WEAK LAW OF LARGE NUMBERS FOR STATIONARY GRAPH PROCESSES

Fernando Gama and Alejandro Ribeiro

Department of Electrical and Systems Engineering, University of Pennsylvania

ABSTRACT

The ability to obtain accurate estimators from a set of measurements is a key factor in science and engineering. Typically, there is an inherent assumption that the measurements were taken in a sequential order, be it in space or time. However, data is increasingly irregular so this assumption of sequentially obtained measurements no longer holds. By leveraging notions of graph signal processing to account for these irregular domains, we propose an unbiased estimator for the mean of a wide sense stationary graph process based on the diffusion of a single realization. We also provide a bound on the estimation error and determine the conditions for a specific rate of convergence of the estimator to the mean, in a weak law of large numbers fashion.

***Index Terms*—** Graph signal processing, wide sense stationary, estimation theory, unbiased estimator, law of large numbers

1. INTRODUCTION

A basic element within statistical estimation theory is the existence of a sample set of measurements with a certain probabilistic model. In many useful results there is an implicit assumption that the measurements are taken sequentially, which often manifests as the assumption that the samples are drawn from a stationary process. One such important result is the Law of Large Numbers (LLN) which states the convergence of the sample mean to the true mean provided the validity of some simple conditions [1, 2]. In many modern contexts, however, data is supported on irregular domains so that the sequential assumption intrinsic to measurements in time or space no longer holds [3–6]. The theory of graph signal processing (GSP) provides tools to study settings in which the measurements' support is described by a weighted graph that encodes relationships between the signals' components [7–10]. Instrumental in GSP is the notion of a shift operator that is formally defined as a matrix having the same sparsity pattern of the graph – the graph's adjacency matrix, for example. The shift operator is interpreted as a linear operator that can be applied to a graph signal and generalizes the conventional time shifts that are utilized in classical signal processing.

This paper derives a weak LLN for signals supported on graphs (GLLN). In deriving this GLLN we observe that there are two aspects that have to be considered: (i) What is the graph equivalent of performing a time average and (ii) What is the graph equivalent of a stationary signal. Both of these aspects have answers in the existing literature. To answer (i) observe that time averages can be either written as averages of individual signal components or as averages at a fixed point in time of shifted versions of the signal. Since performing averages of individual signal components of a graph signal disregards the information encoded in the graph it is more natural to consider averages at a given node obtained through successive application of the shift operator; see (2). To answer (ii) we resort to the notion of signals that are wide sense stationary with respect to a shift operator [11–13]. These signals satisfy a form of measure invariance with respect to application of the shift operator and are such that it is possible to define a power spectral density that completely characterizes their covariance matrix.

Supported by NSF CCF-1217963. Email: (fgama, aribeiro)@seas.upenn.edu.

We begin the paper with a discussion of wide sense stationary graph signals (Section 2) and the appropriate notion of mean of a stationary graph signal (Section 2.1). To derive the GLLN we consider sequential application of the graph shift operator and focus on the average of the first n signals that result from application of the shift (Section 3). We show that this is an unbiased estimator of the mean save for a normalization constant. The GLLN states that this unbiased estimate is concentrated around the mean as long as the eigenvalues of the graph satisfy some mild properties (Proposition 3). This general result is particularized to directed cycles – where we recover the weak LLN for stationary processes –, and Erdős-Rényi (ER) graphs. Graph whose spectra do not satisfy the conditions in Proposition 3 are discussed in Section 3.1. The paper closes with numerical results (Section 4) and concluding remarks (Section 5).

2. STATIONARY GRAPH SIGNALS

Consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ that encodes the irregular structure that supports a given signal $\mathbf{x}$. That is, let $\mathcal{V}$ be a finite set of n nodes, $\mathcal{E}$ be the set of edges such that $(k, \ell) \in \mathcal{E}$ if and only if node k is connected to node ℓ; and $\mathcal{W} : \mathcal{E} \to \mathbb{R}$ the function that gives different weights to the edges of the network. Define $\mathcal{N}_k = \{\ell \in \{1, \ldots, n\} : (\ell, k) \in \mathcal{E}\}$ as the incoming neighborhood of node k. Let $\mathbf{x} : \mathcal{V} \to \mathbb{R}$ be the graph signal that attaches a certain value to each node. By fixing an arbitrary ordering of the nodes we can view $\mathbf{x}$ as a vector in $\mathbb{R}^n$ where each element represents the value of the signal at each node [7,8]. Also, the graph structure and its impact on the signal can be captured by the graph shift operator $\mathbf{S} \in \mathbb{R}^{n \times n}$ [14]. Matrix $\mathbf{S}$ is such that $[\mathbf{S}]_{k,\ell}$ can be nonzero only if $k = \ell$ or if $(k, \ell) \in \mathcal{E}$. We further assume that $\mathbf{S}$ is a normal matrix, i.e. it satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$, so that it has an eigendecomposition in terms of an orthonormal basis of eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, that is $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}_s\mathbf{V}^H$ where $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ is a unitary matrix and $\mathbf{\Lambda}_s = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalue λ_k associated to the eigenvector $\mathbf{v}_k$, $k = 1, \ldots, n$. Examples of normal graph shift operators are the adjacency matrices of some graphs and the Laplacian of any undirected graph, and their respective normalized counterparts [15]. The eigendecomposition of $\mathbf{S}$ allows for the definition of a Graph Fourier Transform (GFT) given by $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$ and its inverse GFT (iGFT) given by $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$.

In the present paper we focus on signals that can be described as wide sense stationary (WSS) graph processes [12]. This class of signals satisfy that their covariance matrix $\mathbf{C}_x$ is invariant when shifting the signal c steps backward and forward from reference points a and b. Formally, it holds that $\mathbf{S}^a\mathbf{C}_x\mathbf{S}^b = \mathbf{S}^{a+c}\mathbf{C}_x\mathbf{S}^{b-c}$ for any set of non-negative integers a, b, and $c \leq b$. This restriction is equivalent to $\mathbf{C}_x$ and $\mathbf{S}$ being commutative, which implies that they have the same set of eigenvectors. Additionally, the direction of the mean $\boldsymbol{\mu}$ has to remain unchanged after successive applications of the shift. Specifically, $\mathbf{S}\boldsymbol{\mu}$ has to be proportional to $\boldsymbol{\mu}$. These invariance restrictions on both the first and second order moments lead to the following definition.

Definition 1 (WSS graph process)**.** *Let $\mathcal{G}$ be a graph with a graph shift operator $\mathbf{S}$. Let $\mathbf{v}_m$ be an eigenvector of $\mathbf{S}$. The graph signal $\mathbf{x}$ is a*

WSS graph process if it is a random vector with mean $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ *and covariance matrix* $\mathbf{C}_x = \mathbb{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]$ *that satisfy*

(i) $\boldsymbol{\mu} = \mu\,\mathbf{v}_m$,

(ii) $\mathbf{C}_x$ *and* $\mathbf{S}$ *are simultaneously diagonalizable.*

In the case when all eigenvalues of $\mathbf{S}$ are distinct, Def. 1 is equivalent to understanding the WSS graph process $\mathbf{x}$ as being generated from a graph-filtered white sequence $\mathbf{w}$ [12]. Formally, $\mathbf{x} = \mathbf{H}\mathbf{w} + \mu\mathbf{v}_m$ for some graph filter $\mathbf{H} = \sum_{t=0}^{n-1} h_t \mathbf{S}^t$. Then, the covariance matrix can also be written as $\mathbf{C}_x = \mathbf{H}\mathbf{H}^T$. From now on, it is assumed that all eigenvalues are distinct.

The fact that $\mathbf{C}_x$ and $\mathbf{S}$ are simultaneously diagonalizable implies that they have an eigendecomposition by the same unitary matrix, $\mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H$ where $\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}^T$ is a $\mathbb{R}^n$ vector with nonnegative elements which are indeed the eigenvalues of the covariance matrix. The vector $\mathbf{p}$ is defined as the power spectral density (PSD) of the WSS graph process $\mathbf{x}$. Moreover, the process given by the GFT coefficients $\tilde{\mathbf{x}}$ is also WSS, has uncorrelated elements and its covariance matrix is given by $\mathbf{C}_{\tilde{x}} = \text{diag}(\mathbf{p})$ [12].

Remark 1 (Alternative definitions of WSS graph processes)**.** There are two other definitions of WSS graph processes. The first one, found in [11] identifies a *shift* based on an isometrization of the Laplacian, and uses this shift to impose restrictions on the first and second order moment of the process. These restrictions are equivalent to (i) and (ii) in Def. 1. An important observation is that the proposed *shift* is not local since it involves operations between nodes that do not share an edge. The second definition provided in [13] adopts the use of a localization operator based on the graph Laplacian to force conditions on the first and second moment of the graph random process. These conditions are also equivalent to the ones in Def. 1. A major issue with this operator is that it does not conform a mathematical group, making it senseless to repeatedly apply this operator to the signal [16]. Furthermore, both definitions require a Laplacian which implies the need of an undirected graph.

2.1. The concept of mean in a graph process

In traditional signal processing, the mean is associated with the DC (or constant) component of the signal. More specifically, it is understood as the contribution corresponding to the zero-frequency coefficient of the Fourier expansion (slowest time-varying eigenfunction) [17, 18]. In order to characterize the notion of the mean of a random graph signal the concept of total variation (TV) turns out to be useful. TV measures how much the value of the signal at each node changes with respect to the value of the signal at its neighbors. The formal definition is as follows [9].

Definition 2 (Total variation on a graph)**.** *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ *be a graph with* n *vertices and adjacency matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$, $[\mathbf{A}]_{k,\ell} = w_{k,\ell}$, *where* $w_{k,\ell}$ *is the weight corresponding to the edge* $(k,\ell) \in \mathcal{E}$ *by the weight function* $\mathcal{W}$*. Then, the total variation TV*$(\mathbf{x})$ *of a graph signal* $\mathbf{x} \in \mathbb{R}^n$ *on* $\mathcal{G}$ *is defined as*

$$TV(\mathbf{x}) = \sum_{k=1}^{n} \left| x_k - \sum_{\ell\in\mathcal{N}_k} \frac{w_{\ell,k}}{|\lambda_{\max}|} x_\ell \right| = \left\| \mathbf{x} - \frac{1}{|\lambda_{\max}|}\mathbf{A}^T\mathbf{x} \right\|_1 \quad (1)$$

where $\lambda_{\max}$ *is the largest eigenvalue of adjacency matrix* $\mathbf{A}$ *and serves the purpose of normalizing the energy.*

Using TV as a way of measuring the variability of a graph signal allows us to determine which is the slowest graph-varying of all the eigenvectors, and thus assign it to $\mathbf{v}_m$, the eigenvector corresponding to the mean of the WSS process. In this respect, the following proposition comes in handy.

Proposition 1 (Ordering of frequencies)**.** *Let* $\lambda_{\max} = re^{j\theta} \in \mathbb{C}$ *be the eigenvalue of the adjacency matrix* $\mathbf{A}$ *with highest value* $|\lambda_{\max}| = r > 0$, $\mathfrak{Re}\{\lambda_{\max}\} > 0$*. Let* $\lambda_k, \lambda_\ell \in \mathbb{C}$ *be two other eigenvalues. If the eigenvalue* λ_k *is closer in the complex plane to* $|\lambda_{\max}|$ *than* λ_ℓ *is, then TV*$(\mathbf{v}_k) <$ *TV*$(\mathbf{v}_\ell)$ *where* $\mathbf{v}_k$ *is the eigenvector associated to* λ_k *and* $\mathbf{v}_\ell$ *is the eigenvector associated to* λ_ℓ*.*

Proof. See [9], Theorem 2. □

This proposition states that the eigenvalue $\lambda_{\max}$ of the adjacency matrix $\mathbf{A}$ with the highest magnitude has associated an eigenvector with the slowest total variation. From there onwards, the total variation of the eigenvectors increase as the corresponding eigenvalues are located farther away from $|\lambda_{\max}|$. Furthermore, for a connected graph there exists a real nonnegative eigenvalue such that the magnitude of all other eigenvalues does not exceed this one [19, 20]. This implies that the eigenvalue λ_m corresponding to the eigenvector $\mathbf{v}_m$ with the least total variation is real and nonnegative.

Remark 2 (Use of the Laplacian of the graph)**.** A popular choice of graph shift operator in GSP is the Laplacian of a graph $\mathbf{S} = \mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is the diagonal degree matrix and $\mathbf{A}$ is the adjacency. Besides being only useful for undirected graphs, the fact that the Laplacian is a difference operator makes it useless for diffusion processes. More specifically, note that the slowest graph-varying eigenvector of $\mathbf{L}$ is indeed the constant $(1/\sqrt{n})\,\mathbf{1}$, making it *ideal* for adoption as $\mathbf{v}_m$. However, this eigenvector corresponds to the zero eigenvalue and therefore all successive operations of the shift do not further contribute, in mean sense, to the estimation, $\mathbb{E}[\mathbf{S}^t\mathbf{x}] = \mathbf{S}^t\boldsymbol{\mu} = (\mu/\sqrt{n})\mathbf{S}^t\mathbf{1} = \mathbf{0}$ for all $t = 1, 2, \ldots, n-1$.

From now on, it is assumed that the graph shift operator under consideration is the adjacency matrix $\mathbf{S} = \mathbf{A}$, that the eigenvalues are all distinct $\lambda_k \neq \lambda_\ell$ whenever $k \neq \ell$, that they ordered as $\lambda_1 \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$ and that the eigenvector associated to the mean of the WSS graph process is the one with lowest total variation $\mathbf{v}_m = \mathbf{v}_1$, corresponding to $\lambda_m = \lambda_1 \in \mathbb{R}$.

3. WEAK LAW OF LARGE NUMBERS

We now study the use of the aggregated diffusion process $\mathbf{y}_n = \sum_{t=0}^{n-1} \mathbf{S}^t\mathbf{x}$ as an estimator for the mean of the WSS process $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} = \mu\mathbf{v}_1$. With access to the aggregated diffusion process $\mathbf{y}_n$ we can obtain an unbiased estimator as

$$\hat{\boldsymbol{\mu}}_n = \frac{1}{\sum_{t=0}^{n-1}\lambda_1^t} \sum_{t=0}^{n-1} \mathbf{S}^t\mathbf{x} = \frac{1}{\sum_{t=0}^{n-1}\lambda_1^t}\mathbf{y}_n. \quad (2)$$

Proposition 2 (Unbiased estimator)**.** *Given a WSS graph process* $\mathbf{x}$ *with mean* $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} = \mu\mathbf{v}_1$*, then the estimator* $\hat{\boldsymbol{\mu}}_n$ *given in* (2) *is an unbiased diffusion estimator for the mean of the process.*

Proof. Simply calculate the expectation of $\hat{\boldsymbol{\mu}}_n$ and recall that $\mathbf{S}\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ because $\mathbf{v}_1$ is an eigenvector of $\mathbf{S}$. □

The unbiased diffusion estimator (2) is also a WSS process since it results from filtering a WSS process [12]. Then, its PSD $\mathbf{q}$ is

$$q_k = p_k \frac{|\sum_{t=0}^{n-1}\lambda_k^t|^2}{|\sum_{t=0}^{n-1}\lambda_1^t|^2}, \quad k = 1, \ldots, n \quad (3)$$

which leads to a covariance matrix given by $\mathbf{C}_{\hat{\mu}_n} = \mathbf{V}\text{diag}(\mathbf{q})\mathbf{V}^H$. Observe that for $k = 1$, that is, for the eigenvalue corresponding to the mean eigenvector, we always have a PSD $q_1 = p_1$.

With the covariance matrix, performance bounds on the estimation error can be obtained. In particular, if we look at the value of the estimator in a specifc node ℓ we get the following Lemma.

Lemma 1 (Error bound)**.** *Let $\mathbf{x}$ be a WSS graph process on a graph $\mathcal{G}$ characterized by a normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}_s\mathbf{V}^H$, $\mathbf{\Lambda}_s = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Let $\mathbb{E}[\mathbf{x}] = \mu\mathbf{v}_1$ be the mean of the process and $\mathbf{C}_x = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^H$ the covariance matrix, with $p_k < \infty$ for all $k = 1, \ldots, n$. Then,*

$$\mathbb{P}\left(\left|[\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}]_\ell\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} q_k |v_{\ell,k}|^2. \tag{4}$$

Proof. Let $\mathbf{e}_\ell$ be the vector selecting node ℓ, that is a vector with the value 1 in position ℓ and zeros elsewhere. Observe that $\mathbb{E}[\mathbf{e}_\ell^H \hat{\boldsymbol{\mu}}_n] = \mu_\ell$ and $\text{var}\left[\mathbf{e}_\ell^H(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu})\right] = \mathbf{e}_\ell^H \mathbf{C}_{\hat{\mu}_n}\mathbf{e}_\ell = \mathbf{e}_\ell^H \mathbf{V}\text{diag}(\mathbf{q})\mathbf{V}^H\mathbf{e}_\ell < \infty$ since $p_k < \infty$ for all $k = 1, \ldots, n$. Then apply Chebyshev's inequality [2, Theorem 1.6.4] to get (4). □

The bound given in Lemma 1 is reminiscent of the basic inequality needed to obtain the classical weak law of large numbers [2, Theorem 2.2.3]. This intuition comes from the fact if we let $\mathbf{x}$ be comprised of i.i.d. random variables, then $p_k = 0$ for all $k \neq 1$ making only q_1 survive in (4). Thus, only $|v_{\ell,k}|^2$ remains and this value typically has a dependence on $1/n$. See the directed cycle case later in this section. Also, observe that $\{v_{\ell,k}, k = 1, \ldots, n\}$ are the values contained in the ℓ-th row of $\mathbf{V}$. Interpreting this in the context of the GFT, we note that the performance of the estimation at a single node ℓ depends on the variance of all nodes and on all frequency coefficients (all eigenvalues) of the graph shift operator, since the signal is passed over all the nodes. Additionally, it is affected by the value of each frequency component (each eigenvector) on node ℓ alone.

To push the analogy forward, we observe that (4) depends on both q_k and $\mathbf{v}_k$. Then, we first proceed to study the behavior of q_k, which depends on the spectrum of the graph as specified by the following lemma.

Lemma 2 (Behavior of q_k)**.** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ be a weighted graph that admits a normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}_s\mathbf{V}^H$. Let $\lambda_1 \in \mathbb{R}$ be the largest positive eigenvalue such that $|\lambda_k| \leq \lambda_1$ for all $k = 2, \ldots, n$. If $\lambda_1 > 1$ and $|\lambda_k|/\lambda_1 = o(n^{-\delta/2n})$ for some $\delta > 0$, or if $\lambda_1 = 1$, then*

$$q_k = o(n^{-\delta}) \quad , \; k = 2, \ldots, n \tag{5}$$

For $k = 1$ we always have $q_1 = p_1$.

Proof. Let $\lambda_k = re^{j\theta}$. Assume that $\lambda_1 > 1$. Then, using the geometric sum on (3)

$$q_k = p_k \frac{|1-\lambda_1|^2}{|1-\lambda_k|^2}\frac{|1-\lambda_k^n|^2}{|1-\lambda_1^n|^2} \leq p_k \frac{|1-\lambda_1|^2}{|1-r|^2}\frac{1+r^{2n}}{|1-\lambda_1^n|^2}$$

First observe that, since $r \leq \lambda_1$, then it always holds that $|1-\lambda_1|^2/|1-r|^2 \leq 1$ so that the term is always bounded. Now, because $\lambda_1 > 1$ then $|1-\lambda_1^n|^2 = \mathcal{O}(\lambda_1^{2n})$. Likewise, if $r > 1$ then $1 + r^{2n} = \mathcal{O}(r^{2n})$ so that

$$\frac{1+r^{2n}}{|1-\lambda_1^n|^2} = \mathcal{O}\left(\frac{|\lambda_k|^{2n}}{\lambda_1^{2n}}\right) = o(n^{-\delta}).$$

If $r \leq 1$, then

$$\frac{1+r^{2n}}{|1-\lambda_1^n|^2} = \mathcal{O}\left(\frac{1}{\lambda_1^{2n}}\right) = o(n^{-\delta})$$

For $\lambda_1 = 1$, we have that $|\sum_{t=0}^{n-1}\lambda_1^t|^2 = n^2$ so that

$$q_k = \frac{p_k}{n^2}\left|\sum_{t=0}^{n-1}\lambda_k^t\right|^2 = \frac{p_k}{n^2}\frac{|1-r^n e^{jn\theta}|^2}{|1-\lambda_k|^2}$$

For $r < \lambda_1 = 1$ we have that $|1 - r^n e^{jn\theta}|^2 = \mathcal{O}(1)$. If $r = o(1)$, then $|1 - re^{j\theta}|^2 = \mathcal{O}(1)$ and the $1/n^2$ guarantees that $q_k = o(1/n)$. If $r = 1$ then $|1 - e^{jn\theta}|^2/|1 - e^{j\theta}|^2 = o(1/n)$ making $q_k = o(1/n)$ completing the proof. □

Lemma 2 states that, under certain conditions on the spectrum of the graph, all terms are polynomially decreasing except for $q_1 = p_1$; it is shown later on that these conditions hold for some practical graphs such as the directed cycle and Erdős-Rényi graphs. We can also obtain a rate of convergence for the probability of error that is evocative of the classical weak law of large numbers.

Proposition 3 (Weak law of large numbers)**.** *Let $\mathbf{x}$ be a WSS graph process on a graph $\mathcal{G}$ characterized by a normal graph shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}_s\mathbf{V}^H$ with eigenvalues that satisfy the conditions on Lemma 2. Then, if $p_k < \infty$ for all k, we have that*

$$\min_{\ell=1,\ldots,n} \mathbb{P}\left(\left|[\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}]_\ell\right| > \epsilon\right) \leq \frac{p_1}{n\epsilon^2} + o(n^{-\delta}). \tag{6}$$

Proof. First, observe that because $\|\mathbf{v}_1\|^2 = 1$ there is always a node ℓ for which $|v_{\ell,1}| \leq 1/\sqrt{n}$ so that

$$\min_{\ell=1,\ldots,n} |v_{\ell,1}|^2 \leq \frac{1}{n}.$$

Note that if the eigenvectors are highly localized, then there exists a node ℓ for which $v_{\ell,1} = 0$. Let $q_{\max} = \max_{k=2,\ldots,n}\{q_k\}$. Then, from (4) we have that

$$\min_{\ell=1,\ldots,n} \mathbb{P}\left(\left|[\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}]_\ell\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\left(\frac{p_1}{n} + q_{\max}\right) \tag{7}$$

where the fact that $\sum_{k=2}^{n}|v_{\ell,k}|^2 \leq 1$ for all ℓ was used (because the rows of $\mathbf{V}$ also form an orthonormal basis). Now because the eigenvalues of $\mathbf{S}$ satisfiy the assumptions of Lemma 2 by hypothesis and because $q_{\max}$ is some value of q_k for $k = 2, \ldots, n$, then we know that $q_{\max} = o(n^{-\delta})$ thus completing the proof. □

Proposition 3 is remindful of the traditional WLLN because it states that the probability that the average of shifts is far from the mean decreases polynomially as the size of the graph increases. Alternatively, if we make $\epsilon = 1/\sqrt{n}$ we obtain a concentration inequality by which the average of shifts gets arbitrarily close to the mean at a rate of $1/\sqrt{n}$ with constant probability.

In what follows we consider the application of Proposition 3 to different well-known graphs.

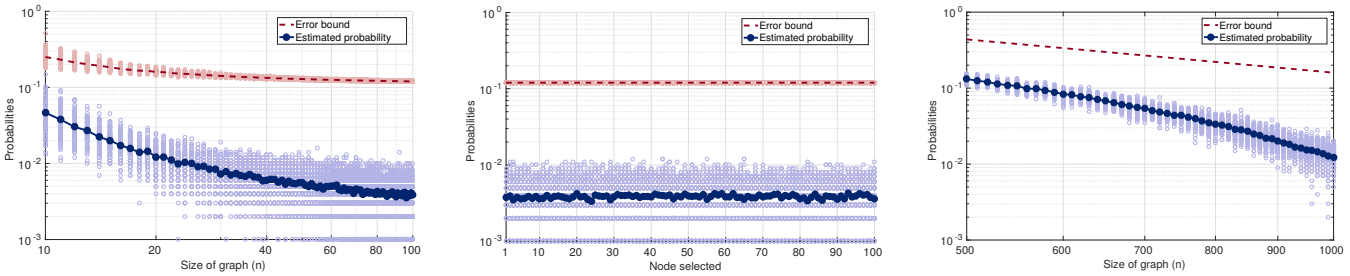
Directed cycle (Classical WLLN). The directed cycle $\mathcal{G}_{dc}$ represents the graph support for time-stationary signals. Then, by applying Proposition 3 we expect to recover the traditional WLLN.

Corollary 1 (Convergence of Directed Cycle)**.** *Let $\mathcal{G}_{dc}$ be the directed cycle graph. Then, for any node ℓ the error bound is*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} x_k - \mu\right| > \epsilon\right) \leq \frac{p_1}{n\epsilon^2}. \tag{8}$$

Proof. Note that $\lambda_1 = 1$ so that Proposition 3 holds. More specifically, $q_k = 0$ for all $k \neq 1$ and $v_{\ell,1} = 1/\sqrt{n}$ for all ℓ so that $\boldsymbol{\mu} = \mu\mathbf{1}$ is the constant vector. Finally, $\sum_{t=0}^{n-1}\mathbf{S}^t\mathbf{x} = \sum_{k=1}^{n} x_k\mathbf{1}$ because after n shifts the values of the signal have been aggregated at all nodes due to the nature of the directed cycle. □

Corollary 1 is a statement of the weak LLN for signals that are stationary in time. The result in this case is stronger than the one in (6) because it lacks the order term $o(n^{-\delta})$. This term vanishes because in the case of a cycle graph the kth component of the estimator's PSD is $q_k = 0$, $k = 2, \ldots, n$. It is also stronger in that the minimum disappeared since, after n shifts, all nodes have aggregated the sample mean, so any node yields the same estimator and thus the same probability of error. Finally, note that if $\mathbf{x}$ are i.i.d. r.v. then the DC component of the signal is $p_1 = \sigma^2$ and (8) is the Chebyshev's bound that leads to the classical WLLN.



(a) (b) (c)

Fig. 1: Estimated probability and error bounds. (a) Probability of error for the ER graph as a function of the size of the graph; the larger the graph, the less probability of error. (b) Probability of error for the ER graph as a function of the node selected ℓ; every node yields the same estimation error. (c) Probability of error for the path graph as a function of the size.

Erdős-Rényi (ER) graphs. Another family of graphs that satisfies Proposition 3 are ER graphs. These graphs have the particularity that the largest eigenvalue grows linearly with the size whereas the rest of the eigenvalues have a growth rate that does not exceed $\sqrt{n}$. This means that the graph is well-suited for estimation since the PSD of the estimator concentrates around the largest eigenvalue corresponding to the mean of the process. This shows in the proof of the following corollary.

Corollary 2 (WLLN for Erdős-Rényi graphs)**.** *Let $\mathcal{G}_{ER}$ be an ER graph of size n with p such that $pn \to \alpha \geq 1$. Then, for any node ℓ and any $0 < \delta < n$ we have that*

$$\mathbb{P}\left(\left|[\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}]_\ell\right| > \epsilon\right) \leq \frac{p_1}{n\epsilon^2} + o(n^{-\delta}). \tag{9}$$

Proof. First, note that $\lambda_1 = pn + o(n)$ and that by the semi-circle law, with probability $1 - o(1)$ all eigenvalues except the largest one lie in the interval $(-c\sqrt{n}, c\sqrt{n})$ for any $c > 2p(1-p)$ [21, 22]. Then, we have that $\lambda_2 \leq c\sqrt{n}$ so that $\lambda_2/\lambda_1 = o(n^{-\delta/2n})$ for any $0 < \delta < n$, satisfying Proposition 3. Additionally, because $\sqrt{n}v_{\ell,1} = 1 + o(n^{-(1/2-r)})$, $0 < r < 1/2$ with probability $1 - o(1)$, see [23], then any node ℓ yields similar probability of error. □

Corollary 2 states that the estimator obtained at any node is arbitrarily close to the true mean at that node, with a convergence rate that is polynomial on the size of the graph.

3.1. Non-convergent graphs

In cases where the graph spectra does not fall under the conditions of Lemma 2 and thus Proposition 3 cannot be applied, there are still interesting analyses to make.

Lemma 3. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ be a weighted graph that admits a normal graph shift operator $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}_s\mathbf{V}^H$. Let $\lambda_1 \in \mathbb{R}$ be the largest positive eigenvalue such that $|\lambda_k| \leq \lambda_1$ for all $k = 2, \ldots, n$. Let $\mathcal{K}$ be the set of indices such that $|\lambda_k|/\lambda_1$ does not satisfy $o(n^{-\delta/2n})$ for any $\delta > 0$. If $\lambda_1 > 1$ and $\mathcal{K}$ is nonempty, or if $\lambda_1 < 1$, then*

$$\mathbb{P}\left(\left|[\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}]_\ell\right| > \epsilon\right) \leq \frac{p_1}{\epsilon^2}|v_{\ell,1}|^2 + o(1) + \sum_{k\in\mathcal{K}} p_k \left|\frac{1-\lambda_1}{1-\lambda_k}\right|^2 |v_{\ell,k}|^2(1+o(1)). \tag{10}$$

If $\lambda_1 < 1$, then $\mathcal{K} = \{2, \ldots, n\}$.

Proof. In analogy with the proof of Lemma 2 we prove that for the conditions of $\lambda_1 < 1$ for which $\mathcal{K} = \{2, \ldots, n\}$ or for the case when $\lambda_1 > 1$ and $\mathcal{K}$ is nonempty, then

$$q_k = p_k \left|\frac{1-\lambda_1}{1-\lambda_k}\right|^2 (1 + o(1)) \tag{11}$$

First, let $k \in \mathcal{K}$ and $\lambda_k = re^{j\theta}$ with $\lambda_1 > 1$. Then, since r/λ_1 does not decrease any faster than $n^{-\delta/2n}$, we have

$$\frac{|1-\lambda_k^n|^2}{|1-\lambda_1^n|^2} = \frac{1 - 2r^n\cos(n\theta) + r^{2n}}{1 - 2\lambda_1^n + \lambda_1^{2n}} = 1 + o(1).$$

For $\lambda_1 < 1$ we have that, since $r/\lambda_1 \leq 1$ and $r^n = o(1)$ and $\lambda_1^n = o(1)$, then $|1 - r^n e^{j\theta n}|^2 = 1 + o(1)$ and $|1 - \lambda_1^n|^2 = 1 + o(1)$, completing the proof. □

For the case in which $\lambda_1 < 1$ we reach a fundamental limit under which is not possible to achieve a better estimation. This situation occurs because on each successive step of the diffusion process the information harnessed from neighboring nodes is less and less (because all the eigenvalues are less than 1), eventually making it impossible to accurately estimate the mean. Alternatively, when $\lambda_1 > 1$, if $|\mathcal{K}| = o(n)$ then then the estimator is still consistent since at most finitely many values of q_k do not follow the rate $o(n^{-\delta})$. When $|\mathcal{K}| = \mathcal{O}(n)$, then no assertions about the convergence of the estimator can be done.

4. SIMULATIONS

First, we consider an ER graph with n nodes and probability $p(n) = 3/n$ as support for a WSS graph process, for which realizations are obtained. The mean is equal to the eigenvector associated to the largest eigenvalue. An estimate of the probability of error is obtained by averaging 1000 realizations. First, the size n varies from 10 to 100. For each n, 100 different graphs are used and the error probability on all of them is averaged. It can be observed from Fig 1a that both the bound and the error probability (7) decrease as n increases. For the second simulation we fix $n = 100$ and $p = 0.03$ and we analyze the bound as a function of the node ℓ selected, see Fig 1b. As discussed earlier, all nodes tend to be equal in their estimation of the mean.

We also consider a second simulation of the same WSS process described before but now supported on a path graph. In this case $\lambda_k = 2\cos(\pi k/(n+1))$ so that Proposition 3 is not satisfied. Also $\mathcal{K} = \mathcal{O}(n)$ so that no assertions about the bound can be made. Yet, the simulation results in Fig. 1c show that the probability of error still decreases with the size of the graph. This is indicative that there is room for obtaining better bounds or less stringent conditions on which a convergence rate is achieved. This will be our focus in future works.

5. CONCLUSIONS

In the present paper, an unbiased estimator for the mean of a WSS graph process based on the diffusion of a single realization was presented. Error bounds were obtained and convergence rates for the error probability were provided. These results are reminiscent of the weak law of large numbers. Specific expressions for the directed cycle and the Erdős-Rényi graphs were obtained. Also, situations in which the said convergence rate is not achieved were discussed.

6. REFERENCES

[1] P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics: An Introduction with Applications*, Springer Science+Business Media, Dordrecht, Netherlands, 1993.

[2] R. Durrett, *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, NY, 4th edition, 2010.

[3] C. Guestrin, P. Bodik, R. Thibaux, M. Paskin, and S. Madden, "Distributed regression: an efficient framework for modeling sensor network data," in *3rd ACM/IEEE Int. Conf. Inform. Process. Sensor Networks*, Berkeley, CA, 26-27 Apr. 2004, ACM/IEEE.

[4] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmonic Anal.*, vol. 21, no. 1, pp. 53–94, July 2006.

[5] R. Wagner, V. Delouille, and R. Baraniuk, "Distributed wavelet de-noising for sensor networks," in *45th IEEE Conf. Decision and Control*, San Diego, CA, 13-15 Dec. 2006, IEEE.

[6] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *37th IEEE Int. Conf. Acoust., Speech and Signal Process.*, Kyoto, Japan, 25-30 March 2012, IEEE.

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

[8] 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.

[9] A. Sandyhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[10] A. Sandryhaila and J. M. F. 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, Sep. 2014.

[11] B. Girault, P. Gonçalves, and E. Fleury, "Translation and Stationarity for Graph Signals," Research Report RR-8719, École Normale Supérieure de Lyon, Inria Rhône-Alpes, Apr. 2015.

[12] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *arXiv:1603.04667v1 [cs.SY]*, 14 March 2016.

[13] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *arXiv:1601.02522v4 [cs.DS]*, 8 July 2016.

[14] M. Püschel and J. M. F. Moura, "Algebraic signal processing theory: Foundation and 1-d time," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3575–3585, Aug. 2008.

[15] F. R. K. Chung, *Spectral Graph Theory*, Number 92 in Conf. Board Math. Sci.: Regional Conf. Series Math. Amer. Math. Soc., Providence, RI, 1997.

[16] D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Appl. Comput. Harmonic Anal.*, vol. 40, no. 2, pp. 260–291, March 2016.

[17] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1996.

[18] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Pearson, Upper Saddle River, NJ, 3rd edition, 2010.

[19] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Application*, Academic Press, New York, NY, 1979.

[20] F. R. Gantmacher, *The Theory of Matrices*, vol. 2, AMS Chelsea Publishing, Providence, RI, 1989, Translated by K. A. Hirsch.

[21] Z. Füredi and J. Komlós, "The eigenvalues of random symmetric matrices," *Combinatorica*, vol. 1, no. 3, pp. 233–241, March 1981.

[22] M. Krivelevich and B. Sudakov, "The largest eigenvalue of sparse random graphs," *arXiv:math/0106066v1 [math.CO]*, 10 June 2001.

[23] P. Mitra, "Entrywise bounds for eigenvectors of random graphs," *Electron. J. of Combinatorics*, vol. 16, pp. 1–18, Oct. 2009.