[TV(\mathbf{X})] < \delta_4$.

Proof. The idea behind the proof is similar to the proof for Proposition 3. We expand $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as done in (7) and (8). Proposition 1 can then be used to obtain the appropriate higher-order moments. This gives $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as a scalar

quantity that depends on $\mathbb{E}\left[r^{+}_{ij|\mathscr{O}_M}\right]$ and ρ^{+}_{ij}. Finally, these two terms can be bounded appropriately to obtain the desired bounds on $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. □

Using Propositions 3 and 4, we can compute the worst-case bounds for weak outliers in distribution of level k, given by $\mathrm{TV}(\mathbf{x}) \notin \left[\max\left\{0, \delta_1 - k\sqrt{\delta_4}\right\}, \delta_2 + k\sqrt{\delta_4}\right] \implies \mathbf{x}$ is a weak outlier in distribution.

C. Bounds for outliers in scale

We derived bounds for $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ in the previous subsection in order to examine weak outliers in distribution. Here we focus on bounds for outlier in scale. Recall that we are assuming $\mathbf{x}$ has non-negative entries, allowing for the redefinition of the 1-norm as $\|\mathbf{X}\|_1 = \sum_i X_i$. We have that the expectation of $\|\mathbf{X}\|$ is:

$$\mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i] = \sum_i \mu_i. \tag{19}$$

Similarly, the variance of $\|\mathbf{X}\|$ is:

$$\mathrm{Var}\left[\sum_i X_i\right] = \sum_i \mathrm{Var}[X_i] + \sum_{i\neq j} \rho_{ij}\sigma_i\sigma_j. \tag{20}$$

In the setting with complete information, since the correlations are known, the expression for the variance simplifies to $\mathrm{Var}[\sum_i X_i] = \mathbf{1}^\intercal \Sigma \mathbf{1}$. In the setting with partial information, if we know that the correlations are bounded by $0 \leq \nu_{ij} < \rho_{ij} < \varepsilon_{ij} \leq 1$, we get that

$$\sum_i \sigma_i^2 + \sum_{i\neq j} \nu_{ij}\sigma_i\sigma_j < \mathrm{Var}\left[\sum_i X_i\right] < \sum_i \sigma_i^2 + \sum_{i\neq j} \varepsilon_{ij}\sigma_i\sigma_j. \tag{21}$$

Finally, we note that the TV can be bounded by the Rayleigh quotient since the graph Laplacian is a Hermitian matrix. Let $\lambda_{\max}$ be the largest eigenvalue of $\mathfrak{L}$, then we have

$$\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\intercal \mathfrak{L} \mathbf{x} \leq \lambda_{\max} \|\mathbf{x}\|_2^2 \leq \lambda_{\max} \|\mathbf{x}\|_1^2. \tag{22}$$

However, the upper bound provided by the Rayleigh quotient is not useful for outlier detection as it is extremely conservative in practice. The bounds that we have obtained are much tighter than the ones that would be derived via the Rayleigh quotient.

V. General Network Simulation Results

We aim to convey two ideas via simulations: First, we compute the bounds on strong distribution outliers using simulations, and compare them against the theoretically-derived weak distribution outlier bounds. We observe that the difference between these two bounds depends on the underlying $\boldsymbol{\mu}$ and Σ of the data, and we show two examples to highlight that dependency. Second, we provide some more intuition on the partial information case by empirically evaluating the mean and variance of the TV for a range of ρ. We observe non-monotonic variations with ρ which are difficult to predict *a priori*, highlighting the importance of our analytical bounds from Propositions 3 and 4.

A. Strong and weak bounds on TV in simulated networks

Using two simulations, we examine the performance gap between bounds on strong outliers in distribution versus bounds on weak outliers in distribution. We also demonstrate the utility of distinguishing outliers using TV rather than the underlying distribution, as the former can provide much more useful interpretations, particularly in a networked setting. The idea for both simulations is to assume a value of $\boldsymbol{\mu}$ and Σ and generate normally-distributed $M = 1 \times 10^6$ observations. The data points are plotted on a $\mathrm{TV}(\mathbf{x})$ versus $\|\mathbf{x}\|$ plot, and the quantities $\hat{\mathbb{E}}[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]$ and $\widehat{\mathrm{Var}}[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]$ required for obtaining the bounds on strong outliers in distribution are computed empirically by binning $\|\mathbf{x}\|$ and conditioning on each bin. The theoretically-derived bounds for weak distribution outliers and scale outliers are also plotted. Additionally, we color each observation by the density obtained from evaluating the probability density function value of the realized $\mathbf{x}$. Note that if only the underlying distribution was used for outlier detection, all black-colored trials could be considered as outliers. This stands in contrast to the bounds provided by our derivations.

In the first simulation, we choose $N = 2$, $\sigma_1 = \sigma_2 = 1$, $\rho_{12} = 0.5$ and $\boldsymbol{\mu} = (545.34, 582.13)^\intercal$. The results in Figure 2 shows that the bounds for the weak and strong outliers in distribution are very close. Thus, the computation of the weak distribution bounds theoretically would be a good approximation to the strong distribution bounds.

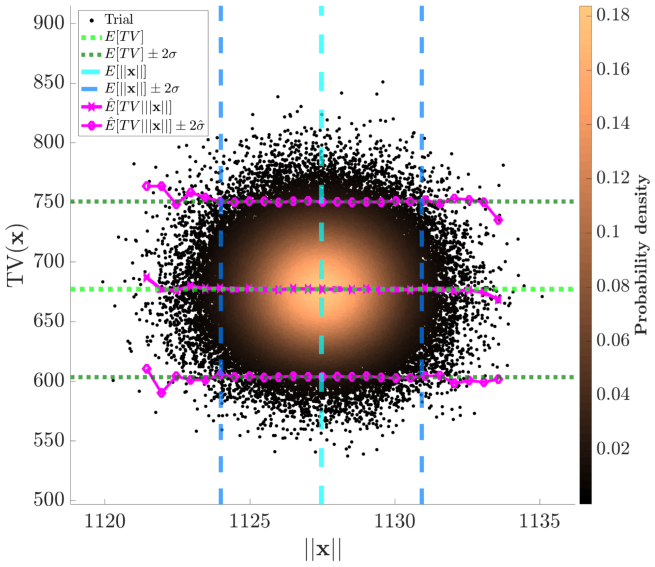

Fig. 2. TV versus 1-norm of the graph signal for a generic bi-vertex graph, with scale outlier, weak outlier in distribution, and empirically-derived strong outlier bounds. Each observation $(\|\mathbf{x}\|, \mathrm{TV}(\mathbf{x}))$ is colored by its probability density $f_{\mathbf{X}}(\mathbf{X} = \mathbf{x})$.

For the second simulation, we have that $N = 30$, and we estimate $\boldsymbol{\mu}$ and Σ from a real-world data set of air traffic delay signals (see Section VI). We see that the weak outliers in distribution bounds tend to be overly-liberal (for low values of $\|\mathbf{x}\|$) or overly-conservative (for large values of $\|\mathbf{x}\|$) in terms of distinguishing outliers. This behavior is

expected, as the weak bounds are not conditioned on $\|\mathbf{x}\|$. For future work, we are interested in deriving an analytical strong outlier in distribution bound. Figure 3 also shows the usefulness of TV as an outlier-distinguishing metric over simply using the density of the underlying distribution; the former takes into account pairwise interactions and the strength of that interaction, whereas the latter does not.

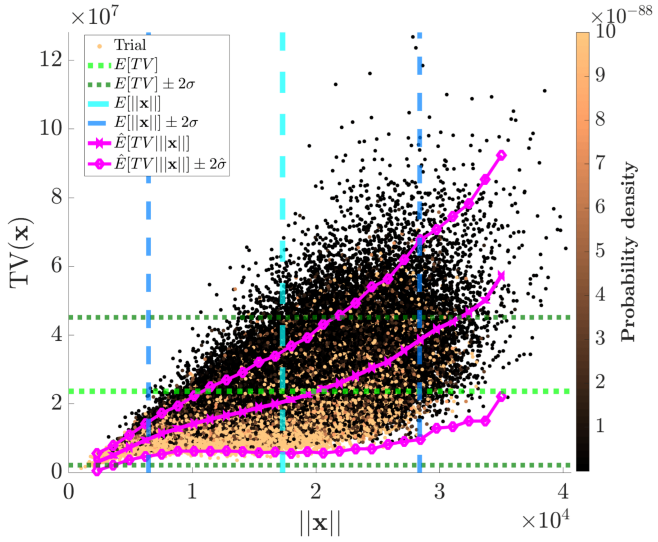

Fig. 3. TV versus 1-norm of simulated graph signals within a 30-vertex graph; data set from the US air transportation network.

B. Expectation and variance of TV as a function of ρ

Through simulations, we demonstrate the utility of our analytical bounds on $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ for the setting with partial information regarding correlations. Using the results from Section IV-A, it is theoretically possible – although computationally intractable – to evaluate the exact expectation and variance of the TV for a set of discretized values of $\rho_{ij} \in [\nu_{ij}, \varepsilon_{ij}] \subseteq [0,1]$ given the signal mean and variance at each vertex. The intractability is apparent in two locations: First, the search space is exponential in the number of edges, and a discretization of ρ_{ij} into N_ρ intervals for each edge requires $N_\rho^{N\times(N-1)}$ evaluations for $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. Secondly, the evaluated functions are not convex in ρ_{ij}, indicating that methods such as gradient descent cannot be used to obtain worst-case bounds. We demonstrate this non-monotonic behavior in Figure 4. However, since our bounds derived in Propositions 3 and 4 are tight, this allows for a computationally efficient strategy to evaluate the optimization over the search space of $0 \le \nu_{ij} < \rho_{ij} < \varepsilon_{ij} \le 1$.

For the simulations depicted in Figure 4, we draw $M = 5\times10^4$ data points from a multivariate Gaussian distribution with $N = 5, \sigma_i = 10$, $\forall i$, and $\rho_{ij} = \rho$, $\forall i,j$ where we vary ρ independently within the interval $[0,1]$. We consider four examples, each initialized with a different $\boldsymbol{\mu}$. These four examples have differing ranges for the values of the vertex signal means, parameterized by a "tolerance factor" η. We choose $\eta \in \{0, 0.1, 0.25, 1.5\}$ and $\mu_i \overset{iid}{\sim} 100(1-\eta)+200\eta X_U$, where $X_U \overset{iid}{\sim} \mathrm{Unif}(0,1)$. Higher η indicates that signals have higher

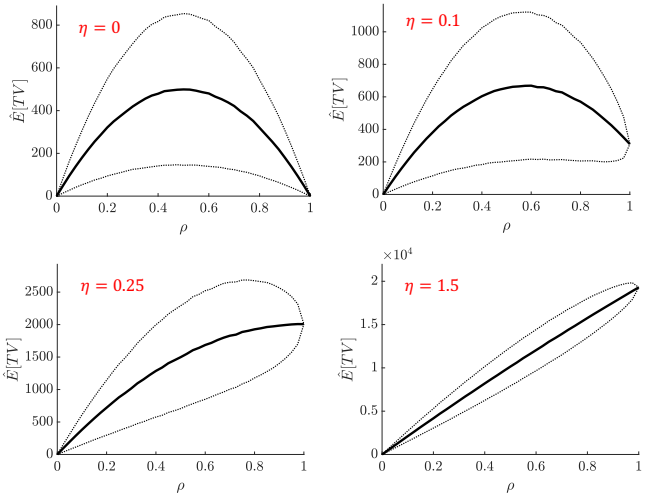

Fig. 4. Empirically-derived curves for the expectation and variance of TV as a function of correlation and parameterized by $\boldsymbol{\mu}$ via η.

baseline difference in terms of magnitudes across a pair of vertices.

If $\eta = 0$, then $\mu_i = \mu_j$ for all unique pairs of vertices $i, j \in V$, and we degenerate to the case depicted in (6) where $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ is quadratic in ρ. For larger tolerances η, we see that $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ becomes monotonic and closer to a linear function, as depicted in Figure 4. The dotted lines in Figure 4 show the $\pm\sqrt{\mathrm{Var}[\mathrm{TV}(\mathbf{X})]}$ bounds around the expected TV. Figure 4 highlights the dynamic behavior of the bounds as a function of ρ, even in our relatively constrained setting of $\rho_{ij} = \rho$, $\forall i,j \in V$.

We conclude this section with a few remarks. The simulations confirm that empirical results shown in Figure 4 match theoretical predictions (not shown for simplicity) from the previous section. At $\rho = 1$, all four examples had zero variance since $\sigma_i = \sigma$, $\forall i \in V$. If the variances were allowed to vary, *i.e.* $\sigma_i \neq \sigma_j$, then $\mathrm{Var}[\mathrm{TV}(\mathbf{X})] \neq 0$ at $\rho = 1$. Furthermore, we emphasize that our derived bounds are tight and characterize the exploration of the entire search space of variations in correlation. However, not all choices of ρ_{ij} lead to a valid positive semi-definite covariance matrix Σ. Enforcing this constraint analytically may result in tighter bounds, and is an open problem.

VI. Air Traffic Delay Network Example

High delays at major airports are known to have significant economic impacts; when analyzing system performance involving multiple airports and routes, airport-centric delay metrics, such as the sum of all inbound arrival delays and outbound departure delays, are commonly used. However, due to the strong underlying network connectivity, the delays at some airports are highly correlated with delays at other airports. This real-world context is isomorphic to the problem setting of this paper: Identifying signal outliers within a networked and interdependent system.

Airport delays can take on two forms: (1) The obvious problematic scenario when delays are high across all airports; (2) a subtler scenario when delays are high, but with an

unexpected distribution across a specific set of airports. The former corresponds to outliers in scale and can be identified via classical metrics, whereas the latter corresponds to outliers in distribution and is significantly more difficult to identify.

To derive the theoretical weak outlier bounds for this case study, we use $\boldsymbol{\mu}$ and Σ estimated from a data set of the daily total delay (in minutes) at the top 30 US airports (by enplanement) from 2008 to 2017. The resultant weak outlier bounds are plotted in Figure 5. Out of a total 3,653 days (data points), 101 days (2.8%) were classified as outliers in scale, 550 days (15.1%) were classified as weak outliers in distribution, and 60 days (1.6%) were classified as outliers in both scale and distribution, all at a level of $k = 2$. The days we have identified as outliers – particularly ones with high total delays but low TV, and vice versa – are very interesting operationally. The insights gained through this type of outlier detection could help air traffic flow managers make better decisions regarding network flow management and develop robust models for delay dynamics [15].

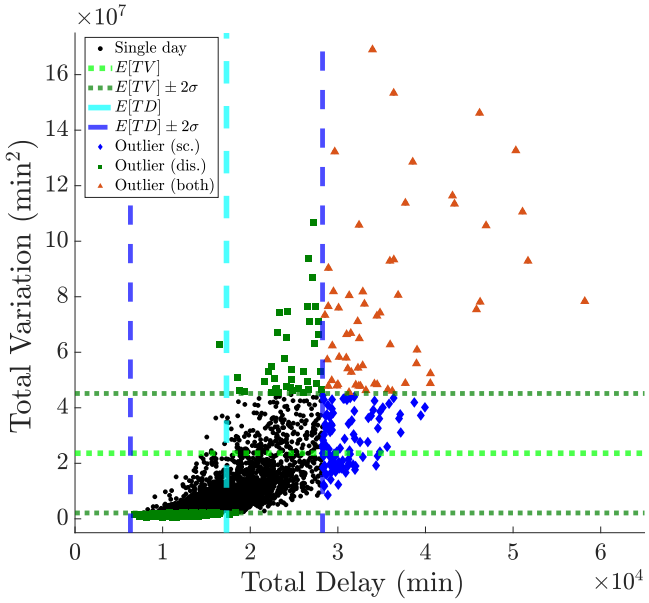

Fig. 5. Identifying outlier days in air traffic delay data from the US air transportation network. Days with large amounts of total delay are outliers in scale (sc.), days with an unexpected spatial distribution of delays are (weak) outliers in distribution (dis.), with some days exhibiting both outlier properties.

VII. Conclusion and Future Work

We defined and derived bounds for identifying outliers with respect to both scale and spatial distribution for networked data. We used the Total Variation (TV) of a graph signal as the metric to perform outlier detection. In a Gaussian setting, we derive analytical expressions and bounds for the expectation and variance of the TV. We demonstrated the usefulness of our bounds and definitions using simulations, and showed the applicability of the proposed methodology through a case study of the air transportation network. While our work takes preliminary steps towards providing a theoretical characterization of TV as a useful measure for outlier detection, there are several interesting research directions and open problems to be solved. One such direction is the derivation of analytical expressions for the bounds on strong distribution outliers, given by $\mathbb{E}[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|]$. Another direction arises from the simulations in terms of quantifying the gap between the bounds on strong versus weak outliers as a function of $\boldsymbol{\mu}$ and Σ. Finally, we would like to investigate using $\mathbf{x}/\|\mathbf{x}\|$ as a scale-independent metric to identify outliers in distribution.

References

[1] A. S. Hadi, "Identifying multiple outliers in multivariate data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 54, no. 3, pp. 761–771, 1992.

[2] P. Filzmoser, *A multivariate outlier detection method*. na, 2004.

[3] D. M. Rocke and D. L. Woodruff, "Identification of outliers in multivariate data," *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1047–1061, 1996.

[4] D. Yu, G. Sheikholeslami, and A. Zhang, "Findout: finding outliers in very large datasets," *Knowledge and Information Systems*, vol. 4, no. 4, pp. 387–412, 2002.

[5] W. Eberle and L. Holder, "Discovering structural anomalies in graph-based data," in *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*. IEEE, 2007, pp. 393–398.

[6] S. Shekhar, C.-T. Lu, and P. Zhang, "Detecting graph-based spatial outliers," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 451–468, 2002.

[7] H. B. Ahmed, D. Dare, and A.-O. Boudraa, "Graph signals classification using total variation and graph energy informations," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 667–671.

[8] R. S. Tsay, D. Pena, and A. E. Pankratz, "Outliers in multivariate time series," *Biometrika*, vol. 87, no. 4, pp. 789–804, 2000.

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

[10] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1/2, pp. 134–139, 1918.

[11] R. Kan, "From moments of sum to moments of product," *Journal of Multivariate Analysis*, vol. 99, no. 3, pp. 542–554, 2008.

[12] T. Schürmann and I. Hoffmann, "On biased correlation estimation," *arXiv preprint arXiv:1707.09037*, 2017.

[13] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.

[14] I. Olkin, J. W. Pratt *et al.*, "Unbiased estimation of certain correlation coefficients," *The Annals of Mathematical Statistics*, vol. 29, no. 1, pp. 201–211, 1958.

[15] M. Z. Li, K. Gopalakrishnan, K. Pantoja, and H. Balakrishnan, "A spectral approach towards analyzing airport performance and disruptions," *13th Air Traffic Management Research and Development Seminar*, Vienna, Austria. June 2019.