

# A Spectral Graph Uncertainty Principle

Ameya Agaskar, *Student Member, IEEE*, and Yue M. Lu, *Senior Member, IEEE*

**Abstract**—The spectral theory of graphs provides a bridge between classical signal processing and the nascent field of graph signal processing. In this paper, a spectral graph analogy to Heisenberg’s celebrated uncertainty principle is developed. Just as the classical result provides a tradeoff between signal localization in time and frequency, this result provides a fundamental tradeoff between a signal’s localization on a graph and in its spectral domain. Using the eigenvectors of the graph Laplacian as a surrogate Fourier basis, quantitative definitions of graph and spectral “spreads” are given, and a complete characterization of the feasibility region of these two quantities is developed. In particular, the lower boundary of the region, referred to as the uncertainty curve, is shown to be achieved by eigenvectors associated with the smallest eigenvalues of an affine family of matrices. The convexity of the uncertainty curve allows it to be found to within  $\varepsilon$  by a fast approximation algorithm requiring  $\mathcal{O}(\varepsilon^{-1/2})$  typically sparse eigenvalue evaluations. Closed-form expressions for the uncertainty curves for some special classes of graphs are derived, and an accurate analytical approximation for the expected uncertainty curve of Erdős-Rényi random graphs is developed. These theoretical results are validated by numerical experiments, which also reveal an intriguing connection between diffusion processes on graphs and the uncertainty bounds.

**Index Terms**—Diffusion on graphs, Fourier transforms on graphs, graph Laplacians, signal processing on graphs, spectral graph theory, uncertainty principles, wavelets on graphs.

## I. INTRODUCTION

HEISENBERG’S uncertainty principle is a cornerstone of signal processing. The simple inequality [1], [2]

$$\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4} \quad (1)$$

in which  $\Delta_t^2$  and  $\Delta_\omega^2$  measure the “time spread” and “frequency spread” of some signal, respectively, is one way to precisely characterize a general principle with far-reaching

consequences: that a signal cannot be concentrated in both time and frequency.

In this paper, we establish analogous uncertainty principles for signals defined on graphs. The study of signals on graphs, and the extension of classical signal processing techniques to such nonstandard domains, has received growing interest in the past few years (see, e.g., [3]–[11]). These studies are often motivated (and enabled) by the deluge of modern data collected on various technological, social, biological, and informational networks [12]. The efficient acquisition, representation, and analysis of such high-dimensional graph-based data present challenges that should be addressed by the development of new signal processing theories and tools.

## A. Related Work

Uncertainty principles date back to Heisenberg [1], who in 1927 proved a result that Weyl and Pauli soon afterward generalized to (1). It was also shown that the bound in (1) is achievable by Gaussian-shaped functions and frequency modulations thereof. A lifetime later, analogous results were found for discrete-time signals as well [13], [14]. Similar uncertainty principles have also been established on the unit sphere  $S^d$  [15] and, in more abstract settings, on compact Riemannian manifolds [16].

In a different line of work, Donoho and Stark [17] introduced a new concept of uncertainty related to signal support size. They showed that a length  $N$  discrete-time signal with support set  $\mathcal{T}$  in the time domain and support set  $\mathcal{W}$  in the frequency domain satisfies  $|\mathcal{T}| |\mathcal{W}| \geq N$ . This bound is a *nonlocal* uncertainty principle—it limits the cardinality of a signal’s time and frequency support sets, even if each is the disjoint union of far-flung subsets. Further studied in, e.g., [18]–[20], these non-local uncertainty principles laid the foundation for sparse signal recovery from partial measurements.

In the same vein of the classical (and local) uncertainty principle stated in (1), we have been studying the following question: given an arbitrary graph, to what extent can a signal be simultaneously localized on that graph and in the “frequency” domain? To obtain the spectral representation of these signals, we use the standard approach of treating the eigenvectors of the graph Laplacian operator [21] as a Fourier basis. The Laplacian encodes a notion of smoothness on a graph [22] and is analogous to the Laplace–Beltrami operator on a manifold [23].

The analogy between the spectral decomposition of graph Laplacians and the standard Fourier transform has been used to extend the concept of bandlimited sampling to signals defined on graphs [9] and in the construction of wavelet transforms on graphs [4], [6], [10]. In the latter case, as pointed out in [10], a desirable property of the wavelet transforms is that the dictionary elements (i.e., wavelets) should be well localized in the graph and spectral domains. Our results provide a way to

Manuscript received May 15, 2012; revised December 19, 2012; accepted February 20, 2013. Date of publication March 12, 2013; date of current version June 12, 2013. This work was supported in part by the Department of the Air Force under Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government. This paper was presented in part at the 2011 SPIE Conference on Wavelets and Sparsity and in part at the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing.

A. Agaskar is with the Signals, Information, and Networks Group, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA, and also with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02421 USA (e-mail: aagaskar@seas.harvard.edu).

Y. M. Lu is with the Signals, Information, and Networks Group, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: yuelu@seas.harvard.edu).

Communicated by D. P. Palomar, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2013.2252233

precisely quantify this desideratum, as well as its fundamental limit.

### B. Contributions and Paper Organization

We begin in Section II with a review of some basic concepts in graph theory, including the definition of the graph Laplacian matrix and its spectral decomposition. After justifying the use of the Laplacian eigenvectors as a Fourier basis on graphs, we define in Section II-C the *graph spread* about a vertex  $u_0$ ,  $\Delta_{g,u_0}^2(\mathbf{x})$ , and the *spectral spread*,  $\Delta_s^2(\mathbf{x})$ , of a signal  $\mathbf{x}$  defined on a graph. These two quantities, which we first introduced in some preliminary work [24], [25], are defined in analogy to the standard time and frequency spreads, respectively.

In [24], we developed a lower bound on the product of  $\Delta_{g,u_0}^2$  and  $\Delta_s^2$  analogous to (1). However, the bound was not tight and applied only under restrictive conditions for the graph and the signal on it. In [25], we took a new approach to characterize a more general and precise relationship between the two kinds of uncertainty. In this paper, we continue this line of investigation, and provide a rigorous basis for the arguments presented in [25], in addition to some new results.

The main contributions of this paper are developed in Section III, where we characterize the uncertainty bound, in Section IV, where we analyze the bound when applied to special families of graphs, and in Section V, where we reveal a connection between diffusion processes and the uncertainty bound. The main results are summarized as follows.

- 1) *Convexity of the feasible region*: We prove that when the underlying graph is connected and contains at least three vertices, the feasibility region of all possible pairs  $(\Delta_s^2(\mathbf{x}), \Delta_{g,u_0}^2(\mathbf{x}))$  is a *bounded* and *convex* set. The feasibility region's convexity was stated without proof in [25].
- 2) *Characterization of the uncertainty curve*: We provide a complete characterization of the curve

$$\gamma_{u_0}(s) = \min_{\mathbf{x}} \Delta_{g,u_0}^2(\mathbf{x}) \text{ subject to } \Delta_s^2(\mathbf{x}) = s$$

which forms the lower boundary of the feasibility region. Studying  $\gamma_{u_0}(s)$ , which we will refer to as the *uncertainty curve*, is important because it is a fundamental bound analogous to the classical uncertainty bound (1). Theorem 1 states that each point on the uncertainty curve is achieved by an eigenvector associated with the smallest eigenvalue of a particular matrix-valued function  $\mathbf{M}(\alpha)$ . Varying the parameter  $\alpha$  allows one to “trace” and obtain the entire curve  $\gamma_{u_0}(s)$ . A rigorous and complete proof of Theorem 1 is provided in this paper, extending the rough argument given in [25]. Based on the convexity of  $\gamma_{u_0}(s)$ , we show in Section III-C that the *sandwich algorithm* [26] can be used to efficiently produce a piecewise linear approximation for the uncertainty curve that differs from the true curve by at most  $\varepsilon$  (under a suitable error metric) and requires solving  $\mathcal{O}(\varepsilon^{-1/2})$  typically sparse eigenvalue problems.

- 3) *Special graph families*: The uncertainty curves for several special families of graphs are investigated in Section IV. For complete graphs and star graphs, we

derive closed-form formulas for the uncertainty curves  $\gamma_{u_0}(s)$ . For Erdős–Rényi random graphs [27], [28], we develop an analytical approximation for the expected value of  $\gamma_{u_0}(s)$ , which is shown through experiment to be very accurate.

- 4) *Diffusion process on a graph*: In Section V, we reveal an intriguing connection between the classical uncertainty principle for functions on the real line and our results for signals on graphs. In the classical case, the solution to the heat equation  $\frac{du}{dt} = \frac{d^2u}{dy^2}$  starting at  $t = 0$  as an impulse is a Gaussian function with a variance that grows linearly with  $t$ ; this solution achieves the Heisenberg uncertainty bound (1). We first show experimental results indicating that a diffusion process starting with an impulse on a graph follows the graph uncertainty curve very closely (though not, in general, exactly.) We then prove in Proposition 4 that the match is exact for the special cases of a star graph or a complete graph. We further prove in Proposition 5 that for general graphs, under a simple condition on the distance function on the graph, the first two derivatives of the uncertainty curve and the curve traced by the diffusion process match at the point corresponding to  $t = 0$ . We conclude the paper in Section VI.

## II. MATHEMATICAL FORMULATION

### A. Graphs, Signals, and Notation

We define a simple, undirected graph as  $G = (V, E)$ , where  $V = \{v_1, v_2, \dots, v_N\}$  is a set of  $N$  vertices and  $E = \{e_1, e_2, \dots, e_M\}$  is the set of  $M$  edges. Each edge is an unordered pair of two different vertices  $u, v \in V$ , and we will use the notation  $u \sim v$  to indicate that  $u$  and  $v$  are connected by an edge. The fundamental structure of a graph  $G$  can be captured by its *adjacency matrix*  $\mathbf{A} = [a_{ij}]_{ij}$ , where  $a_{ij} = 1$  if there is an edge between  $v_i$  and  $v_j$ , and  $a_{ij} = 0$  otherwise. As defined, the diagonal of  $\mathbf{A}$  is always zero because a simple graph may contain no loops (i.e., edges connecting one vertex to itself), and  $\mathbf{A}$  is symmetric because the graph is undirected. (A common generalization is to consider a weighted graph, where each edge  $e_m$  ( $1 \leq m \leq M$ ) is associated with a positive “weight”  $w_m$ . However, in this paper, we only consider unweighted graphs.)

The degree of a vertex  $v$ , denoted by  $\deg(v)$ , is the number of edges incident upon that vertex. We define  $\mathbf{D}$  as the diagonal matrix that has the vertex degrees on its diagonal, i.e.,

$$\mathbf{D} \stackrel{\text{def}}{=} \text{diag} \{ \deg(v_1), \deg(v_2), \dots, \deg(v_N) \}. \quad (2)$$

To quantify the graph-domain spread of a signal, we will need a notion of distance, denoted by  $d(u, v)$ , between any pair of vertices  $u$  and  $v$  on the graph. A simple choice is to use the *geodesic distance* [21], in which case  $d(u, v)$  is the length of the shortest path connecting the two vertices. In this work, we only consider connected graphs, so  $d(u, v)$  is always finite. Other distance functions have been proposed in the literature, including the resistance distance [29] and the diffusion distance [4]. Our subsequent discussions are not confined to any particular choice of the distance function. The only requirement is that  $d(u, v)$

should form a semimetric: namely,  $d(u, v) \geq 0$  with equality if and only if  $u = v$ , and  $d(u, v) = d(v, u)$ .

A finite-energy signal defined on the graph  $\mathbf{x} \in \ell^2(G)$  is a mapping from the set of vertices to  $\mathbb{R}$ . It can be treated as a vector in  $\mathbb{R}^N$ , and so any such signal will be denoted by a boldface variable. There is a natural inner product on  $\ell^2(G)$  defined by  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x}$ , which induces a norm  $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$ . We will denote the value of  $\mathbf{x}$  at vertex  $v$  by  $x(v)$ . An impulse at  $v \in V$ , i.e., a signal that has value 1 at  $v$  and 0 everywhere else, will be denoted as  $\delta_v$ .

### B. Laplacian Matrix and Graph Fourier Transforms

As mentioned in Section I, the graph Laplacian matrix plays an important role in this work. There are several different definitions of the Laplacian matrix commonly used in the literature. The unnormalized Laplacian matrix [21] is given by  $\mathbf{L}_{\text{unnorm}} \stackrel{\text{def}}{=} \mathbf{D} - \mathbf{A}$ , where  $\mathbf{D}$  and  $\mathbf{A}$  are the degree matrix in (2) and the adjacency matrix, respectively. In this paper, we find it more convenient to use the normalized Laplacian matrix [30], defined as

$$\mathbf{L}_{\text{norm}} \stackrel{\text{def}}{=} \mathbf{D}^{-1/2} \mathbf{L}_{\text{unnorm}} \mathbf{D}^{-1/2} \\ = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}.$$

The choice of unnormalized or normalized Laplacian makes no essential difference to our analysis in Section III. The latter is chosen because it leads to simpler expressions in some of our derivations. For notational simplicity, we will drop the subscript in  $\mathbf{L}_{\text{norm}}$ , calling it  $\mathbf{L}$  in what follows.

Intuitively, the Laplacian matrix is analogous to the continuous Laplacian operator  $-\nabla^2$  or  $-\frac{d^2}{dy^2}$  on the real line. In fact, when the underlying graph is a line or a cycle,  $\mathbf{L}$  provides the standard stencil approximation for the second-order differentiation operator. The same holds for higher dimensional lattices. In more general settings where the graphs are formed by sampling an underlying continuous manifold, the Laplacian matrix converges at high sampling densities to the Laplace–Beltrami operator, a differential geometric analogy to the second derivative [23].

By construction,  $\mathbf{L}$  is a real symmetric matrix. We can therefore diagonalize  $\mathbf{L}$  as

$$\mathbf{L} = \mathbf{F} \mathbf{\Lambda} \mathbf{F}^T \quad (3)$$

where  $\mathbf{F}$  is an orthogonal matrix whose columns are the eigenvectors of  $\mathbf{L}$ , and  $\mathbf{\Lambda} \stackrel{\text{def}}{=} \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_N \}$  is a diagonal matrix of eigenvalues, which are all real.  $\mathbf{L}$  can be shown to be positive semidefinite with rank less than  $N$ , so we can order the eigenvalues as  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ .

A large number of the topological properties of a graph can be inferred from the spectrum of its graph Laplacian [30]. For example, a graph is connected (meaning that a path can always be found connecting one vertex to the other) if and only if the smallest eigenvalue ( $\lambda_1 = 0$ ) has multiplicity one. The corresponding unit-norm eigenvector  $\mathbf{f}_1$  is defined by

$$f_1(v) = \sqrt{\frac{\deg(v)}{\sum_{u \in V} \deg(u)}} \quad (4)$$

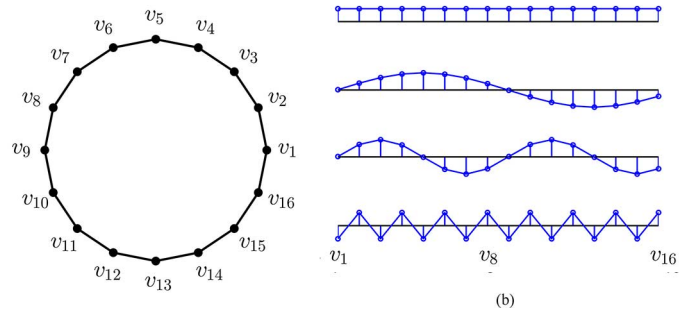


Fig. 1. (a) Cycle graph with 16 vertices. Signals defined on this graph are equivalent to standard discrete, periodic signals. (b) Several eigenvectors of the graph Laplacian. These eigenvectors exhibit the sinusoidal characteristics of the DFT basis.

where  $\deg(v)$  is the degree of the vertex  $v$ . One can also show that the maximum possible eigenvalue of  $\mathbf{L}$  is equal to 2, attained only by bipartite graphs. (These are graphs with two mutually exclusive subsets of vertices  $U_0$  and  $U_1$  such that every edge connects a vertex in  $U_0$  to a vertex in  $U_1$ .)

Given a signal  $\mathbf{x} \in \ell^2(G)$ , we can represent it in terms of the eigenvectors of  $\mathbf{L}$  by computing

$$\hat{\mathbf{x}} = \mathbf{F}^T \mathbf{x} \quad (5)$$

where  $\hat{\mathbf{x}}$  is called the graph Fourier transform of  $\mathbf{x}$ . The matrix  $\mathbf{F}^T$  represents the Fourier transform operator<sup>1</sup>. Since  $\mathbf{F}$  is orthogonal,  $\mathbf{F} \mathbf{F}^T = \mathbf{I}$ . It follows that we can invert the Fourier transform by taking

$$\mathbf{x} = \mathbf{F} \hat{\mathbf{x}}.$$

Using the Laplacian eigenvectors as a surrogate Fourier basis is a standard approach in the literature for defining signal processing operations on graphs [4]–[6], [9], [10]. It may not seem immediately obvious, though, that the analogy is a fair one. In what follows, we provide some justification for this approach.

First, consider the special case of a cycle graph, illustrated in Fig. 1(a). Signals defined on this graph can be thought of as discrete, periodic signals. The Laplacian of this graph is a circulant matrix, and is thus diagonalized by a discrete Fourier transform (DFT) matrix. Thus, in this case, the Laplacian eigenbasis is exactly the common sine/cosine DFT basis. Fig. 1(b) shows several such eigenvectors, which exhibit sinusoidal characteristics with increasing oscillation frequencies.

For general graphs, of course, the Laplacian eigenbasis is no longer the DFT basis. Nonetheless, the eigenvectors still satisfy our intuition about frequency. For example, we would like to say that a signal is “high pass” if its value changes significantly between neighboring vertices, and that it is “low pass” if its value varies very little. To quantify the variation of a signal on a graph, we can construct an  $N \times M$  normalized incidence matrix  $\mathbf{S}$  [21], where each column of  $\mathbf{S}$  corresponds to an edge

<sup>1</sup>There may be eigenvalues of  $\mathbf{L}$  with multiplicity greater than one, so we should really think of the Fourier transform as the set of projections onto the eigenspaces associated with each unique eigenvalue. The Fourier transform defined in this way is unique up to unitary transformations within eigenspaces. We can choose an orthogonal basis in each eigenspace, ensuring that  $\mathbf{F}$  is orthogonal.

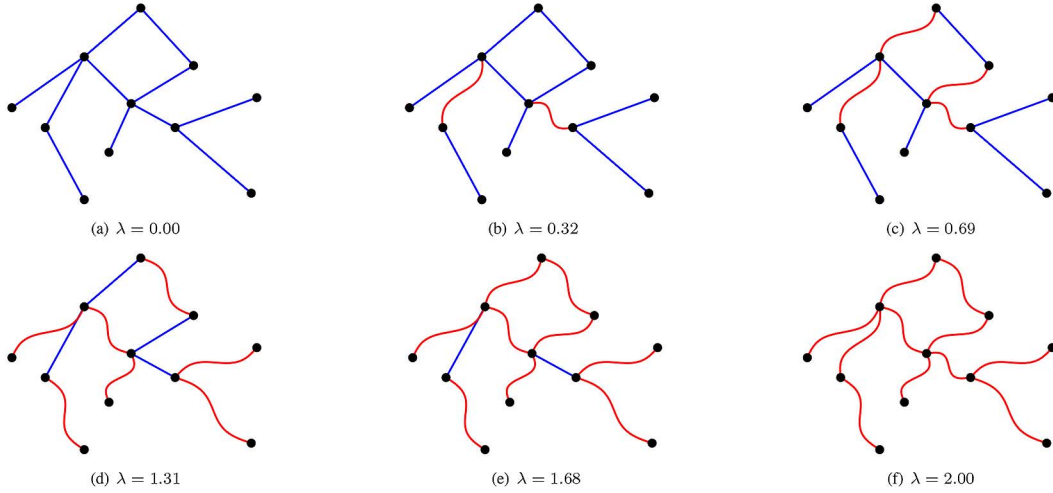


Fig. 2. Some Laplacian eigenvectors of a graph. Straight lines indicate that values on joined vertices have the same sign; wavy lines indicate that there is a sign change between the joined vertices. As is evident, eigenvectors associated with larger eigenvalues correspond to more sign changes and thus faster variation. (a)  $\lambda = 0.00$ . (b)  $\lambda = 0.32$ . (c)  $\lambda = 0.69$ . (d)  $\lambda = 1.31$ . (e)  $\lambda = 1.68$ . (f)  $\lambda = 2.00$ .

$e = (u, v)$  and has exactly two nonzero values:  $+\frac{1}{\sqrt{\deg(u)}}$  in the row corresponding to vertex  $u$ , and  $-\frac{1}{\sqrt{\deg(v)}}$  in the row corresponding to vertex  $v$ . The choice of  $(u, v)$  or  $(v, u)$ , and therefore the signs involved, is arbitrary for each edge (though it is important that each column have one positive and one negative value). For any  $\mathbf{x} \in \ell^2(G)$ , the vector  $\mathbf{y} = \mathbf{S}^T \mathbf{x}$  is a signal on the edges of the graph, where each edge has the difference between the normalized values<sup>2</sup> of  $\mathbf{x}$  on its endpoint vertices. So, in a way,  $\mathbf{y}$  is the “derivative” of  $\mathbf{x}$ . For any nonzero signal  $\mathbf{x}$ , we can then measure its normalized variation on the graph as

$$\begin{aligned} \frac{1}{\|\mathbf{x}\|^2} \sum_{u \sim v} \left( \frac{x(u)}{\sqrt{\deg(u)}} - \frac{x(v)}{\sqrt{\deg(v)}} \right)^2 &= \frac{1}{\|\mathbf{x}\|^2} \|\mathbf{y}\|^2 \\ &= \frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{S} \mathbf{S}^T \mathbf{x} \\ &= \frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{L} \mathbf{x} \end{aligned} \quad (6)$$

where the last equality ( $\mathbf{S} \mathbf{S}^T = \mathbf{L}$ ) is well known and easy to verify [21]. When the signal  $\mathbf{x}$  is the  $i$ th eigenvector  $\mathbf{f}_i$  of  $\mathbf{L}$ , the normalized variation in (6) becomes  $\lambda_i$ , the corresponding eigenvalue. This justifies the usage of Laplacian eigenvalues as frequencies: eigenvectors corresponding to the higher eigenvalues of  $\mathbf{L}$  are the high-variation components, and the lower eigenvalues correspond to low-variation components. We illustrate this fact with an example in Fig. 2.

### C. Graph and Spectral Spreads

We would like to quantify the localization of a signal on a graph in both the graph and spectral domains. To do so, we look

<sup>2</sup>The normalization by  $\frac{1}{\sqrt{\deg(u)}}$  will limit the undue effect on the Laplacian of a vertex with a large number of incident edges.

to the definitions of analogous quantities in classical time–frequency analysis. For a nonzero signal  $\mathbf{x} \in \mathcal{L}^2(\mathbb{R})$ , its time spread about a point  $t_0$  is defined by [2]

$$\Delta_{t, t_0}^2 \stackrel{\text{def}}{=} \frac{1}{\|\mathbf{x}\|^2} \int_{-\infty}^{\infty} (t - t_0)^2 |x(t)|^2 dt. \quad (7)$$

The overall time spread of  $x(t)$  is then obtained by minimizing over  $t_0$ , i.e.,

$$\Delta_t^2 \stackrel{\text{def}}{=} \min_{t_0} \frac{1}{\|\mathbf{x}\|^2} \int_{-\infty}^{\infty} (t - t_0)^2 |x(t)|^2 dt \quad (8)$$

where the minimizing value of  $t_0$  is given by  $t_0 = \frac{1}{\|\mathbf{x}\|^2} \int_{-\infty}^{\infty} t |x(t)|^2 dt$ . Generalizing (7) to signals defined on graphs, we introduce the following definition [24], [25].

**Definition 1 (Graph Spread):** For a nonzero signal  $\mathbf{x} \in \ell^2(G)$ , its graph spread about a vertex  $u_0$  is

$$\begin{aligned} \Delta_{g, u_0}^2(\mathbf{x}) &\stackrel{\text{def}}{=} \frac{1}{\|\mathbf{x}\|^2} \sum_{v \in V} d(u_0, v)^2 x(v)^2 \\ &= \frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x} \end{aligned} \quad (9)$$

where  $d(\cdot, \cdot)$  is the distance metric described in Section II-A, and  $\mathbf{P}_{u_0}$  is a diagonal matrix defined as

$$\mathbf{P}_{u_0} \stackrel{\text{def}}{=} \text{diag} \{d(u_0, v_1), d(u_0, v_2), \dots, d(u_0, v_N)\}. \quad (10)$$

**Remark:** Similar to (8), we can also define the overall (i.e., global) graph spread of  $\mathbf{x}$  as

$$\Delta_g^2(\mathbf{x}) \stackrel{\text{def}}{=} \min_{u_0 \in V} \frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}. \quad (11)$$

For our subsequent analysis on uncertainty principles though, we will focus on the local graph spread (i.e., about a particular vertex  $u_0$ ) as defined in (9). Unlike classical domains such as the real line whose topology is shift-invariant, the “landscape” of a graph can look very different around different vertices. Thus, it

is important to explicitly specify the center vertex  $u_0$  when considering the graph spread and uncertainty principles. If needed, global versions can always be obtained through finite minimization over all  $u_0 \in V$ .

The spectral spread of a signal defined on graphs requires more thought. In the classical case, the frequency spread of a real-valued signal  $x(t) \in \mathcal{L}^2(\mathbb{R})$  is given by [2]

$$\Delta_\omega^2 \stackrel{\text{def}}{=} \frac{1}{\|x\|^2} \int_{-\infty}^{\infty} \omega^2 |\hat{x}(\omega)|^2 \frac{d\omega}{2\pi} \quad (12)$$

where  $\hat{x}(\omega)$  is the Fourier transform of  $x(t)$ . This expression is simpler than that of the time spread in (7) because the frequency center is chosen to be  $\omega_0 = 0$  due to the symmetry of the Fourier transforms of real-valued signals. On recognizing that  $\omega^2 \hat{x}(\omega)$  is the Fourier transform of  $\frac{-d^2}{dt^2} x(t)$  and using Parseval's identity, we can rewrite (12) as

$$\Delta_\omega^2 = \frac{1}{\|x\|^2} \int_{-\infty}^{\infty} x(t) \frac{-d^2}{dt^2} x(t) dt. \quad (13)$$

Generalizing to the graph case, treating  $\mathbf{L}$  as analogous to the operator  $-\frac{d^2}{dt^2}$ , we obtain the following definition [24], [25].

**Definition 2 (Spectral Spread):** For a nonzero signal  $\mathbf{x} \in \ell^2(G)$ , we define its spectral spread as

$$\Delta_s^2(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{L} \mathbf{x} \quad (14)$$

$$= \frac{1}{\|\mathbf{x}\|^2} \sum_{n=1}^N \lambda_n |\hat{x}_n|^2 \quad (15)$$

where the second equality follows from the decomposition of  $\mathbf{L}$  in (3) and the definition of graph Fourier transforms in (5).

**Remark:** The equivalent definitions in (14) and (15) reveal two different facets of the spectral spread: while (15) perhaps more clearly justifies the ‘‘spectral’’ nature of  $\Delta_s^2(\mathbf{x})$ , the form in (14) shows that  $\Delta_s^2(\mathbf{x})$  can also be understood as the normalized variation of  $\mathbf{x}$  introduced in (6).

### III. UNCERTAINTY PRINCIPLES: BOUNDS AND CHARACTERIZATIONS

Intuitively, we can reason that there should exist a tradeoff between the graph and spectral spreads of a signal. If the graph spread  $\Delta_g^2$  is small, then the signal must resemble an impulse centered at some vertex; in this case, the normalized variation (i.e., the spectral spread  $\Delta_s^2$ ) should be high. If instead  $\Delta_s^2$  is small, then the signal cannot vary too quickly; it will thus take a long distance for the signal values to drop significantly from the peak value, in which case the graph spread will be high. How can one precisely quantify the above intuition? What are the signals with a given spectral spread that are maximally localized on the graph? These are the fundamental questions addressed in this section.

#### A. Feasibility Region

In the classical uncertainty principle, not all pairs of time–frequency spreads  $(\Delta_t^2, \Delta_\omega^2)$  are achievable, and the tradeoff is quantified by the celebrated inequality  $\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}$ , which holds for any nonzero function  $x(t) \in \mathcal{L}^2(\mathbb{R})$  [1], [2]. Furthermore,

this bound is tight. In fact, any pair of the form  $(\Delta_t^2, \Delta_\omega^2) = (c, \frac{1}{4c})$  for  $c > 0$  is achievable by a function of the form  $x(t) = \exp(-\frac{t^2}{4c})$ .

In a similar way, we are interested in characterizing the following *feasibility region*:

$$\mathcal{D}_{u_0} \stackrel{\text{def}}{=} \{(s, g) : \Delta_s^2(\mathbf{x}) = s, \Delta_{g, u_0}^2(\mathbf{x}) = g \text{ for some nonzero } \mathbf{x} \in \ell^2(G)\} \quad (16)$$

containing all pairs of the form  $(\Delta_s^2, \Delta_{g, u_0}^2)$  that are achievable on a graph  $G$ , using  $u_0$  as the center vertex.

**Proposition 1:** Let  $\mathcal{D}_{u_0}$  be the feasibility region for a *connected* graph  $G$  with  $N$  vertices. Then, the following properties hold.

- $\mathcal{D}_{u_0}$  is a closed subset of  $[0, \lambda_N] \times [0, \mathcal{E}_G^2(u_0)]$ , where  $\lambda_N \leq 2$  is the largest eigenvalue of graph Laplacian  $\mathbf{L}$ , and  $\mathcal{E}_G(u_0) \stackrel{\text{def}}{=} \max_{v \in V} d(u_0, v)$  is the *eccentricity* of the center vertex  $u_0$ .
- $\mathcal{D}_{u_0}$  intersects the horizontal axis at exactly one point,  $(1, 0)$ , and the vertical axis at exactly one point,  $(0, \mathbf{f}_1^T \mathbf{P}_{u_0}^2 \mathbf{f}_1)$ , where  $\mathbf{f}_1$  is the eigenvector defined in (4).
- The points  $(1, \mathcal{E}_G^2(u_0))$  and  $(\lambda_N, \mathbf{f}_N^T \mathbf{P}_{u_0}^2 \mathbf{f}_N)$ , where  $\mathbf{f}_N$  is any unit-norm eigenvector associated with  $\lambda_N$ , belong to  $\mathcal{D}_{u_0}$ .
- $\mathcal{D}_{u_0}$  is a convex set if the number of vertices  $N \geq 3$ .

*Proof:*

- The graph and spectral spreads of any nonzero signal can be bounded by the largest and smallest eigenvalues of  $\mathbf{L}$  and  $\mathbf{P}_{u_0}^2$ . More precisely, using the Rayleigh inequalities [31], we have

$$0 = \lambda_1 \leq \frac{\mathbf{x}^T \mathbf{L} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_N$$

and, similarly,

$$0 \leq \frac{\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \max_{1 \leq i \leq N} (\mathbf{P}_{u_0}^2)_{ii} = \mathcal{E}_G^2(u_0).$$

$\mathcal{D}_{u_0}$  is compact, and therefore closed, because it is the image of a compact set under a continuous transform [32]. Specifically, if we take the unit sphere in  $\mathbb{R}^N$ , a compact set, and apply the map  $f : \mathbf{x} \mapsto (\Delta_s^2(\mathbf{x}), \Delta_{g, u_0}^2(\mathbf{x}))$ , which is continuous on the unit sphere, we get the whole uncertainty region.

- A signal has zero graph spread (i.e.,  $\Delta_{g, u_0}^2(\mathbf{x}) = 0$ ) if and only if it is an impulse supported on  $u_0$ , i.e.,  $x(v) = c$  if  $v = u_0$  and  $x(v) = 0$  otherwise, for some nonzero scalar  $c$ . Meanwhile, using (14) and (6), one can verify that the normalized variation (and thus the spectral spread  $\Delta_s^2$ ) of such impulse signals is equal to 1. It follows that  $(1, 0)$  is the only point that lies at the intersection of  $\mathcal{D}_{u_0}$  and the horizontal axis. Next, consider the intersection of  $\mathcal{D}_{u_0}$  with the vertical axis. Since  $\Delta_s^2(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x} / \|\mathbf{x}\|^2 \geq \lambda_1 = 0$ , the spectral spread  $\Delta_s^2(\mathbf{x}) = 0$  if and only if  $\mathbf{x}$  is an eigenvector of  $\mathbf{L}$  associated with the smallest eigenvalue  $\lambda_1 = 0$ . (See (4) for an example.) Such eigenvectors are also unique (up to scalar multiplications) since the

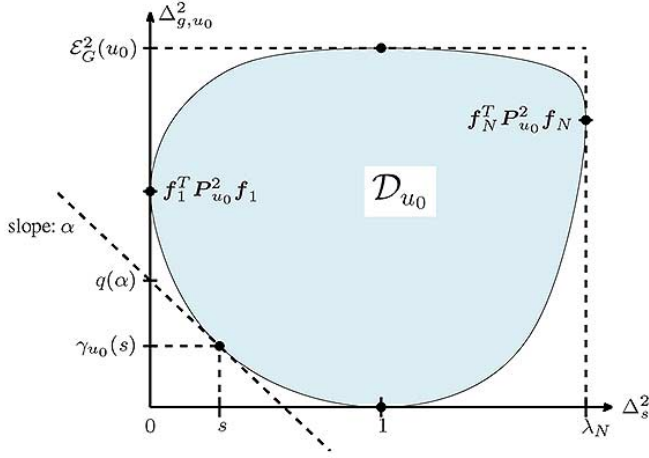


Fig. 3. Feasibility region  $\mathcal{D}_{u_0}$  for the spectral and graph spreads.  $\mathcal{D}_{u_0}$  is a bounded and convex set that intersects the horizontal (and vertical) axis at exactly one point. The lower boundary of  $\mathcal{D}_{u_0}$  can be implicitly computed by considering supporting lines of varying slopes. The achievable region must lie in the half-plane above the supporting line (found by solving an eigenvalue problem.).

smallest eigenvalue  $\lambda_1$  of connected graphs always has multiplicity one [30].

- c) The inclusion of  $(\lambda_N, \mathbf{f}_N^T \mathbf{P}_{u_0}^2 \mathbf{f}_N)$  in  $\mathcal{D}_{u_0}$  is clear. For the first point  $(1, \mathcal{E}_G^2(u_0))$ , consider an impulse function supported at the furthest vertex on the graph from  $u_0$ . Similar to (b), we can compute its spectral and graph spreads as  $\Delta_s^2 = 1$  and  $\Delta_{g,u_0}^2 = \mathcal{E}_G^2(u_0)$ , respectively.
- d) See Appendix A.

*Remark:* Fig. 3 illustrates a typical feasibility region  $\mathcal{D}_{u_0}$  as specified by Proposition 1. The boundedness and convexity of  $\mathcal{D}_{u_0}$  imply that the entire region can be completely characterized by its upper and lower boundaries: any pair between the two boundaries must also be achievable. Furthermore, the lower boundary must be convex and the upper boundary must be concave.

### B. Uncertainty Curve

In what follows, we will describe a technique for computing the lower boundary curve of  $\mathcal{D}_{u_0}$ , which we call the *uncertainty curve*.

*Definition 3:* Given a connected graph  $G$ , the uncertainty curve with respect to a center vertex  $u_0$  is

$$\begin{aligned} \gamma_{u_0}(s) &\stackrel{\text{def}}{=} \min_{\mathbf{x}} \Delta_{g,u_0}^2(\mathbf{x}) \text{ subject to } \Delta_s^2(\mathbf{x}) = s \\ &= \min_{\mathbf{x}} \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x} \text{ subject to } \mathbf{x}^T \mathbf{x} = 1 \text{ and } \mathbf{x}^T \mathbf{L} \mathbf{x} = s \end{aligned} \quad (17)$$

for all  $s \in [0, \lambda_N]$ .

*Remark:* We could also define and study the upper boundary curve of  $\mathcal{D}_{u_0}$  in a similar way. We choose to focus on the lower boundary curve because it provides an uncertainty bound analogous to the classical bound (1). We will say that a signal  $\mathbf{x}$  achieves the uncertainty curve if  $\Delta_{g,u_0}^2(\mathbf{x}) = \gamma_{u_0}(\Delta_s^2(\mathbf{x}))$ .

We note that (17) is a quadratically constrained quadratic program [33]. The equality constraints make the problem

nonconvex. On differentiating the corresponding Lagrangian function

$$\Lambda(\mathbf{x}; \alpha, \lambda) \stackrel{\text{def}}{=} \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x} - \alpha(\mathbf{x}^T \mathbf{L} \mathbf{x} - s) - \lambda(\mathbf{x}^T \mathbf{x} - 1)$$

we see that the optimal solution  $\mathbf{x}^*$  to (17) must satisfy

$$(\mathbf{P}_{u_0}^2 - \alpha \mathbf{L}) \mathbf{x}^* = \lambda \mathbf{x}^*$$

for some  $\alpha, \lambda \in \mathbb{R}$ . If we treat  $\alpha$  as being fixed, then the aforementioned equality becomes an eigenvalue problem. This observation leads us to study the matrix-valued function

$$\mathbf{M}(\alpha) \stackrel{\text{def}}{=} \mathbf{P}_{u_0}^2 - \alpha \mathbf{L}. \quad (18)$$

For any  $\alpha$ , the smallest eigenvalue of  $\mathbf{M}(\alpha)$ , denoted by

$$q(\alpha) \stackrel{\text{def}}{=} \lambda_{\min}(\mathbf{M}(\alpha))$$

and its associated eigenspace, denoted by  $\mathcal{S}(\alpha)$ , are key to our analysis of the uncertainty curve  $\gamma_{u_0}(s)$ .

*Proposition 2:* For any  $\alpha \in \mathbb{R}$  and any unit-norm eigenvector  $\mathbf{v}$  in  $\mathcal{S}(\alpha)$ , the point  $(\mathbf{v}^T \mathbf{L} \mathbf{v}, \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v})$  is on  $\gamma_{u_0}(s)$ .

*Proof:* Let  $\mathbf{x}$  be an arbitrary signal with  $\|\mathbf{x}\| = 1$ . By definition,  $\Delta_{g,u_0}^2(\mathbf{x}) - \alpha \Delta_s^2(\mathbf{x}) = \mathbf{x}^T \mathbf{M}(\alpha) \mathbf{x}$ . Applying Rayleigh's inequality to  $\mathbf{M}(\alpha)$  thus leads to

$$\Delta_{g,u_0}^2(\mathbf{x}) - \alpha \Delta_s^2(\mathbf{x}) \geq q(\alpha) \quad (19)$$

$$= \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v} - \alpha \mathbf{v}^T \mathbf{L} \mathbf{v} \quad (20)$$

where (20) comes from the fact that  $\mathbf{v}$  is an eigenvector associated with  $q(\alpha)$ . Let  $s = \mathbf{v}^T \mathbf{L} \mathbf{v}$ . On specializing the relationship (20) to those signals  $\mathbf{x}$  satisfying  $\Delta_s^2(\mathbf{x}) = s$ , we have

$$\Delta_{g,u_0}^2(\mathbf{x}) \geq \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v}$$

which indicates that the point  $(\mathbf{v}^T \mathbf{L} \mathbf{v}, \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v})$  must lie on the uncertainty curve  $\gamma_{u_0}(s)$ . ■

There is an interesting geometric interpretation of the aforementioned derivations: as illustrated in Fig. 3, for any  $\alpha$ , the inequality in (19) defines a half-plane in which  $\mathcal{D}_{u_0}$  must lie. The boundary of the half-plane, a line of slope  $\alpha$  defined by

$$\Delta_{g,u_0}^2 - \alpha \Delta_s^2 = q(\alpha)$$

provides a tight lower bound to  $\mathcal{D}_{u_0}$ . Varying the values of  $\alpha$  generates a family of such half-planes, the intersection of which contains  $\mathcal{D}_{u_0}$ . For readers familiar with convex analysis, we note that  $q(\alpha)$  is the Legendre transform of  $\gamma_{u_0}(s)$  [33].

Proposition 2 guarantees that any nonzero eigenvector of  $\mathbf{M}(\alpha)$  associated with the smallest eigenvalue  $q(\alpha)$  generates a point on the curve  $\gamma_{u_0}(s)$ . Next, we will show that the converse is also true: every point on  $\gamma_{u_0}(s)$  is achieved by an eigenvector in  $\mathcal{S}(\alpha)$  for some  $\alpha$ . To establish this result, we need to introduce the following two functions:

$$h_+(\alpha) \stackrel{\text{def}}{=} \max_{\mathbf{x} \in \mathcal{S}(\alpha): \|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{L} \mathbf{x} \quad (21)$$

$$h_-(\alpha) \stackrel{\text{def}}{=} \min_{\mathbf{x} \in \mathcal{S}(\alpha): \|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{L} \mathbf{x}$$



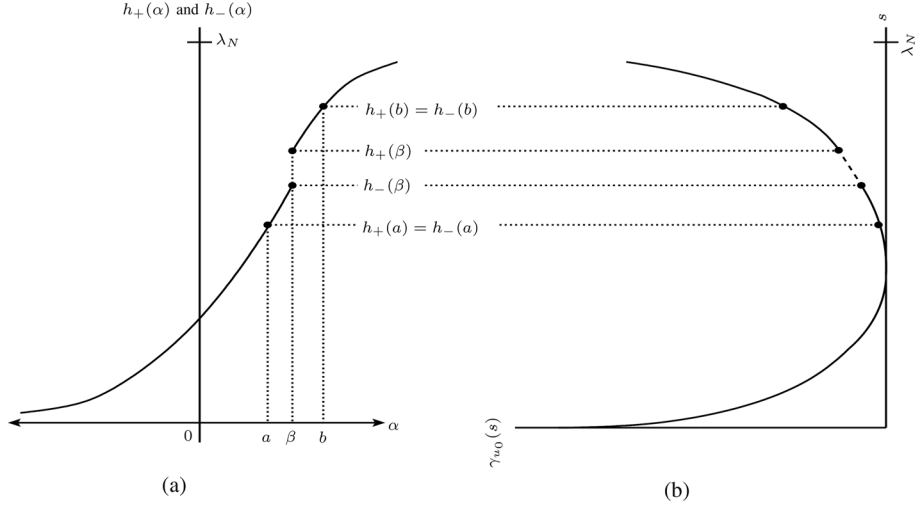


Fig. 4. Mapping of the eigenvectors in  $S(\alpha)$  onto the  $s$ - $g$  plane is shown. (a)  $h_+(\alpha)$  and  $h_-(\alpha)$  are plotted against  $\alpha$  (they coincide except at jumps in the plot.) They are, respectively, the maximum and minimum spectral spreads of elements of the eigenspace  $S(\alpha)$ . Any element of  $S(\alpha)$  determines a point on the graph of  $\gamma_{u_0}(s)$ . (b) When  $S(\alpha)$  is of dimension greater than one, it corresponds to a line segment on  $\gamma_{u_0}(s)$ .

which measure, respectively, the maximum and minimum spectral spread (i.e., the horizontal coordinate on the  $s$ - $g$  plane) that can be achieved by eigenvectors in  $S(\alpha)$ .

**Lemma 1:** The following properties hold for  $h_+(\alpha)$  and  $h_-(\alpha)$ .

- a) They are increasing functions, i.e.,  $h_+(\alpha_1) \leq h_+(\alpha_2)$  and  $h_-(\alpha_1) \leq h_-(\alpha_2)$  for all  $\alpha_1 < \alpha_2$ .
- b) They have the same limits as  $|\alpha|$  tends to infinity

$$\lim_{\alpha \rightarrow -\infty} h_+(\alpha) = \lim_{\alpha \rightarrow -\infty} h_-(\alpha) = 0 \quad (22)$$

and

$$\lim_{\alpha \rightarrow +\infty} h_+(\alpha) = \lim_{\alpha \rightarrow +\infty} h_-(\alpha) = \lambda_N. \quad (23)$$

- c) On any finite interval  $[a, b]$ , the functions  $h_+(\alpha)$  and  $h_-(\alpha)$  differ on at most a finite number of points, denoted by  $\mathcal{B} \stackrel{\text{def}}{=} \{\beta_1, \beta_2, \dots, \beta_k\}$  for some  $k \geq 0$ . Except for these points,  $h_+(\alpha)$  and  $h_-(\alpha)$  coincide, are continuous, and satisfy

$$h_+(\alpha) = h_-(\alpha) = -q'(\alpha), \text{ for all } \alpha \in [a, b] \setminus \mathcal{B} \quad (24)$$

where  $q'(\alpha)$  is the derivative of  $q(\alpha)$ . At the points, if any, where they do differ,  $h_+(\alpha)$  and  $h_-(\alpha)$  have jump discontinuities. Moreover, for all  $\beta \in \mathcal{B}$

$$h_+(\beta) = \lim_{\alpha \rightarrow \beta^+} h_+(\alpha) > \lim_{\alpha \rightarrow \beta^-} h_-(\alpha) = h_-(\beta)$$

where the limits are taken as  $\alpha$  approaches  $\beta$  from the positive and negative sides, respectively.

*Proof:* See Appendix B. ■

The results of Lemma 1 are illustrated in Fig. 4(a), where we plot a typical example of  $h_+(\alpha)$  and  $h_-(\alpha)$ : as  $\alpha$  increases from  $-\infty$  to  $+\infty$ , the values of the functions increase from 0 to  $\lambda_N$ . Within any finite interval,  $h_+(\alpha) = h_-(\alpha)$  except at a finite number of points (e.g., the point  $\beta$  in the figure). At these “jump points,”  $h_+(\alpha)$  is right-continuous, whereas  $h_-(\alpha)$  is left-continuous.

Since we are only considering connected graphs,  $\lambda_1 = 0$  has multiplicity 1, and so,  $\mathbf{f}_1$  is the unique vector (up to scaling) that achieves the uncertainty curve with  $\Delta_s^2 = 0$ . At the other end,  $\lambda_N$  may have multiplicity, but some vector in its eigenspace will achieve the uncertainty curve with  $\Delta_s^2 = \lambda_N$ . For values of  $s \in (0, \lambda_{\max})$ , we can use the following theorem to precisely characterize vectors that achieve the uncertainty curve at  $s$ .

**Theorem 1:** A signal  $\mathbf{x} \in \ell^2(G)$  with  $\Delta_s^2(\mathbf{x}) \in (0, \lambda_{\max})$  achieves the uncertainty curve, i.e.,  $\Delta_{g, u_0}^2(\mathbf{x}) = \gamma(\Delta_s^2(\mathbf{x}))$ , if and only if it is a nonzero eigenvector in  $S(\alpha)$  for some  $\alpha$ .

*Proof:* The “if” direction has been established in Proposition 2. To prove the “only if” direction, we will show that for any signal  $\mathbf{x} \in \ell^2(G)$  that achieves the uncertainty curve, there is an  $\alpha$  and a unit-norm eigenvector  $\mathbf{v} \in S(\alpha)$  such that  $\mathbf{v}^T \mathbf{L} \mathbf{v} = \Delta_s^2(\mathbf{x})$ . Since both  $\mathbf{x}$  and  $\mathbf{v}$  lie on the uncertainty curve (with the former given as an assumption and the latter guaranteed by Proposition 2), we have  $\Delta_{g, u_0}^2(\mathbf{x}) = \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v}$ , and thus

$$\frac{1}{\|\mathbf{x}\|^2} \mathbf{x}^T \mathbf{M}(\alpha) \mathbf{x} = \Delta_{g, u_0}^2(\mathbf{x}) - \alpha \Delta_s^2(\mathbf{x}) = \mathbf{v}^T \mathbf{M}(\alpha) \mathbf{v} = q(\alpha).$$

Now, since  $q(\alpha)$  is the smallest eigenvalue of  $\mathbf{M}(\alpha)$ , the aforementioned equality implies that  $\mathbf{x}$  must also be an eigenvector associated with  $q(\alpha)$ . In fact,  $\mathbf{x}$  will be equal to  $\mathbf{v}$  (up to a scalar multiple) if  $q(\alpha)$  has multiplicity one. The remainder of the proof verifies the claim, namely, for any  $s \in (0, \lambda_N)$  we can find an  $\alpha$  and a unit-norm eigenvector  $\mathbf{v} \in S(\alpha)$  such that  $\mathbf{v}^T \mathbf{L} \mathbf{v} = s$ .

By part (b) of Lemma 1, we can always find some  $a < b$  such that  $h_-(a) < s < h_+(b)$ . Furthermore, part (c) of Lemma 1 ensures that, within the interval  $[a, b]$ , the two functions  $h_+(\alpha)$  and  $h_-(\alpha)$  differ (and are discontinuous) on at most a finite number of points. For notational simplicity, and without loss of generality, we assume that there is only one such discontinuity point, denoted by  $\beta \in [a, b]$ . As shown in Fig. 4, the interval  $[h_-(a), h_+(b)]$  can now be written as the union of three subintervals

$$[h_-(a), h_-(\beta)), [h_-(\beta), h_+(\beta)], \text{ and } (h_+(\beta), h_+(b)]$$

to one of which  $s$  must belong.

We first consider the case where  $s \in [h_-(a), h_-(\beta)]$ . Lemma 1 says that  $h_-(\alpha)$  is a continuous function on  $[a, \beta]$ . By the intermediate value theorem, there exists some  $\alpha_0 \in [a, \beta]$  such that  $h_-(\alpha_0) = s$ . By definition,  $h_-(\alpha_0) = \min_{\mathbf{z} \in \mathcal{S}(\alpha_0): \|\mathbf{z}\|=1} \mathbf{z}^T \mathbf{L} \mathbf{z}$ . Since the eigenspace  $\mathcal{S}(\alpha_0)$  has finite dimensions, the minimization can always be achieved by some unit-norm eigenvector  $\mathbf{v} \in \mathcal{S}(\alpha_0)$ , i.e.,  $s = h_-(\alpha_0) = \mathbf{v}^T \mathbf{L} \mathbf{v}$ . The same line of reasoning can be used when  $s$  belongs to the third subinterval,  $(h_+(\beta), h_+(b)]$ . This leaves us with the remaining case when  $s \in [h_-(\beta), h_+(\beta)]$ . Let

$$\mathbf{v}_+ \stackrel{\text{def}}{=} \operatorname{argmax}_{\mathbf{z} \in \mathcal{S}(\beta): \|\mathbf{z}\|=1} \mathbf{z}^T \mathbf{L} \mathbf{z} \quad \text{and} \quad \mathbf{v}_- \stackrel{\text{def}}{=} \operatorname{argmin}_{\mathbf{z} \in \mathcal{S}(\beta): \|\mathbf{z}\|=1} \mathbf{z}^T \mathbf{L} \mathbf{z}$$

and consider the vector-valued function  $\mathbf{y}(\theta) \stackrel{\text{def}}{=} \frac{\cos(\theta)\mathbf{v}_+ + \sin(\theta)\mathbf{v}_-}{1 + \sin(2\theta)\mathbf{v}_+^T \mathbf{v}_-}$ , defined for  $\theta \in [0, \pi/2]$ . The denominator is nonzero for every  $\theta$ , since  $\mathbf{v}_- \neq -\mathbf{v}_+$  (otherwise, we would have  $h_-(\beta) = h_+(\beta)$ ). So  $\mathbf{y}(\theta)$  is of unit norm and is a continuous function of  $\theta$ . It also must belong to  $\mathcal{S}(\beta)$  since it is a linear combination of two elements of the subspace. Furthermore,  $\mathbf{y}(0)^T \mathbf{L} \mathbf{y}(0) = h_+(\beta)$  and  $\mathbf{y}(\pi/2)^T \mathbf{L} \mathbf{y}(\pi/2) = h_-(\beta)$ . By the intermediate value theorem,  $\mathbf{y}(\theta)$  for  $\theta \in [0, \pi/2]$  achieves all the values in between. In particular, there exists some  $\theta_0$  such that  $\mathbf{y}(\theta_0)^T \mathbf{L} \mathbf{y}(\theta_0) = s$ . We note that since every element of  $\mathcal{S}(\beta)$  achieves a point on the line  $g - \beta s = q(\beta)$ , this interpolation procedure amounts to including the straight line segment between the two endpoints as part of the uncertainty curve. ■

*Remark:* If  $\mathcal{S}(\alpha)$  is one-dimensional for every  $\alpha \in [a, b]$ , or more generally if there is a single distinct eigenvalue function that achieves the minimum on  $[a, b]$ , then from Theorem 1 as well as Lemma 1 and its proof,  $q(\alpha)$  is analytic on  $[a, b]$  and the corresponding portion of the uncertainty curve can be expressed in parametric form as

$$\begin{cases} s(\alpha) = -q'(\alpha) \\ \gamma_{u_0}(s) = q(\alpha) - \alpha q'(\alpha) \end{cases} \quad (25)$$

where the first equality is due to (24) and the second is due to the fact that any vector in  $\mathcal{S}(\alpha)$  must achieve a point on the line  $g - \alpha s = q(\alpha)$ .

In general, Theorem 1 and its proof justify a way to obtain the uncertainty curve: for every  $\alpha$ , we find the eigenvectors associated with the smallest eigenvalue of  $\mathbf{M}(\alpha)$ . These eigenvectors will give us points on  $\gamma_{u_0}(s)$ . By “sweeping” the values of  $\alpha$  from  $-\infty$  to  $\infty$ , the entire curve can then be traced.

### C. Fast Approximation Algorithm

In practice, of course, we must sample and work with a finite set of  $\alpha$ 's, which lead to an approximation of the true curve. In what follows, we describe an efficient algorithm that can compute an approximation—more specifically, an upper and lower bound—of  $\gamma_{u_0}(s)$  with any desired accuracy.

Since  $\gamma_{u_0}(s)$  is the lower boundary of the convex region  $\mathcal{D}_{u_0}$ , it is itself a convex function. We can therefore use the *sandwich* algorithm described in [26] to approximate it. The algorithm can be easily understood by studying Fig. 5(a): consider a segment of the curve  $\gamma_{u_0}(s)$  with two end points  $A$  and  $B$ , whose coordinates are denoted by  $(a, \gamma_{u_0}(a))$  and  $(b, \gamma_{u_0}(b))$ , respectively.

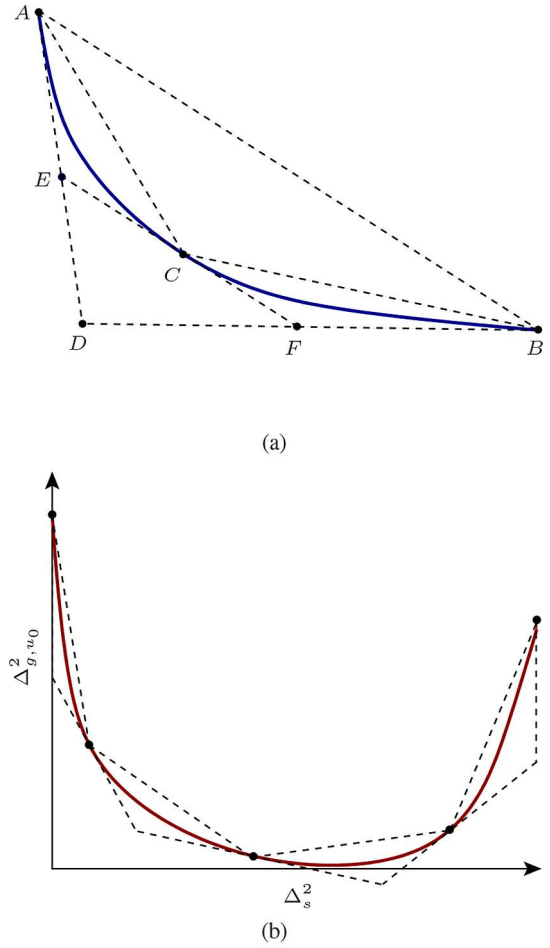


Fig. 5. Illustration of the *sandwich* algorithm. (a) Single refinement step on a segment of the uncertainty curve. (b) Two refinement steps on the full curve.

Also given are supporting lines<sup>3</sup> containing the end points, represented by the line segments  $AD$  and  $BD$ . Due to the convexity of  $\gamma_{u_0}(s)$ , the chord that connects  $A$  to  $B$  must lie entirely above the curve and thus form an upper bound. Similarly, the combination of  $AD$  and  $BD$  forms a piecewise linear lower bound of  $\gamma_{u_0}(s)$ .

To refine these two initial bounds, let  $\alpha$  be the slope of the chord, i.e.,

$$\alpha = \frac{\gamma_{u_0}(b) - \gamma_{u_0}(a)}{b - a}. \quad (26)$$

Computing the smallest eigenvalue  $q(\alpha)$  and the associated eigenvectors of  $\mathbf{M}(\alpha)$ , we can obtain a new point on the curve, denoted by  $C$  in Fig. 5(a). The  $s$ - $g$  coordinates of  $C$  are  $(\mathbf{v}^T \mathbf{L} \mathbf{v}, \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v})$ , where  $\mathbf{v}$  is a unit-norm element in the eigenspace  $\mathcal{S}(\alpha)$ . Our previous analysis in Section III-B—in particular, (19) and (20)—guarantees that the line

$$g - \alpha s = q(\alpha)$$

that passes through  $C$  must be a supporting line of  $\gamma_{u_0}(s)$ . In other words,  $\alpha$  is a subderivative of  $\gamma_{u_0}(s)$  at point  $C$ , and

<sup>3</sup>A supporting line is a line that intersects a curve but does not separate any two points on the curve [33].



is the derivative if it exists. This property, together with the construction of  $\alpha$  in (26), also ensures that  $C$  is always located between  $A$  and  $B$ . As illustrated in the figure, the curve is now bounded above by joining the three points ( $A, C$ , and  $B$ ), and it is bounded below by joining the three supporting lines ( $AE, EF$ , and  $FB$ ).

The procedure above can then be repeated, in a recursive fashion, on the two curve segments  $AC$  and  $CB$ . Each stage of the recursion roughly doubles the number of points in the approximation, and we proceed until a fixed number of refinements have been computed. Fig. 5(b) shows the lower and upper bounds of  $\gamma_{u_0}(s)$  obtained by starting from two initial points  $(0, \mathbf{f}_1^T \mathbf{P}_{u_0}^2 \mathbf{f}_1)$  and  $(\lambda_N, \mathbf{f}_N^T \mathbf{P}_{u_0}^2 \mathbf{f}_N)$  and running the algorithm for two refinement iterations, involving a total of five eigenvalue evaluations (each corresponding to a single point drawn on the curve.) We can see that the proposed algorithm starts producing reasonable approximations of  $\gamma_{u_0}(s)$  after just a small number of steps.

Let  $\eta_u^{(n)}(\cdot)$  and  $\eta_\ell^{(n)}(\cdot)$  denote, respectively, the upper and lower bounds the algorithm generates after  $n$  eigenvalue evaluations. We measure the quality of approximation by computing the Hausdorff distance [26] between these two bounds, defined as

$$d(n) = \sup_{s_1} \inf_{s_2} \left[ (s_1 - s_2)^2 + (\eta_u^{(n)}(s_1) - \eta_\ell^{(n)}(s_2))^2 \right]^{\frac{1}{2}}.$$

Informally, the Hausdorff distance  $d(n)$  is small if the two bounding curves are close to each other. The following theorem, which follows directly from [26, Th. 3], shows that  $d(n)$  is of order  $1/n^2$ .

**Theorem 2:** Let  $\varepsilon > 0$  be any preset precision level. To get  $d(n) \leq \varepsilon$ , it is sufficient to run the approximation algorithm until we have  $n \geq \max \left\{ 4, \sqrt{9W/\varepsilon} + 2 \right\}$ , where  $W = \sqrt{\lambda_N^2 + \mathcal{E}_G^4(u_0)}$ .

**Remark:** In many practical applications, the underlying graph  $G$  is large but sparse. Correspondingly,  $\mathbf{M}(\cdot)$  are sparse matrices. Obtaining an approximation of  $\gamma_{u_0}(s)$  within a given precision  $\varepsilon$  then boils down to computing (e.g., via iterative power methods) the smallest eigenvalue and an associated eigenvector of about  $\mathcal{O}(1/\sqrt{\varepsilon})$  sparse matrices.

Instead of approximating the whole curve, we may wish to find  $\gamma_{u_0}(s)$  only for some particular value of  $s$ , as well as the signal that achieves it. The sandwich algorithm can be modified slightly to this end. At each step of the approximation procedure, we can choose to refine only the segment containing  $s$ , ignoring all other segments. Iterating in this way, we will find both  $\gamma_{u_0}(s)$  and the vector with spectral spread  $s$  that achieves the bound.

#### IV. UNCERTAINTY CURVE FOR SPECIAL GRAPH FAMILIES

The uncertainty curves for several standard graph families are analyzed in this section. The structure and regularity of complete graphs and star graphs make it possible to find closed-form expressions for their corresponding curves. For Erdős-Rényi random graphs [27], [28], we will derive and compute analytical approximations for the expected (i.e., mean) curves under different parameters. Throughout this section, the distance metric  $d(\cdot, \cdot)$  is assumed to be the geodesic distance.

##### A. Complete Graphs

A complete graph is a fully connected graph in which every pair of distinct vertices is connected by an edge [12]. It is often used to model fully connected subgraphs, or *cliques*, in real-world networks [34]. The Laplacian matrix of a complete graph with  $N$  vertices is given by

$$\mathbf{L}_{ij} = \begin{cases} 1, & \text{if } i = j \\ -\frac{1}{N-1}, & \text{otherwise} \end{cases} \quad (27)$$

i.e., the diagonal of  $\mathbf{L}$  is all 1, and the off-diagonal elements are all equal to  $-\frac{1}{N-1}$ . It is easy to verify that  $\mathbf{L}$  has eigenvalue 0 with multiplicity 1, and eigenvalue  $\frac{N}{N-1}$  with multiplicity  $N-1$ . Without loss of generality, we can choose the first vertex as the center. The diagonal distance matrix is then

$$\mathbf{P}_{u_0} = \text{diag}\{0, 1, 1, \dots, 1\}. \quad (28)$$

We would like to compute the uncertainty curve  $\gamma(s)$  for a complete graph for  $s \in [0, \frac{N}{N-1}]$ . First, we will show that any vector that achieves the uncertainty curve has a special form.

**Proposition 3:** For a complete graph, suppose  $\tilde{\mathbf{x}}$  achieves the uncertainty curve. Then,  $\tilde{\mathbf{x}}$  is of the form

$$\tilde{\mathbf{x}} = [x_1, x_2, x_2, \dots, x_2]^T. \quad (29)$$

*Proof:* See Appendix C. ■

The result in Proposition 3 suggests that, for complete graphs, we need only consider vectors of the form in (29). Enforcing the unit-norm constraint on (29), we can further simplify these eigenvectors as  $\tilde{\mathbf{x}}(\theta) = [\cos \theta, \frac{\sin \theta}{\sqrt{N-1}}, \frac{\sin \theta}{\sqrt{N-1}}, \dots, \frac{\sin \theta}{\sqrt{N-1}}]^T$  for some parameter  $\theta$ . The graph spread in this case is given by

$$\Delta_{g,u_0}^2 = \sum_{i=1}^{N-1} 1 \cdot \frac{\sin^2 \theta}{N-1} = \frac{1}{2} - \frac{1}{2} \cos 2\theta$$

where the second equality is due to a standard trigonometric identity. Meanwhile, by using the variational form in (6), we can compute the spectral spread as

$$\begin{aligned} \Delta_s^2 &= (N-1) \left( \frac{\cos \theta}{\sqrt{N-1}} - \frac{\sin \theta}{N-1} \right)^2 \\ &= \frac{N}{2N-2} - \frac{1}{\sqrt{N-1}} \sin 2\theta + \frac{N-2}{2N-2} \cos 2\theta. \end{aligned} \quad (30)$$

Combining these two expressions and using the identity  $\sin^2 2\theta + \cos^2 2\theta = 1$ , we can see that the uncertainty curve  $\gamma_{u_0}(s)$  is part of the ellipse given by

$$(2\Delta_{g,u_0}^2 - 1)^2 + (N-1) \left( \Delta_s^2 + \frac{N-2}{N-1} \Delta_{g,u_0}^2 - 1 \right)^2 = 1. \quad (31)$$

For fixed  $s = \Delta_s^2$ , solving for  $\gamma_{u_0}(s) = \Delta_{g,u_0}^2$  (by picking the smaller of the two solutions to (31)) leads to

$$\begin{aligned} \gamma_{u_0}(s) &= \\ &= \frac{N - s(N-2) - 2\sqrt{1 - (N-2)(s-1) - (N-1)(s-1)^2}}{4 + (N-2)^2/(N-1)} \end{aligned} \quad (32)$$

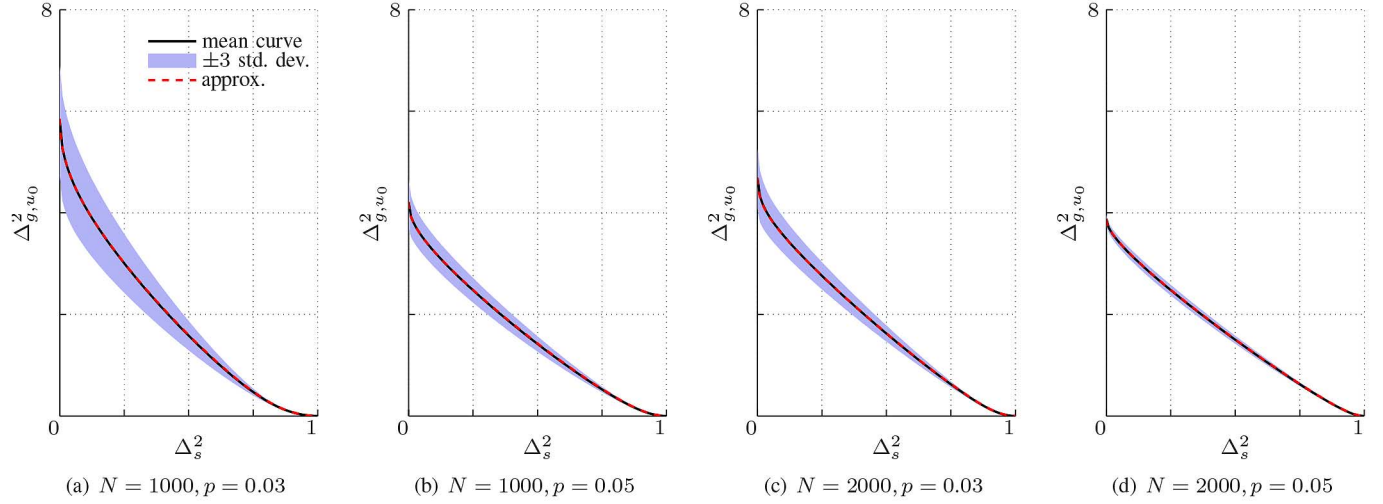


Fig. 6. Uncertainty curves for Erdős-Rényi graphs. For each choice of  $(N, p)$  parameters, 1000 Erdős-Rényi graphs were generated and their uncertainty curves for  $s \in [0, 1]$  were computed using the sandwich approximation procedure described in Section III. The geodesic distance function is used. Each curve was interpolated to generate comparable curves on a regular grid. For each parameter choice, the mean and standard deviation of the interpolated curve were computed over the ensemble. The mean curve is plotted on the graphs as a solid line, with shaded areas illustrating the three standard deviation levels. Meanwhile, the approximate expected value computed before generating the curves is plotted as a dashed red line. The shape of the uncertainty curve is clearly quite stable across each ensemble, especially as  $N$  and  $p$  increase, and the approximate expectation curve is quite accurate. (a)  $N = 1000, p = 0.03$ . (b)  $N = 1000, p = 0.05$ . (c)  $N = 2000, p = 0.03$ . (d)  $N = 2000, p = 0.05$ .

for  $0 \leq s \leq \frac{N}{N-1}$ . Thus, the curve is the entire lower half of the ellipse given by (31). When the graph is large (i.e.,  $N \gg 1$ ), this curve converges to a straight line  $\gamma_{u_0}(s) = 1 - s$  in the  $s$ - $g$  plane.

### B. Star Graphs

A star graph [21] with  $N$  vertices has one central vertex and  $N-1$  leaves, each connected by a single edge to the center. It is a prototypical example of a hub in a network [34]. The Laplacian matrix can be expressed in block form as

$$\mathbf{L} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{N-1}} \mathbf{1}_{N-1}^T \\ -\frac{1}{\sqrt{N-1}} \mathbf{1}_{N-1} & \mathbf{I}_{N-1} \end{pmatrix} \quad (33)$$

where  $\mathbf{1}_{N-1}$  is the  $(N-1)$ -vector of all ones, and  $\mathbf{I}_{N-1}$  is the  $(N-1) \times (N-1)$  identity matrix. Since the graph is bipartite, the largest eigenvalue of  $\mathbf{L}$  is always equal to 2 [30]. Let  $u_0$  be the center of the star; the diagonal distance matrix is again given by  $\mathbf{P}_{u_0} = \text{diag}\{0, 1, 1, \dots, 1\}$ .

Just as for the complete graph, we can always represent signals that achieve the uncertainty curve on star graphs as  $\tilde{\mathbf{x}}(\theta) = [\cos \theta, \frac{\sin \theta}{\sqrt{N-1}}, \frac{\sin \theta}{\sqrt{N-1}}, \dots, \frac{\sin \theta}{\sqrt{N-1}}]^T$  for some  $\theta$  (see the remark in Appendix C for justification). Now, the graph spread is given by  $\Delta_{g,u_0}^2 = \sin^2 \theta = \frac{1}{2} - \frac{1}{2} \cos 2\theta$ ; again, by using (6), the spectral spread can be computed as

$$\begin{aligned} \Delta_s^2 &= (N-1) \left( \frac{\cos \theta}{\sqrt{N-1}} - \frac{\sin \theta}{\sqrt{N-1}} \right)^2 \\ &= 1 - \sin 2\theta. \end{aligned}$$

The lower bound curve is thus the lower part of the ellipse defined by

$$(\Delta_s^2 - 1)^2 + (2\Delta_{g,u_0}^2 - 1)^2 = 1.$$

Written explicitly, the curve is

$$\gamma_{u_0}(s) = \frac{1}{2} \left( 1 - \sqrt{s(2-s)} \right), \quad \text{for } 0 \leq s \leq 2. \quad (34)$$

We note that, unlike the complete graph case, this curve does not depend on the size of the graph.

### C. Erdős-Rényi Random Graphs

An Erdős-Rényi random graph  $G$  is generated by taking  $N$  vertices and selecting each pair of vertices to be an edge with probability  $p$ , independent of all other potential edges. We denote by  $\mathcal{G}_p(N, p)$  the statistical ensemble of the resulting graphs. First studied by Erdős and Rényi [27], [28],  $\mathcal{G}_p(N, p)$  may be the simplest random graph model. Although they do not capture all of the behaviors of real networks, Erdős-Rényi graphs are an excellent theoretical model because they lend themselves to tractable analysis.

To study the properties of the uncertainty curves for Erdős-Rényi graphs, we generated several realizations drawn from  $\mathcal{G}_p(N, p)$  and used the approximation algorithm described in Section III-C to compute their uncertainty curves. It quickly emerged that the curves for different realizations generated with the same parameters were, for reasonable sizes of  $N$ , tightly clustered around a common mean curve. This is illustrated in Fig. 6, which shows the mean curves and estimated standard deviations for several parameter values. In what follows, we develop an analytic approximation for computing the expected (i.e., mean) uncertainty curve for different choices of parameters  $N$  and  $p$ .

Recall from the definition of the uncertainty curve that we are trying to approximate the expectation of

$$\begin{aligned} \gamma_{u_0}(s) &= \min_{\mathbf{x} \in \ell^2(G)} \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x} \text{ subject to } \|\mathbf{x}\|^2 = 1 \text{ and } \mathbf{x}^T \mathbf{L} \mathbf{x} = s \end{aligned} \quad (35)$$

over random graphs drawn from  $\mathcal{G}_p(N, p)$ . The matrices  $\mathbf{P}_{u_0}^2$  and  $\mathbf{L}$  and the optimal vector  $\mathbf{x}$  that solves the minimization problem are all random quantities. Since  $\gamma_{u_0}(s)$  is obtained through a nonconvex quadratic program, there is generally no closed-form expressions linking  $\gamma_{u_0}(s)$  to  $\mathbf{P}_{u_0}^2$  and  $\mathbf{L}$ . As a result, directly computing the expectation of  $\gamma_{u_0}(s)$  will be difficult. To make the problem tractable, we proceed by replacing  $\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}$  and  $\mathbf{x}^T \mathbf{L} \mathbf{x}$  in (35) with their respective expected values and minimizing after the fact. Later, we will see that this strategy turns out to be very effective in generating accurate approximations.

Another observation that emerged from our numerical experiment was a characteristic of the vectors that achieved the bound with  $s \leq 1$ : these vectors were all approximately radial functions, i.e., the value at any vertex  $v$  was a function of  $d(u_0, v)$ . Because this simplifies the analysis greatly, we will only consider the part of the curve with  $s \leq 1$ , which corresponds to signals that are maximally localized in both the graph and spectral domains. We will explicitly incorporate this assumption by focusing on vectors whose values depend only on distance from  $u_0$ . In this case, the original  $N$ -dimensional vector  $\mathbf{x} \in \ell^2(G)$  can be represented by a smaller vector  $\mathbf{y}$ , with  $x(v) = y(d(u_0, v))$ . The dimensionality of  $\mathbf{y}$  is equal to  $\mathcal{E}_G(u_0) + 1$ , where  $\mathcal{E}_G(u_0)$  is the eccentricity of the center vertex. We note that  $\mathcal{E}_G(u_0)$  is a random variable that in principle can take any value between 0 and  $N - 1$ . When  $N$  is large, however, we can find a small number  $d_{\max} \sim \mathcal{O}(\log N)$  such that  $\mathcal{E}_G(u_0) \leq d_{\max}$  with high probability [35]. So, in what follows, we will treat  $\mathbf{y}$  as a vector in  $\mathbb{R}^{d_{\max}+1}$ .

For a given, deterministic  $\mathbf{y}$ , we will compute the expectations (over the randomness of the graph model) of  $\|\mathbf{x}\|^2$  and  $\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}$ . To that end, we define  $f_d$  as the probability that a vertex  $v$  chosen uniformly at random from  $V \setminus \{u_0\}$  has a distance  $d(u_0, v) = d$ . The special case  $f_1 = p$  is easy to verify. For the other cases, we will use the results of Blondel *et al.* [36], who developed a recursive formula<sup>4</sup> to find (approximate) analytical expressions of the entire sequence  $\{f_d\}$ . The expected number of vertices at a distance  $d \geq 1$  is  $(N - 1)f_d$ . It follows that for fixed  $\mathbf{y}$

$$\begin{aligned} \mathbb{E} [\|\mathbf{x}\|^2] &= \mathbb{E} \left[ \sum_{v \in V} y(d(u_0, v))^2 \right] \\ &\approx y^2(0) + \sum_{k=1}^{d_{\max}} (N - 1) f_k y^2(k) \end{aligned} \quad (36)$$

and

$$\begin{aligned} \mathbb{E} [\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}] &= \mathbb{E} \left[ \sum_{v \in V} d(u_0, v)^2 x(v)^2 \right] \\ &\approx \sum_{k=1}^{d_{\max}} k^2 (N - 1) f_k y^2(k) \end{aligned} \quad (37)$$

<sup>4</sup>Unlike our construction, they allowed  $v$  to be any vertex in  $V$ , including  $u_0$ ; thus, in their result,  $f_0 = \frac{1}{N}$ , and all other values of  $f_d$  differ from ours by a factor of  $\frac{N-1}{N}$ . For large  $N$ , the difference is negligible.

where the approximations are due to the truncation of  $\mathbf{y}$  at dimension  $d_{\max}$ .

The spectral spread is more complicated. We start with the expression

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{u \sim v} \left( \frac{x(u)}{\sqrt{\deg(u)}} - \frac{x(v)}{\sqrt{\deg(v)}} \right)^2.$$

By assuming that the degree of every vertex is approximately equal to its expectation  $(N - 1)p$ , we write

$$\mathbf{x}^T \mathbf{L} \mathbf{x} \approx \frac{1}{(N - 1)p} \sum_{u \sim v} (x(u) - x(v))^2. \quad (38)$$

Recall that  $x(v) = y(d(u_0, v))$ . Consequently, the only edges that contribute to (38) are those between vertices at *different* distances from  $u_0$ . Since a vertex at distance  $k$  can only be connected to vertices at a distance of  $k - 1$  and  $k + 1$ , we simply need to characterize  $M_{k,k+1}$ , the expected number of edges from vertices at a distance  $k$  to vertices at a distance  $k + 1$ , for  $k = 0$  to  $d_{\max} - 1$ . The expected value of the spectral spread can then be obtained as

$$\mathbb{E} [\mathbf{x}^T \mathbf{L} \mathbf{x}] \approx \frac{1}{(N - 1)p} \sum_{k=0}^{d_{\max}-1} M_{k,k+1} (y(k+1) - y(k))^2. \quad (39)$$

It is easy to see that  $M_{0,1} = (N - 1)p$ , since that is simply the expected number of edges incident upon  $u_0$ . The other terms of  $M_{k,k+1}$  can be approximated through a recurrence relation. First, we observe that the expected number of vertices at distance  $k$  is  $(N - 1)f_k$  and the expected number of vertices *not* at distance  $k$  (not counting  $u_0$ ) is  $(N - 1)(1 - f_k)$ . Thus, we can approximate that the total number of *potential* edges between these two disjoint sets of vertices is  $(N - 1)^2 f_k (1 - f_k)$ . Since each potential edge will be chosen with probability  $p$ , we get that  $M_{k-1,k} + M_{k,k+1} \approx (N - 1)^2 p f_k (1 - f_k)$ , which leads to the following approximate recurrence relation:

$$\begin{cases} M_{0,1} = (N - 1)p \\ M_{k,k+1} \approx (N - 1)^2 p f_k (1 - f_k) - M_{k-1,k}, \text{ for } k \geq 1. \end{cases} \quad (40)$$

The expressions in (36), (37), and (39) show that the expected values of the squared norm, graph spread, and spectral spread are all nonnegative quadratic forms involving the vector  $\mathbf{y} \in \mathbb{R}^{d_{\max}+1}$ . It follows that we can write

$$\begin{aligned} \mathbb{E} [\|\mathbf{x}\|^2] &\approx \mathbf{y}^T \mathbf{H}_a \mathbf{y}, \quad \mathbb{E} [\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}] \approx \mathbf{y}^T \mathbf{P}_a^2 \mathbf{y}, \text{ and} \\ \mathbb{E} [\mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x}] &\approx \mathbf{y}^T \mathbf{L}_a \mathbf{y} \end{aligned} \quad (41)$$

for some positive semidefinite matrices  $\mathbf{H}_a, \mathbf{P}_a^2, \mathbf{L}_a$ , respectively. Substituting these expectations for their (random) counterparts in (35), we compute our approximation of the expected uncertainty curve  $\tilde{\gamma}_{u_0}(s)$  as

$$\begin{aligned} \tilde{\gamma}_{u_0}(s) &= \\ &\min_{\mathbf{y} \in \mathbb{R}^{d_{\max}+1}} \mathbf{y}^T \mathbf{P}_a^2 \mathbf{y} \text{ subject to } \mathbf{y}^T \mathbf{H}_a \mathbf{y} = 1 \text{ and } \mathbf{y}^T \mathbf{L}_a \mathbf{y} = s. \end{aligned} \quad (42)$$

We note that this minimization problem (a quadratic program with quadratic constraints) has exactly the same mathematical structure as the one previously studied in (17). Using the same techniques derived in Section III-B, we can show that any solution to (42) satisfies the (generalized) eigenvalue problem

$$(\mathbf{P}_a^2 - \alpha \mathbf{L}_a) \mathbf{y} = \tau_{\min}(\alpha) \mathbf{H}_a \mathbf{y} \quad (43)$$

for some value of  $\alpha$ , where  $\tau_{\min}(\alpha)$  is the smallest (generalized) eigenvalue. As earlier, we can construct a sandwich approximation to the curve by solving (43) for a sequence of  $\alpha$ 's.

Despite the various approximations made along the way, the analytical solution obtained in (42) fits experiment remarkably well. As illustrated in Fig. 6, the resulting analytic curves (shown in dashed lines) match almost perfectly with the observed sample average (shown in solid lines). We note that the matrices in (42) are of size  $d_{\max} \times d_{\max}$ , which is much smaller than  $N \times N$ . For example, for the  $\mathcal{G}_p(10^6, 10^{-4})$  model, we would have  $d_{\max} = 4$  (the smallest  $d$  such that  $1 - \sum_{k=1}^d f_k < 10^{-7}$ ).

Thus, the analytic approximation derived here can be computed far faster than the actual uncertainty curve for any realization of the model, and does not itself require any realization to be generated.

## V. DIFFUSION PROCESSES AND UNCERTAINTY BOUNDS

In constructing dictionaries to represent signals on graphs, one would like the dictionary elements to be localized in both graph and spectral domains. Quantifying the signal localization in these two domains and studying their fundamental tradeoff have been one of the motivations of this work. To test the theoretical results and the computational algorithm presented in Section III, we consider two graph wavelet transforms in the literature: the diffusion wavelets of Coifman and Maggioni [4] and the spectral graph wavelet transform of Hammond *et al.* [6]. The localization properties of these two constructions are studied on a graph visualized in Fig. 7(a) based on the network of football games played in the 2000 regular season by NCAA Division I-A teams [37]. While the spectral graph wavelet transform does not downsample the graph, the diffusion wavelet transform does. In our experiment, the center vertex  $u_0$  is chosen to be one of the vertices that remain in the downsampled graph at the coarsest level of the diffusion wavelet transform.

Fig. 7(b) shows several scaling functions from both constructions plotted against the uncertainty curve  $\gamma_{u_0}(s)$ , with the latter obtained by using the sandwich algorithm in Section III-C. In this and all subsequent experiments, we use eight refinement iterations (for a total of 257 sparse eigenvalue evaluations) to plot the uncertainty curves. At this level, we find the lower and upper approximations of  $\gamma_{u_0}(s)$  to be visually indistinguishable. As predicted, both the spectral graph wavelet and diffusion wavelet constructions result in basis elements that obey the computed bound. In fact, they follow the curve quite well.

The diffusion wavelets are based on the evolution of a discrete time diffusion process on a graph. In the classical setting, where the signal domain is the real line, there is a strong connection between the *continuous time* diffusion process and the

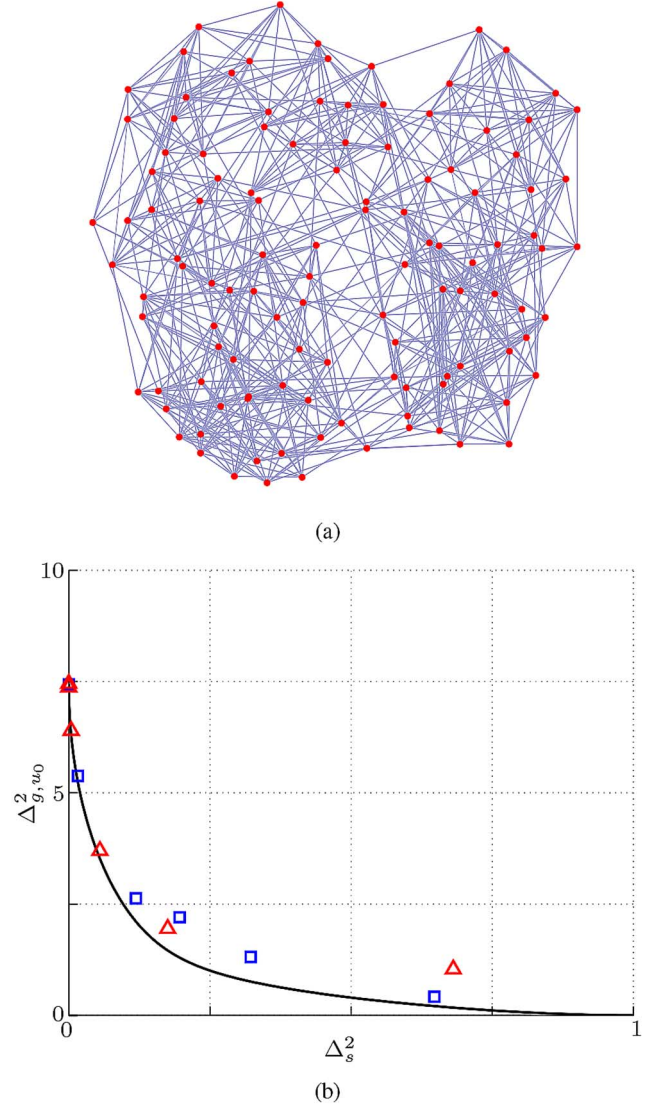


Fig. 7. (a) Network of football games between NCAA Division I-A teams in the 2000 regular season [37]. (b) Spectral spread versus graph spread on this graph. (Solid line) Computed uncertainty curve  $\gamma_{u_0}(s)$ . (Triangles) Scaling functions in diffusion wavelets [4]. (Squares) Scaling functions in spectral graph wavelet transform (SGWT) [6]. (The true SGWT scaling functions are not related to the wavelet functions by a two-scale relation; here, we simply take the cumulative sum of the coarsest-level scaling function and higher level wavelet functions.)

Heisenberg uncertainty curve: to see this, consider a diffusion (i.e., heat) equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial y^2} \quad (44)$$

where  $u(y, t)$  is a function of  $y, t \in \mathbb{R}$ . This equation governs the conduction of heat in physical processes, and its solution was the original motivation for Fourier analysis. The fundamental solution to (44), i.e., the solution with the initial condition that  $u(y, 0) = \delta(y - y_0)$  for a given  $y_0$ , is the Gaussian kernel

$$K(t, y, y_0) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{(y-y_0)^2}{4t}}.$$

Thus, if we start with an impulse and evolve according to (44), at time  $t$ , we get a function with time spread  $t$  and frequency

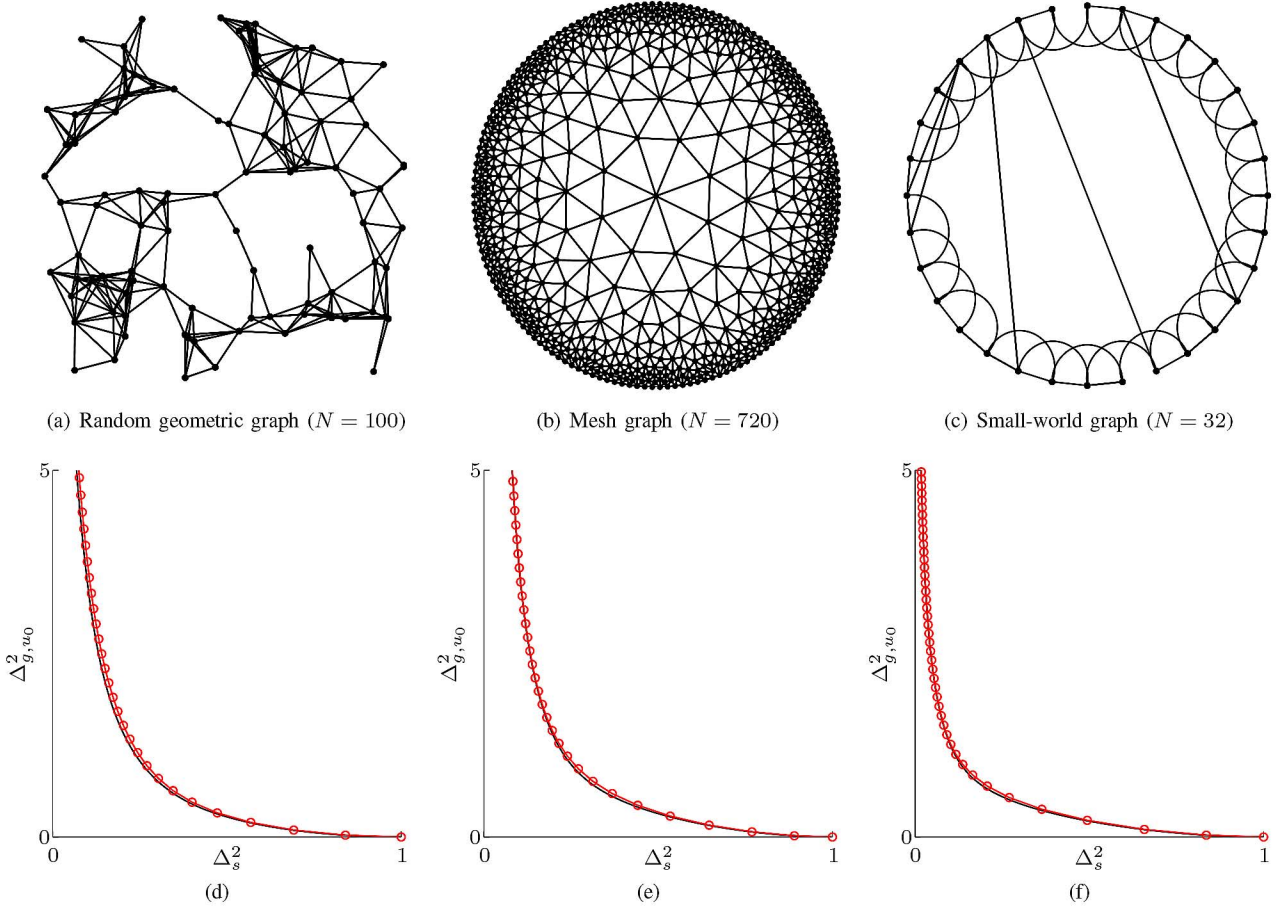


Fig. 8. Diffusion process versus the uncertainty curve for three types of graph. (a) Random geometric graph [38]. (b) Triangular mesh [39]. (c) Small-world graph [40]. (d)–(f) Associated uncertainty curves shown below the graphs (solid black line). A continuous-time diffusion process is run on each graph, beginning with an impulse at one vertex, and the resulting spreads are plotted (solid red line with circles). The circles are evenly spaced in time. The diffusion process tracks the curve closely, though close examination reveals that the match is not exact.

spread  $\frac{1}{4t}$ , achieving the classical Heisenberg uncertainty  $\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}$  with equality. In other words, the diffusion kernels on the real line are exactly the signals that achieve the time–frequency uncertainty bound.

This line of thought motivated us to consider a continuous-time diffusion process on graphs, governed by an equation analogous to (44)

$$\frac{d\mathbf{x}}{dt} = -\mathbf{L}\mathbf{x} \quad (45)$$

where  $\mathbf{L}$  is the graph Laplacian. With the initial condition  $\mathbf{x}(0) = \delta_{u_0}$ , the solution to (45) is [5]

$$\mathbf{x}(t) = e^{-t\mathbf{L}}\delta_{u_0} = \sum_{i=1}^N e^{-t\lambda_i} \mathbf{f}_i \mathbf{f}_i^T \delta_{u_0} \quad (46)$$

where  $e^{-t\mathbf{L}}$  is the matrix exponential of  $\mathbf{L}$ ,  $\{\lambda_i\}$  are the eigenvalues of  $\mathbf{L}$ , and  $\{\mathbf{f}_i\}$  are the corresponding eigenvectors. Denote by  $\eta_{u_0}(s)$  the curve in the  $s$ – $g$  plane traced out by the diffusion process. The curve can be given in parametric form as

$$\begin{cases} s(t) = \frac{\mathbf{x}(t)^T \mathbf{L} \mathbf{x}(t)}{\|\mathbf{x}(t)\|^2} \\ \eta_{u_0}(s) = \frac{\mathbf{x}(t)^T \mathbf{P}_{u_0}^2 \mathbf{x}(t)}{\|\mathbf{x}(t)\|^2} \end{cases} \quad (47)$$

We show in Appendix E that  $s(t)$  is a strictly decreasing function of  $t$ ; therefore, it is one to one. Furthermore,  $s(0) = 1$  and  $\lim_{t \rightarrow \infty} s(t) = 0$ . All together, this guarantees that the function  $\eta_{u_0}(s)$  is well defined for every  $s \in (0, 1]$ .

We plot in Fig. 8 the diffusion curve  $\eta_{u_0}(s)$  and the uncertainty curve  $\gamma_{u_0}(s)$  for three different graphs: a random geometric graph [38] that can capture the connectivity of wireless sensor networks; an unstructured triangular mesh<sup>5</sup> for finite element analysis [39]; and a small-world graph [40] that serves as the mathematical model for social and various other empirical networks. The geodesic distance function is used. In all three cases, the spreads of the diffusion process, though not exactly achieving the bounds as in the classical setting, match the uncertainty curves remarkably well.

The following proposition, proved in Appendix D, asserts that for certain special graphs, the match between  $\eta_{u_0}(s)$  and  $\gamma_{u_0}(s)$  is exact.

**Proposition 4:** For all  $s \in (0, 1]$ ,  $\eta_{u_0}(s) = \gamma_{u_0}(s)$  if (a)  $G$  is a complete graph with  $N$  vertices and  $u_0$  is any vertex; or (b)  $G$  is a star graph with  $N$  vertices and  $u_0$  is the vertex with degree  $N - 1$ .

<sup>5</sup>This graph was generated using the Mesh2D MATLAB toolbox written by D. Engwirda, available online at MATLAB Central (<http://www.mathworks.com/matlabcentral/fileexchange/25555>).



For general graphs, we can show that, under certain conditions, the low-order derivatives of the uncertainty curve and the diffusion curve match.

*Proposition 5:* Let  $G$  be any connected graph and  $u_0$  be any vertex on  $G$ . Then,  $\eta_{u_0}(1) = \gamma_{u_0}(1) = 0$ ,  $\frac{d\eta_{u_0}}{ds}\Big|_{s=1} = \frac{d\gamma_{u_0}}{ds}\Big|_{s=1} = 0$ , and

$$\frac{d^2\eta_{u_0}}{ds^2}\Big|_{s=1} = \frac{\deg u_0}{2 \sum_{v \sim u_0} \frac{1}{d(v, u_0)^2 \deg v}} \leq \frac{d^2\gamma_{u_0}}{ds^2}\Big|_{s=1} = \frac{\deg u_0}{2} \frac{\sum_{v \sim u_0} \frac{d(v, u_0)^2}{\deg v}}{\left(\sum_{v \sim u_0} \frac{1}{\deg v}\right)^2} \quad (48)$$

with equality if and only if  $d(v, u_0)$  is identical for every  $v \sim u_0$ .

This proposition is proved in Appendix E. It is easy to verify that the geodesic distance satisfies the condition required for equality in (48). Extrapolating the observations in Fig. 8 and results in Propositions 4 and 5 leads us to believe that diffusion kernels on arbitrary graphs will always be close to optimal in graph and spectral localizations. We leave further rigorous study of this tantalizing conjecture as an important line of future work.

## VI. CONCLUSION

Analogous to the classical Heisenberg uncertainty principle in time–frequency analysis, an uncertainty principle for signals defined on graphs was developed in this work. After presenting quantitative definitions of the signal “spreads” in the graph and spectral domains, we provided a complete characterization of the feasibility region achieved by these two quantities. The lower boundary of the region, which is analogous to the classical uncertainty bound (1), was shown to be achieved by eigenvectors associated with the smallest eigenvalues of a particular matrix-valued function. Furthermore, the convexity of the uncertainty curve allows it to be efficiently approximated by solving a sequence of eigenvalue problems. We derived closed-form formulas of the uncertainty curves for complete graphs and star graphs, and developed a fast analytical approximation for the expected uncertainty curve for Erdős–Rényi random graphs. The localization properties of two existing wavelet transforms were evaluated. Finally, numerical experiments and analytical results led us to an intriguing connection between diffusion processes on graphs and the uncertainty bounds.

## APPENDIX A

We would like to prove that the set  $\mathcal{D}_{u_0}$  is convex as long as the number of vertices  $N \geq 3$ . (The need for such a condition will be made clear shortly.) This is equivalent to showing the following result.

*Proposition 6:* Suppose that there exist two vectors  $\mathbf{x}_1, \mathbf{x}_2$  in  $\mathbb{R}^N$  with  $N \geq 3$ , such that

$$\mathbf{x}_i^T \mathbf{x}_i = 1, \quad \mathbf{x}_i^T \mathbf{L} \mathbf{x}_i = s_i, \quad \text{and} \quad \mathbf{x}_i^T \mathbf{P}_{u_0}^2 \mathbf{x}_i = g_i, \quad \text{for } i = 1, 2. \quad (49)$$

Then for any  $\beta \in [0, 1]$ , we can always find a vector  $\mathbf{x}$  in  $\mathbb{R}^N$  satisfying

$$\mathbf{x}^T \mathbf{x} = 1, \quad \mathbf{x}^T \mathbf{L} \mathbf{x} = s, \quad \text{and} \quad \mathbf{x}^T \mathbf{P}_{u_0}^2 \mathbf{x} = g \quad (50)$$

where  $s \stackrel{\text{def}}{=} \beta s_1 + (1 - \beta) s_2$  and  $g \stackrel{\text{def}}{=} \beta g_1 + (1 - \beta) g_2$ .

We will prove the aforementioned proposition by recasting the problem in  $\text{Sym}_N$ , the Hilbert space of real, symmetric  $N \times N$  matrices. The space is endowed with the Hilbert–Schmidt inner product defined by  $\langle \mathbf{A}, \mathbf{B} \rangle_{\text{HS}} \stackrel{\text{def}}{=} \text{tr}(\mathbf{A}^T \mathbf{B}) = \text{tr}(\mathbf{A} \mathbf{B})$ , where  $\text{tr}(\cdot)$  is the trace of a matrix. Every  $\mathbf{x} \in \mathbb{R}^N$  can be mapped onto a matrix  $\mathbf{X} = \mathbf{x} \mathbf{x}^T$  in  $\text{Sym}_N$ . Finding a vector  $\mathbf{x}$  satisfying the conditions in (50) then boils down to finding a rank-one positive semidefinite matrix  $\mathbf{X} = \mathbf{x} \mathbf{x}^T$  satisfying the following three constraints:

$$\text{tr}(\mathbf{X}) = 1, \quad \text{tr}(\mathbf{L} \mathbf{X}) = s, \quad \text{and} \quad \text{tr}(\mathbf{P}_{u_0}^2 \mathbf{X}) = g. \quad (51)$$

The requirement that  $\mathbf{X}$  be a rank-one matrix makes this a hard problem, because the cone of rank-one matrices is not convex. Instead, we will use the following theorem to relax the problem to the cone of positive semidefinite matrices  $\mathcal{S}_+^N$ , which is convex.

*Theorem 3 (Barvinok [41]):* Suppose that  $R > 0$  and  $N \geq R + 2$ . Let  $\mathcal{H} \subset \text{Sym}_N$  be an affine subspace such that  $\text{codim}(\mathcal{H}) \leq \binom{R+2}{2}$ . If the intersection  $\mathcal{S}_+^N \cap \mathcal{H}$  is nonempty and bounded, then there is a matrix  $\mathbf{X}$  in  $\mathcal{S}_+^N \cap \mathcal{H}$  such that  $\text{rank}(\mathbf{X}) \leq R$ .

*Proof of Proposition 6:* First, we note that the three equalities in (51) are all affine constraints on  $\mathbf{X}$ . Together, they define a hyperplane  $\mathcal{H} \subset \text{Sym}_N$  with  $\text{codim}(\mathcal{H}) \leq 3 = \binom{1+2}{2}$ . (In fact,  $\mathbf{I}, \mathbf{L}$ , and  $\mathbf{P}_{u_0}^2$  are linearly independent, so  $\text{codim}(\mathcal{H}) = 3$ .) To apply Theorem 3, we verify next that  $\mathcal{S}_+^N \cap \mathcal{H}$  is nonempty and bounded.

First we show that it is bounded: let  $\mathbf{X}$  be an arbitrary matrix in the intersection  $\mathcal{S}_+^N \cap \mathcal{H}$  (assuming one exists), and let  $\{\nu_1, \nu_2, \dots, \nu_N\}$  be its eigenvalues. The equalities  $1 = \text{tr}(\mathbf{X}) = \sum_{n=1}^N \nu_n$ , together with the nonnegativity of the eigenvalues, imply that

$$\|\mathbf{X}\|_{\text{HS}}^2 = \text{tr}(\mathbf{X}^2) = \sum_{n=1}^N \nu_n^2 \leq \sum_{n=1}^N \nu_n = 1.$$

Therefore,  $\mathcal{S}_+^N \cap \mathcal{H}$  is a subset of the unit ball in  $\text{Sym}_N$  and is thus bounded.

To show that  $\mathcal{S}_+^N \cap \mathcal{H}$  is nonempty, we explicitly construct a member of the set. Let  $\mathbf{x}_1, \mathbf{x}_2$  be the two vectors satisfying (49). On mapping the vectors to two matrices  $\mathbf{X}_1 \stackrel{\text{def}}{=} \mathbf{x}_1 \mathbf{x}_1^T$  and  $\mathbf{X}_2 \stackrel{\text{def}}{=} \mathbf{x}_2 \mathbf{x}_2^T$ , the constraints in (49) can be rewritten as

$$\text{tr}(\mathbf{X}_i) = 1, \quad \text{tr}(\mathbf{L} \mathbf{X}_i) = s_i \quad \text{and} \quad \text{tr}(\mathbf{P}_{u_0}^2 \mathbf{X}_i) = g_i, \quad \text{for } i = 1, 2.$$

$\mathbf{X}_1$  and  $\mathbf{X}_2$  are both in  $\mathcal{S}_+^N$ . Now set  $\mathbf{X}' = \beta \mathbf{X}_1 + (1 - \beta) \mathbf{X}_2$ . It is easy to see that  $\mathbf{X}' \in \mathcal{H}$  and, because  $\mathcal{S}_+^N$  is convex,  $\mathbf{X}' \in \mathcal{S}_+^N$  as well. To be sure, the matrix  $\mathbf{X}' \in \mathcal{S}_+^N \cap \mathcal{H}$  is not necessarily of rank one. However, the result of Theorem 3 (for the case when  $R = 1$ ) guarantees the existence of a rank one matrix  $\mathbf{X}$  in  $\mathcal{S}_+^N \cap \mathcal{H}$ . Decomposing this matrix as  $\mathbf{X} = \mathbf{x} \mathbf{x}^T$  and using the equivalence between (50) and (51), we can conclude that the resulting vector  $\mathbf{x}$  satisfies all the constraints in (50). ■

*Remark:* The proof of proposition 6 uses Theorem 3 for the case when  $R = 1$ . Consequently, we need to work with



$N \geq R + 2 = 3$ . This requirement is sharp in that the achievable region  $\mathcal{D}_{u_0}$  for a graph with two vertices (i.e.,  $N = 2$ ) is not convex. The only connected graph with  $N = 2$  is the complete graph. All unit-norm signals on this graph can be parametrized as  $(\cos \theta, \sin \theta)$ . By computing the corresponding graph Laplacian and distance matrices, it is easy to show that the achievable region is only the boundary of an ellipse (not including its interior) and hence is not convex.

#### APPENDIX B

- a) For any  $\alpha_1 < \alpha_2$ , let  $\mathbf{v}_1$  and  $\mathbf{v}_2$  be two unit-norm eigenvectors in  $\mathcal{S}(\alpha_1)$  and  $\mathcal{S}(\alpha_2)$ , respectively. Applying Rayleigh's inequality, we get  $\mathbf{v}_2^T \mathbf{M}(\alpha_1) \mathbf{v}_2 \geq q(\alpha_1) = \mathbf{v}_1^T \mathbf{M}(\alpha_1) \mathbf{v}_1$ . Similarly, we have  $-\mathbf{v}_2^T \mathbf{M}(\alpha_2) \mathbf{v}_2 \geq -\mathbf{v}_1^T \mathbf{M}(\alpha_2) \mathbf{v}_1$ . A combination of these two inequalities leads to

$$\mathbf{v}_2^T (\mathbf{M}(\alpha_1) - \mathbf{M}(\alpha_2)) \mathbf{v}_2 \geq \mathbf{v}_1^T (\mathbf{M}(\alpha_1) - \mathbf{M}(\alpha_2)) \mathbf{v}_1. \quad (52)$$

Recall that  $\mathbf{M}(\alpha) = \mathbf{P}_{u_0}^2 - \alpha \mathbf{L}$ , and therefore,  $\mathbf{M}(\alpha_1) - \mathbf{M}(\alpha_2) = (\alpha_2 - \alpha_1) \mathbf{L}$ . Replacing this identity into (52), we thus have

$$\mathbf{v}_2^T \mathbf{L} \mathbf{v}_2 \geq \mathbf{v}_1^T \mathbf{L} \mathbf{v}_1.$$

Note that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  can be arbitrary unit-norm elements in  $\mathcal{S}(\alpha_1)$  and  $\mathcal{S}(\alpha_2)$ , respectively. If, in particular, we choose  $\mathbf{v}_1, \mathbf{v}_2$  to be those that attain the maximization in (21), we get  $h_+(\alpha_2) = \mathbf{v}_2^T \mathbf{L} \mathbf{v}_2 \geq \mathbf{v}_1^T \mathbf{L} \mathbf{v}_1 = h_+(\alpha_1)$ . Similarly, we can show that  $h_-(\alpha_2) \geq h_-(\alpha_1)$ .

- b) We will only consider the limits when  $\alpha$  tends to  $-\infty$  as given in (23). The other case, when  $\alpha$  tends to  $+\infty$ , can be analyzed in a similar way, and its proof will be omitted. Let  $\alpha > 0$  be any positive number. By definition

$$h_+(\alpha) \geq h_-(\alpha) \geq 0 \quad (53)$$

where the second inequality is due to the Laplacian matrix  $\mathbf{L}$  being positive semidefinite. Next, we show that  $h_+(\alpha)$  can be made arbitrarily close to 0 as  $\alpha \rightarrow -\infty$ . To that end, let  $\mathbf{v}$  be any unit-norm eigenvector in  $\mathcal{S}(\alpha)$ , and  $\mathbf{f}_1$  be the first eigenvector of  $\mathbf{L}$  as defined in (4). Since  $\mathcal{S}(\alpha)$  is associated with the smallest eigenvalue  $q(\alpha)$ , we have from Rayleigh's inequality

$$\mathbf{v}^T (\mathbf{P}_{u_0}^2 - \alpha \mathbf{L}) \mathbf{v} \leq \mathbf{f}_1^T (\mathbf{P}_{u_0}^2 - \alpha \mathbf{L}) \mathbf{f}_1 = \mathbf{f}_1^T \mathbf{P}_{u_0}^2 \mathbf{f}_1$$

with the equality coming from the identity  $\mathbf{L} \mathbf{f}_1 = 0$ . For any  $\alpha < 0$ , rearranging the aforementioned expression leads to

$$\mathbf{v}^T \mathbf{L} \mathbf{v} \leq -\frac{1}{\alpha} (\mathbf{f}_1^T \mathbf{P}_{u_0}^2 \mathbf{f}_1 - \mathbf{v}^T \mathbf{P}_{u_0}^2 \mathbf{v}) \leq -\frac{\mathcal{E}_G^2(u_0)}{\alpha} \quad (54)$$

where the second inequality uses the bound of the graph spread as provided in Proposition 1. Since (54) holds for any nonzero element  $\mathbf{v}$  from  $\mathcal{S}(\alpha)$ , we must have  $h_+(\alpha) \leq -\mathcal{E}_G^2(u_0)/\alpha$ , which, when combined with (53), completes the proof.

- c) First, using eigenvalue perturbation results, we will derive a candidate set  $\mathcal{A}$  of points such that  $q(\alpha)$  is certainly analytic on  $[a, b] \setminus \mathcal{A}$ . We will show that  $\mathcal{A}$  is finite

so that the set of nonanalytic points of  $q(\alpha)$  is finite as well. Then, we will compute  $h_-(\alpha)$  and  $h_+(\alpha)$  explicitly, and show that they are left- and right-continuous, respectively, and that they are equal to the negative left- and right-derivatives of  $q(\alpha)$ , respectively. We will then show that  $h_-(\alpha) = h_+(\alpha)$  everywhere except a subset  $\mathcal{B} \subseteq \mathcal{A}$ ; therefore, they satisfy (24). Since  $\mathcal{A}$  is finite, it follows that  $\mathcal{B}$  is finite as well.

The starting point of our analysis is the following result.

*Proposition 7:* There exist  $N$  analytic functions  $\lambda_1(\cdot), \dots, \lambda_N(\cdot)$  and  $N$  analytic vector-valued functions  $\mathbf{x}_1(\cdot), \dots, \mathbf{x}_N(\cdot)$  such that

$$\mathbf{M}(\alpha) \mathbf{x}_i(\alpha) = \lambda_i(\alpha) \mathbf{x}_i(\alpha) \quad (55)$$

and  $\mathbf{x}_i(\alpha)^T \mathbf{x}_j(\alpha) = \delta_{ij}$ .

*Proof:* Standard perturbation results [31, p. 404] guarantee the existence of such functions for any matrix function that is analytic and whose value is always Hermitian. The function  $\mathbf{M}(\cdot)$  as defined in (18) is affine in  $\alpha$ , and thus analytic; it is symmetric and real for every  $\alpha$ , and thus Hermitian. Therefore, functions with the properties listed in the proposition do exist. ■

From Proposition 7, we can write  $q(\alpha)$  as

$$q(\alpha) = \min_{1 \leq i \leq N} \lambda_i(\alpha) \quad (56)$$

where  $\{\lambda_i(\cdot)\}_i$  are the eigenvalue functions guaranteed by the proposition. For any  $\alpha_0 \in \mathbb{R}$ , if  $\mathcal{S}(\alpha_0)$  has dimension one, then precisely one of the eigenvalue functions is equal to  $q(\cdot)$  at  $\alpha_0$ , say  $\lambda_k(\alpha_0) = q(\alpha_0)$ . Pick some  $\varepsilon < \frac{1}{2} \min_{j \neq k} |\lambda_j(\alpha_0) - \lambda_k(\alpha_0)|$ . Since every  $\lambda_j(\cdot)$  is analytic, we can find some neighborhood  $\mathcal{N}$  of  $\alpha_0$  for which  $|\lambda_j(\alpha) - \lambda_j(\alpha_0)| < \varepsilon$  for every  $j$ . This guarantees that  $\lambda_k(\alpha) < \lambda_j(\alpha)$  on  $\mathcal{N}$  for every  $j \neq k$ . Thus,  $q(\alpha) = \lambda_k(\alpha)$  on  $\mathcal{N}$ . Since  $\lambda_k(\cdot)$  is analytic on  $\mathcal{N}$ , we have that  $q(\cdot)$  is analytic on  $\mathcal{N}$  and therefore at  $\alpha_0$ . We can make this more general. Suppose instead of only one eigenvalue function attaining the minimum at  $\alpha_0$ , there are multiple eigenvalue functions (e.g., two, denoted by  $\lambda_{k_1}(\cdot)$  and  $\lambda_{k_2}(\cdot)$ ) that attain the minimum, and that they are all equal on a neighborhood  $\mathcal{N}$  of  $\alpha_0$ . All the other eigenvalue functions are larger at  $\alpha_0$ . Again, the analyticity allows us to find a neighborhood  $\mathcal{N}' \subseteq \mathcal{N}$  on which all the other eigenvalue functions are larger than  $\lambda_{k_1}(\cdot) = \lambda_{k_2}(\cdot)$ . Now, since  $q(\alpha) = \lambda_{k_1}(\alpha) = \lambda_{k_2}(\alpha)$ , the function  $q(\alpha)$  is analytic on  $\mathcal{N}'$  as well.

Thus, a necessary condition for  $q(\cdot)$  to be nonanalytic at  $\alpha_0$  is that two (or more) *distinct* eigenvalue functions must intersect at  $\alpha_0$ . Define  $\mu_j(\cdot), j = 1, \dots, N', N' \leq N$  as the set of distinct eigenvalue functions, and let  $n_j$  be the multiplicity of the eigenvalue function  $\mu_j(\cdot)$ . Now consider an arbitrary finite interval  $[a, b]$  and define

$$\mathcal{A} = \bigcup_{1 \leq i < j \leq N'} \{\alpha \in [a, b] : \mu_i(\alpha) = \mu_j(\alpha)\}.$$

It is a well-known property of analytic functions that if they are equal on more than a finite set of points in an interval, then they are identical. Since the  $\mu_i(\cdot)$  are distinct analytic functions, no two of them can be equal on more than a finite set of points in  $[a, b]$ . Thus,  $\mathcal{A}$  is the finite union of finite sets, and therefore contains only a finite number of points

Next, we connect  $q(\alpha)$  to  $h_+(\alpha)$  and  $h_-(\alpha)$ . At any point  $\alpha_0 \in [a, b]$ , there can be  $k \geq 1$  distinct eigenvalue functions that achieve the minimum in (56). Without loss of generality, we shall assume they are the first  $k$  functions,  $\mu_1(\cdot), \dots, \mu_k(\cdot)$ . The associated eigenvectors,  $\mathbf{x}_{ij}(\alpha_0)$ , for  $i = 1, \dots, k$  and  $j = 1, \dots, n_i$ , form an orthonormal basis for the eigenspace  $\mathcal{S}(\alpha_0)$ . Any unit-norm element  $\mathbf{v} \in \mathcal{S}(\alpha_0)$  can then be written as  $\mathbf{v} = \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mathbf{x}_{ij}(\alpha_0)$ , for some constant coefficients  $\{c_{ij}\}$  satisfying  $\sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij}^2 = 1$ .

We now define an analytic function  $\mathbf{v}(\alpha) \stackrel{\text{def}}{=} \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mathbf{x}_{ij}(\alpha)$ , with  $\mathbf{v}(\alpha_0) = \mathbf{v}$ . The eigenvalue identity in (55) implies that  $\mathbf{M}(\alpha)\mathbf{v}(\alpha) = \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mu_i(\alpha) \mathbf{x}_{ij}(\alpha)$ . Differentiating both sides of this equality yields

$$\mathbf{M}'(\alpha)\mathbf{v}(\alpha) + \mathbf{M}(\alpha)\mathbf{v}'(\alpha) = \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mu_i'(\alpha) \mathbf{x}_{ij}(\alpha) + \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mu_i(\alpha) \mathbf{x}_{ij}'(\alpha). \quad (57)$$

Evaluating (57) at  $\alpha = \alpha_0$ , premultiplying it by  $\mathbf{v}^T(\alpha_0)$  and using the substitutions  $\mathbf{M}'(\alpha) = -\mathbf{L}$ ,  $\mathbf{M}(\alpha_0)\mathbf{v}(\alpha_0) = q(\alpha_0)\mathbf{v}(\alpha_0)$ , and  $\mu_i(\alpha_0) = q(\alpha_0)$  for every  $i$ , we get

$$-\mathbf{v}^T(\alpha_0)\mathbf{L}\mathbf{v}(\alpha_0) + q(\alpha_0)\mathbf{v}^T(\alpha_0)\mathbf{v}'(\alpha_0) = \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij}^2 \mu_i'(\alpha_0) + q(\alpha_0) \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij} \mathbf{v}^T(\alpha_0) \mathbf{x}_{ij}'(\alpha_0). \quad (58)$$

The second terms on the left-hand and right-hand sides of (58) are equal, leaving us with

$$\mathbf{v}^T(\alpha_0)\mathbf{L}\mathbf{v}(\alpha_0) = - \sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij}^2 \mu_i'(\alpha_0). \quad (59)$$

By definition,  $h_+(\alpha_0)$  and  $h_-(\alpha_0)$  are the two extreme values of  $\mathbf{v}^T(\alpha_0)\mathbf{L}\mathbf{v}(\alpha_0)$ . Maximizing (and minimizing) the quantity in (59) subject to the unit-norm constraint  $\sum_{i=1}^k \sum_{j=1}^{n_i} c_{ij}^2 = 1$ , we have

$$h_+(\alpha_0) = \max_{1 \leq i \leq k} (-\mu_i'(\alpha_0)) \text{ and } h_-(\alpha_0) = \min_{1 \leq i \leq k} (-\mu_i'(\alpha_0)). \quad (60)$$

Now, there must exist some  $\ell, m \in \{1, \dots, k\}$  such that

$$q(\alpha) = \begin{cases} \mu_\ell(\alpha) & \text{if } \alpha \leq \alpha_0 \\ \mu_m(\alpha) & \text{if } \alpha \geq \alpha_0 \end{cases} \quad (61)$$

on some neighborhood  $\mathcal{N}$  of  $\alpha_0$ , which can be chosen to be small enough that  $\mathcal{N} \cap \mathcal{A} = \{\alpha_0\}$  if  $\alpha_0 \in \mathcal{A}$  or  $\mathcal{N} \cap \mathcal{A} = \emptyset$  otherwise. We must have  $\mu_m'(\alpha_0) = \min_{1 \leq i \leq k} \mu_i'(\alpha_0)$ , since if  $\mu_j'(\alpha_0) < \mu_m'(\alpha_0)$  for some  $j$ , then on a sufficiently small neighborhood of  $\alpha_0$ , we would have  $q(\alpha) = \mu_j(\alpha) < \mu_m(\alpha)$  for  $\alpha > \alpha_0$ , contradicting (61).<sup>6</sup> Meanwhile, away from  $\alpha_0$  there are no other points in  $\mathcal{N}$  at which multiple distinct eigenvalue functions intersect. Thus, from (60), we have that  $h_+(\alpha) = -\mu_m'(\alpha)$  on  $\mathcal{N} \cap [\alpha_0, \infty)$ . Since the  $\mu_i(\cdot)$  are all

<sup>6</sup>The requirement that  $\mu_m'(\alpha_0) = \min_{1 \leq i \leq k} \mu_i'(\alpha_0)$  might not always be sufficient to uniquely determine  $m$ , however. In the case that multiple distinct eigenvalue functions achieve the minimum derivative,  $\mu_m(\cdot)$  is then determined by comparing the higher order derivatives. This nuance does not affect our proof, which only depends on the first derivative.

analytic,  $h_+(\alpha)$  is right-continuous at  $\alpha_0$ . Furthermore, since  $q(\alpha) = \mu_m(\alpha)$  on  $\mathcal{N} \cap [\alpha_0, \infty)$ ,  $h_+(\alpha_0)$  is equal to the negative right-derivative of  $q(\alpha)$  at  $\alpha_0$ . By similar arguments, we can show that  $h_-(\alpha)$  is left-continuous at  $\alpha_0$  and is equal to the negative left-derivative of  $q(\alpha)$  at  $\alpha_0$ .

A necessary condition for  $h_-(\alpha_0) \neq h_+(\alpha_0)$  is that  $k > 1$ , i.e., there are multiple distinct eigenvalue functions achieving the minimum in (56). Thus, the set of points  $\mathcal{B}$  at which they differ satisfies  $\mathcal{B} \subseteq \mathcal{A}$ , so  $\mathcal{B}$  is finite. Meanwhile, if  $h_+(\alpha_0) = h_-(\alpha_0)$ , then the equality must hold for all  $\alpha \in \mathcal{N}$  as well because of the way we constructed the neighborhood  $\mathcal{N}$ . Since  $h_-(\alpha)$  is left-continuous and  $h_+(\alpha)$  is right-continuous at  $\alpha_0$ , both functions are continuous at  $\alpha_0$ . Equality also means the left- and right-derivatives of  $q(\alpha)$  are equal at  $\alpha_0$ , and thus,  $q'(\alpha_0)$  is well defined with  $h_+(\alpha_0) = h_-(\alpha_0) = -q'(\alpha_0)$ .

## APPENDIX C

For  $N = 2$ , the proposition is trivial, so let us assume  $N > 2$ . By Theorem 1,  $\tilde{\mathbf{x}}$  must be an eigenvector associated with the smallest eigenvalue of  $\mathbf{M}(\alpha) = \mathbf{P}_{u_0}^2 - \alpha \mathbf{L}$  for some  $\alpha$ , where  $\mathbf{L}$  and  $\mathbf{P}_{u_0}$  are given by (27) and (28), respectively.  $\mathbf{M}(\alpha)$  is given in block form as

$$\mathbf{M}(\alpha) = \left[ \begin{array}{c|c} -\alpha & \frac{\alpha}{N-1} \mathbf{1}_{N-1}^T \\ \hline \frac{\alpha}{N-1} \mathbf{1}_{N-1} & \mathbf{B} \end{array} \right]$$

where  $\mathbf{B}$  is the  $(N-1) \times (N-1)$  circulant matrix  $\mathbf{B} = \left(1 - \frac{N}{N-1}\alpha\right) \mathbf{I}_{N-1} + \frac{\alpha}{N-1} \mathbf{1}_{N-1} \mathbf{1}_{N-1}^T$ . Let  $\{\mathbf{w}_1, \dots, \mathbf{w}_{N-2}\}$  be an orthonormal set of vectors in  $\mathbb{R}^{N-1}$  such that  $\mathbf{w}_i \perp \mathbf{1}_{N-1}$ . This set spans the subspace of vectors in  $\mathbb{R}^{N-1}$  orthogonal to  $\mathbf{1}_{N-1}$ . It is easy to verify that  $\mathbf{B}\mathbf{w}_i = \left(1 - \frac{N}{N-1}\alpha\right) \mathbf{w}_i$ . Furthermore, if we set  $\mathbf{v}_i = (0, \mathbf{w}_i^T)^T$ , then we can see that  $\mathbf{v}_i$  are all eigenvectors of  $\mathbf{M}(\alpha)$  with eigenvalue  $1 - \frac{N}{N-1}\alpha$ .

If we can show that this is not the smallest eigenvalue of  $\mathbf{M}(\alpha)$ , i.e., that  $q(\alpha) \neq 1 - \frac{N}{N-1}\alpha$ , then it follows that  $\tilde{\mathbf{x}}$  (an eigenvector of  $\mathbf{M}(\alpha)$  corresponding to  $q(\alpha)$ ) must be orthogonal to every  $\mathbf{v}_i$  for  $i = 1, \dots, N-2$ . This will then guarantee that  $\tilde{\mathbf{x}}$  is of the form (29).

To show that  $q(\alpha) \neq 1 - \frac{N}{N-1}\alpha$ , we let  $\mathbf{y} = [y_1, \dots, y_N]$  be chosen such that  $\|\mathbf{y}\| = 1$ ,  $y_1 \neq 0$ , and  $\mathbf{y}^T \mathbf{1}_N = 0$ . This last property makes  $\mathbf{y}$  an eigenvector of  $\mathbf{L}$  with eigenvalue  $\frac{N}{N-1}$ . We have  $\mathbf{y}^T \mathbf{P}_{u_0}^2 \mathbf{y} = 1 - y_1^2$  and  $\mathbf{y}^T \mathbf{L} \mathbf{y} = \frac{N}{N-1}$ . Thus,  $\mathbf{y}^T \mathbf{M}(\alpha) \mathbf{y} = 1 - y_1^2 - \frac{N}{N-1}\alpha < 1 - \frac{N}{N-1}\alpha$ . It follows from the Rayleigh inequality that  $q(\alpha) \leq \mathbf{y}^T \mathbf{M}(\alpha) \mathbf{y} < 1 - \frac{N}{N-1}\alpha$ , proving the proposition.

*Remark:* With small modifications, this proof can be used to demonstrate that the same property holds for star graphs, i.e., any vector achieving the uncertainty curve must be of the form in (29). For a star graph with  $N$  vertices, we have

$$\mathbf{M}(\alpha) = \left[ \begin{array}{c|c} -\alpha & \frac{\alpha}{\sqrt{N-1}} \mathbf{1}_{N-1}^T \\ \hline \frac{\alpha}{\sqrt{N-1}} \mathbf{1}_{N-1} & (1 - \alpha) \mathbf{I}_{N-1} \end{array} \right]. \quad (62)$$

Again, there is an  $(N-2)$ -dimensional eigenspace spanned by the same set  $\{\mathbf{v}_i\}_{i=1}^{N-2}$  as in the complete graph case above. In this case, the eigenvalue associated with that subspace is  $1 - \alpha$ . Thus, to show that the smallest eigenvector is of the desired

form, we must simply show that there is some unit norm vector  $\mathbf{y}$  for which  $\mathbf{y}^T \mathbf{M}(\alpha) \mathbf{y} < 1 - \alpha$ , guaranteeing that the eigenvector associated with the smallest eigenvalue is orthogonal to the eigenspace spanned by  $\{\mathbf{v}_i\}_{i=1}^{N-1}$ . Our test vector here is  $\mathbf{y} = (1, 0, \dots, 0)^T$ , which gives us  $\mathbf{y}^T \mathbf{M}(\alpha) \mathbf{y} = -\alpha < 1 - \alpha$ , so the same property holds for the star graph as the complete graph.

#### APPENDIX D

- a) Let  $\{\mathbf{f}_1, \dots, \mathbf{f}_N\}$  be an orthonormal basis of  $\mathbb{R}^N$  with  $\mathbf{f}_1 = \frac{1}{\sqrt{N}} \mathbf{1}_N$ . It is easy to verify that these are eigenvectors of  $\mathbf{L}$  [given in (27)] with corresponding eigenvalues  $\lambda_1 = 0$  and  $\lambda_k = \frac{N}{N-1}$  for  $k = 2, \dots, N$ . It follows from (46) that the diffusion process starting from  $\mathbf{x}_0 = \delta_{u_0}$  can be obtained as

$$\mathbf{x}(t) = \mathbf{f}_1 \mathbf{f}_1^T \delta_{u_0} + e^{-\lambda_2 t} (\mathbf{I} - \mathbf{f}_1 \mathbf{f}_1^T) \delta_{u_0}. \quad (63)$$

Assuming without loss of generality that  $u_0 = 1$  and using the fact that  $\mathbf{f}_1 = \frac{1}{\sqrt{N}} \mathbf{1}$ , we have  $\mathbf{x}(t) = \frac{1}{N} [1 + (N-1)e^{-\lambda_2 t}, 1 - e^{-\lambda_2 t}, \dots, 1 - e^{-\lambda_2 t}]^T$ . Using our knowledge of  $\mathbf{L}$  and the fact that  $\mathbf{P}_{u_0} = \text{diag}(0, 1, \dots, 1)$ , it is now straightforward to compute the spreads as  $\Delta_s^2(\mathbf{x}(t)) = \frac{Ne^{-2\lambda_2 t}}{1 + (N-1)e^{-2\lambda_2 t}}$  and  $\Delta_{g,u_0}^2(\mathbf{x}(t)) = \frac{\frac{N-1}{N} (1 - e^{-\lambda_2 t})^2}{1 + (N-1)e^{-2\lambda_2 t}}$ . We can verify that these spreads satisfy (32). Thus, for all  $t \geq 0$ ,  $\mathbf{x}(t)$  achieves the uncertainty curve.  $\Delta_s^2(\mathbf{x}(t))$  is continuous and  $\lim_{t \rightarrow \infty} \Delta_s^2(\mathbf{x}(t)) = 0$ , so  $\eta_{u_0}(s) = \gamma_{u_0}(s)$  for  $s \in (0, 1]$ .

- b) Here, we assume without loss of generality that the center of the star, i.e., the vertex with degree  $N-1$  is  $u_0 = 1$ . Again, we explicitly construct an orthonormal eigenbasis for  $\mathbf{L}$ , given in this case by (33). In what follows, we will assume that  $N > 2$ ; the star graph with two vertices is the same as the complete graph with two vertices, so the proof from (a) will apply to that case. Let  $\mathbf{f}_1 = \left[ \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2(N-1)}}, \dots, \frac{1}{\sqrt{2(N-1)}} \right]^T$ ,  $\mathbf{f}_N = \left[ -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2(N-1)}}, \dots, \frac{1}{\sqrt{2(N-1)}} \right]^T$ , and  $\mathbf{f}_k = [0, \mathbf{g}_k^T]^T$  for  $k = 2, \dots, N-1$ , where  $\{\mathbf{g}_k\}_{k=1}^{N-1}$  is any orthonormal basis for  $\mathbb{R}^{N-1}$  satisfying  $\mathbf{g}_1 = \frac{1}{\sqrt{N-1}} \mathbf{1}_{N-1}$ . It is easy to verify that  $\{\mathbf{f}_k\}_{k=1}^N$  forms an orthonormal basis for  $\mathbb{R}^N$ , and that the  $\mathbf{f}_k$  are eigenvectors of  $\mathbf{L}$  with corresponding eigenvalues  $\lambda_1 = 0, \lambda_2 = \dots = \lambda_{N-1} = 1$ , and  $\lambda_N = 2$ .

Similar to (63), we can compute the diffusion process explicitly as

$$\mathbf{x}(t) = \mathbf{f}_1 \mathbf{f}_1^T \delta_{u_0} + e^{-t} (\mathbf{I} - \mathbf{f}_1 \mathbf{f}_1^T - \mathbf{f}_N \mathbf{f}_N^T) \delta_{u_0} + e^{-2t} \mathbf{f}_N \mathbf{f}_N^T \delta_{u_0} \quad (64)$$

$$= (1 - e^{-t}) \mathbf{f}_1 \mathbf{f}_1^T \delta_{u_0} + (e^{-2t} - e^{-t}) \mathbf{f}_N \mathbf{f}_N^T \delta_{u_0} + e^{-t} \delta_{u_0}. \quad (65)$$

Using the expressions for  $\mathbf{f}_1$  and  $\mathbf{f}_N$ , we find that  $\mathbf{x}(t) = \left[ \frac{1}{2} + \frac{1}{2}e^{-2t}, \frac{1}{2\sqrt{N-1}}(1 - e^{-2t}), \mathbf{1}_{N-1}^T \right]^T$ . From this, we can compute the graph spread as  $\Delta_{g,u_0}^2(\mathbf{x}(t)) = \frac{(1 - e^{-2t})^2}{2(1 + e^{-4t})}$  and the spectral spread as  $\Delta_s^2(\mathbf{x}(t)) = \frac{2e^{-4t}}{1 + e^{-4t}}$ . It is easy to verify that these spreads satisfy (34), and so  $\mathbf{x}(t)$  achieves the uncertainty curve for  $t \geq 0$ . Once again,  $\Delta_s^2(\mathbf{x}(t))$  is continuous and  $\lim_{t \rightarrow \infty} \Delta_s^2(\mathbf{x}(t)) = 0$ , so  $\eta_{u_0}(s) = \gamma_{u_0}(s)$  for  $s \in (0, 1]$ .

#### APPENDIX E

We know from Theorem 1 that every point on the uncertainty curve is achieved by an eigenvector associated with the smallest eigenvalue  $q(\alpha)$  of  $\mathbf{M}(\alpha) = \mathbf{P}_{u_0}^2 - \alpha \mathbf{L}$ . In particular, the point  $(1, 0)$  is achieved by  $\delta_{u_0}$ , which is the eigenvector associated with the matrix  $\mathbf{M}(0) = \mathbf{P}_{u_0}^2$  and eigenvalue  $q(0) = 0$ . Since  $d(u_0, v) = 0$  if and only if  $u_0 = v$  and  $d(v, u_0) > 0$  otherwise, the eigenspace  $\mathcal{S}(0)$  is one-dimensional. Thus, from the proof of Lemma 1 in Appendix B, there is some neighborhood  $\mathcal{N}$  of  $\alpha = 0$  on which  $\mathcal{S}(\alpha)$  is one-dimensional, and therefore,  $q(\alpha)$  is analytic. In this case, there exists some neighborhood of  $s = 1$  for which we can use the parametric form of the uncertainty curve given in (25), namely  $(s, \gamma_{u_0}(s)) = (s_u(\alpha), g_u(\alpha))$  where  $s_u(\alpha) = -q'(\alpha)$  and  $g_u(\alpha) = q(\alpha) - \alpha q'(\alpha)$  for  $\alpha \in \mathcal{N}$ .

We can thus compute the derivative of the uncertainty curve parametrically as

$$\frac{d\gamma_{u_0}}{ds} = \frac{g'_u(\alpha)}{s'_u(\alpha)} = \frac{-\alpha q''(\alpha)}{-q''(\alpha)} = \alpha \quad (66)$$

where  $\alpha$  is chosen so that  $s(\alpha)$  is the argument at which we wish to evaluate the derivative. Similarly, the second derivative is

$$\frac{d^2\gamma_{u_0}}{ds^2} = \frac{\frac{d}{d\alpha} \left( \frac{g'_u(\alpha)}{s'_u(\alpha)} \right)}{s'_u(\alpha)} = \frac{1}{-q''(\alpha)}. \quad (67)$$

Both (66) and (67) require that  $q''(\alpha)$  be nonzero. In what follows, we will explicitly compute  $q''(0)$  and show that  $q''(\alpha) \neq 0$  for  $\alpha \in \mathcal{N}'$ , where  $\mathcal{N}' \subseteq \mathcal{N}$ . As described in the proof of Lemma 1, there is an analytic eigenvector function  $\mathbf{v}(\alpha)$  defined in a neighborhood of  $\alpha = 0$  such that

$$\mathbf{M}(\alpha) \mathbf{v}(\alpha) = q(\alpha) \mathbf{v}(\alpha) \quad (68)$$

with  $\mathbf{v}(0) = \delta_{u_0}$  and  $\|\mathbf{v}(\alpha)\|^2 = 1$ . The spectral spread function is given by  $s_u(\alpha) = \mathbf{v}(\alpha) \mathbf{L} \mathbf{v}(\alpha) = -q'(\alpha)$ , where the second equality is due to (25). So we can compute

$$q''(\alpha) = -2\mathbf{v}(\alpha) \mathbf{L} \mathbf{v}'(\alpha). \quad (69)$$

To compute  $\mathbf{v}'(\alpha)$ , we differentiate both sides of (68) and after rearranging terms obtain

$$[\mathbf{M}(\alpha) - q(\alpha) \mathbf{I}] \mathbf{v}'(\alpha) = \mathbf{L} \mathbf{v}(\alpha) + q'(\alpha) \mathbf{v}(\alpha). \quad (70)$$

From (68) and the fact that  $\mathcal{S}(\alpha)$  is 1-D on  $\mathcal{N}$ ,  $\mathbf{M}(\alpha) - q(\alpha) \mathbf{I}$  has a one-dimensional nullspace spanned by  $\mathbf{v}(\alpha)$ . Since  $0 = \frac{d}{d\alpha} \|\mathbf{v}(\alpha)\|^2 = 2\mathbf{v}(\alpha)^T \mathbf{v}'(\alpha)$ , when we multiply both sides of

(70) by the Moore–Penrose pseudoinverse of  $\mathbf{M}(\alpha) - q(\alpha)\mathbf{I}$ , we obtain

$$\mathbf{v}'(\alpha) = [\mathbf{M}(\alpha) - q(\alpha)\mathbf{I}]^+ \mathbf{L}\mathbf{v}(\alpha) \quad (71)$$

where we have also used the fact that  $[\mathbf{M} - q(\alpha)\mathbf{I}]^+ \mathbf{v}(\alpha) = 0$  to simplify the right-hand side of (71).

Setting  $\alpha = 0$  and using the fact that  $q(0) = 0$  and  $\mathbf{v}(0) = \delta_{u_0}$ , we have  $\mathbf{v}'(0) = (\mathbf{P}_{u_0}^2)^+ \mathbf{L}\delta_{u_0}$ . Substituting this into (69), we get  $q''(0) = -2\delta_{u_0}^T \mathbf{L} (\mathbf{P}_{u_0}^2)^+ \mathbf{L}\delta_{u_0} = -2 \sum_{v \sim u_0} \frac{(\mathbf{L}\delta_{u_0})_v^2}{d^2(v, u_0)}$ , where  $(\mathbf{L}\delta_{u_0})_v$  is the  $v$ th entry of  $\mathbf{L}\delta_{u_0}$ . From the definition of the graph Laplacian, we have that for every  $v \sim u_0$ ,  $(\mathbf{L}\delta_{u_0})_v = \frac{-1}{\sqrt{\deg u_0}} \frac{1}{\sqrt{\deg v}}$ . Thus

$$q''(0) = \frac{-2}{\deg u_0} \left( \sum_{v \sim u_0} \frac{1}{d(v, u_0)^2 \deg v} \right). \quad (72)$$

Since the graph is connected,  $q''(0) \neq 0$ , and since  $q(\alpha)$  is analytic on  $\mathcal{N}$ , there exists a neighborhood  $\mathcal{N}' \subseteq \mathcal{N}$  containing 0 on which  $q''(\alpha) \neq 0$  as well. Thus, our expressions for the first and second derivatives (66) and (67) are valid at  $s = 1$ , which corresponds to  $\alpha = 0$ . We obtain  $\frac{d\gamma_{u_0}}{ds} \Big|_{s=1} = 0$  and the expression for  $\frac{d^2\gamma_{u_0}}{ds^2} \Big|_{s=1}$  given in (48).

To compute the derivatives of the curve  $\eta_{u_0}(s)$  traced out by the diffusion process  $\mathbf{x}(t)$ , we express it parametrically in terms of  $t$ , with  $(s, \eta_{u_0}(s)) = (s_d(t), g_d(t))$  where  $s_d(t) = \frac{\mathbf{x}(t)^T \mathbf{L}\mathbf{x}(t)}{\|\mathbf{x}(t)\|^2}$  and  $g_d(t) = \frac{\mathbf{x}(t)^T \mathbf{P}_{u_0}^2 \mathbf{x}(t)}{\|\mathbf{x}(t)\|^2}$ .

We first show that  $\dot{s}_d(t) < 0$ . To simplify the computation of this and other derivatives, we introduce the function  $R_{\mathbf{Z}}(t) = \frac{\mathbf{x}(t)^T \mathbf{Z} \mathbf{x}(t)}{\|\mathbf{x}(t)\|^2}$  for any fixed matrix  $\mathbf{Z}$ . It is easy to verify that since  $\dot{\mathbf{x}}(t) = -\mathbf{L}\mathbf{x}(t)$ ,  $\dot{R}_{\mathbf{Z}}(t) = 2R_{\mathbf{Z}}(t)R_{\mathbf{L}}(t) - R_{\mathbf{L}\mathbf{Z}}(t) - R_{\mathbf{Z}\mathbf{L}}(t)$ , where the last two terms in the sum are equal if  $\mathbf{Z}$  is symmetric. Since we have an explicit solution  $\mathbf{x}(t) = e^{-\mathbf{L}t}\delta_{u_0}$ , we can see that  $\|\mathbf{x}(t)\| \neq 0$  for all  $t$  so that  $R_{\mathbf{Z}}(t)$  and its derivative is well defined.

Since  $s_d(t) = R_{\mathbf{L}}(t)$ , we have  $\dot{s}_d(t) = 2(s_d(t)^2 - R_{\mathbf{L}^2}(t)) = 2 \left[ (\mathbf{x}(t)^T \mathbf{L}\mathbf{x}(t))^2 - \mathbf{x}(t)^T \mathbf{L}^2 \mathbf{x}(t) \right] < 0$  by the Cauchy–Schwarz inequality. Equality would hold only if  $\mathbf{L}\mathbf{x}(t)$  were a multiple of  $\mathbf{x}(t)$ —i.e., if  $\mathbf{x}(t)$  were an eigenvector. From (46), we can see that this could only occur if  $\delta_{u_0}$  itself were an eigenvector, which is impossible for a connected graph. We can directly evaluate  $s_d(0) = 1$  and  $\lim_{t \rightarrow \infty} s_d(t) = 0$ ; combining this with the aforementioned result guarantees that  $s_d(t)$  is a one-to-one function with range  $(0, 1]$ . Thus,  $\eta_{u_0}(s)$  is well defined on that domain.

Since  $g_d(t) = R_{\mathbf{P}_{u_0}^2}(t)$ , we can compute the derivative  $\dot{g}_d(0) = g_d(0)s_d(0) - R_{\mathbf{L}\mathbf{P}_{u_0}^2}(0) = 0$ . Thus, the diffusion curve's derivative at  $s = 1$  is given by

$$\frac{d\eta_{u_0}}{ds} \Big|_{s=1} = \frac{\dot{g}_d(t)}{\dot{s}_d(t)} = 0 = \frac{d\gamma_{u_0}}{ds} \Big|_{s=1}. \quad (73)$$

Meanwhile, we can simplify the second derivative evaluated at  $s = 1$ , obtaining

$$\frac{d^2\eta_{u_0}}{ds^2} \Big|_{s=1} = \frac{\ddot{g}_d(0)}{\dot{s}_d^2(0)}. \quad (74)$$

The first derivative of  $s_d(t)$  at  $t = 0$  can be computed as

$$\begin{aligned} \dot{s}_d(0) &= 2(s_d(0)^2 - R_{\mathbf{L}^2}(0)) \\ &= 2(1 - \|\mathbf{L}\delta_{u_0}\|^2) = -2 \sum_{v \sim u_0} \frac{1}{\deg u_0 \deg v}. \end{aligned} \quad (75)$$

The second derivative of  $g_d(t)$  is

$$\begin{aligned} \ddot{g}_d(t) &= 2\dot{g}_d(t)s_d(t) + 2g_d(t)\dot{s}_d(t) - 4s_d(t)R_{\mathbf{L}\mathbf{P}_{u_0}^2}(t) \\ &\quad + 2R_{\mathbf{L}^2\mathbf{P}_{u_0}^2}(t) + 2R_{\mathbf{L}\mathbf{P}_{u_0}^2\mathbf{L}}(t). \end{aligned} \quad (76)$$

At  $t = 0$ , the only nonzero term in (76) is the last one

$$\begin{aligned} \ddot{g}_d(0) &= 2R_{\mathbf{L}\mathbf{P}_{u_0}^2\mathbf{L}}(0) = 2\delta_{u_0}^T \mathbf{L}\mathbf{P}_{u_0}^2 \mathbf{L}\delta_{u_0} \\ &= 2 \sum_{v \sim u_0} \frac{d(v, u_0)^2}{\deg u_0 \deg v}. \end{aligned} \quad (77)$$

Now we can combine (74), (75), and (77) to obtain the expression for  $\frac{d^2\eta_{u_0}}{ds^2} \Big|_{s=1}$  given in (48). By the Cauchy–Schwarz inequality,  $\left( \sum_{v \sim u_0} \frac{1}{\deg v} \right)^2 \leq \left( \sum_{v \sim u_0} \frac{d(v, u_0)^2}{\deg v} \right) \left( \sum_{v \sim u_0} \frac{1}{d(v, u_0)^2 \deg v} \right)$  with equality if and only if  $d(v, u_0)^2 = c$  for every  $v \sim u_0$ , where  $c$  is some constant. Comparing the expressions for the second derivatives of the uncertainty curve and diffusion curve, we can see that  $\frac{d^2\eta_{u_0}}{ds^2} \Big|_{s=1} \geq \frac{d^2\gamma_{u_0}}{ds^2} \Big|_{s=1}$ , with equality if and only if  $d(v, u_0)$  is identical for every  $v \sim u_0$ .

#### ACKNOWLEDGMENT

We thank the associate editor and the anonymous referees, especially referee #1, for their constructive criticism and valuable comments, which corrected several errors in the original manuscript and greatly improved the presentation of this paper.

#### REFERENCES

- [1] G. B. Folland and A. Sitaram, “The uncertainty principle: A mathematical survey,” *J. Fourier Anal. Appl.*, vol. 3, no. 3, pp. 207–238, May 1997.
- [2] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.
- [3] A. Buades, B. Coll, and J. Morel, “A non-local algorithm for image denoising,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, pp. 60–65.
- [4] R. R. Coifman and M. Maggioni, “Diffusion wavelets,” *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 53–94, Jul. 2006.
- [5] R. Coifman and S. Lafon, “Diffusion maps,” *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.
- [6] D. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Apr. 2010.

- [7] U. A. Khan, S. Kar, and J. M. Moura, "Higher dimensional consensus: Learning in large-scale networks," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2836–2849, May 2010.
- [8] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [9] I. Pesenson, "Sampling in Paley–Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.
- [10] S. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter-banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.
- [11] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.
- [12] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*, 1st ed. New York, NY, USA: Springer-Verlag, 2009.
- [13] R. Ishii and K. Furukawa, "The uncertainty principle in discrete signals," *IEEE Trans. Circuits Syst.*, vol. 33, no. 10, pp. 1032–1034, Oct. 1986.
- [14] L. C. Calvez and P. Vilbé, "On the uncertainty principle in discrete signals," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 39, no. 6, pp. 394–395, Jun. 1992.
- [15] S. S. Goh and T. N. Goodman, "Uncertainty principles and asymptotic behavior," *Appl. Comput. Harmon. Anal.*, vol. 16, no. 1, pp. 19–43, Jan. 2004.
- [16] E. Wolfgang, "Uncertainty principles on compact Riemannian manifolds," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 2, pp. 182–197, Sep. 2010.
- [17] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Math. Anal.*, vol. 49, no. 3, pp. 906–931, 1989.
- [18] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [19] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. Comput. Math.*, vol. 6, pp. 227–254, Dec. 2005.
- [20] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [21] C. Godsil and G. Royle, *Algebraic Graph Theory*. New York, NY, USA: Springer-Verlag, 2001.
- [22] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *Learning Theory*. Berlin, Germany: Springer, 2004, vol. 3120, pp. 624–638.
- [23] M. Belkin, "Problems of learning on manifolds," Ph.D. dissertation, Univ. Chicago, Chicago, IL, USA, 2003.
- [24] A. Agaskar and Y. M. Lu, "An uncertainty principle for functions defined on graphs," *SPIE Conf. Wavelets Sparsity XIV*, 2011.
- [25] A. Agaskar and Y. M. Lu, "Uncertainty principles for signals defined on graphs: Bounds and characterizations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 3493–3496.
- [26] G. Rote, "The convergence rate of the sandwich algorithm for approximating convex functions," *Computing*, vol. 48, no. 3–4, pp. 337–361, Sep. 1992.
- [27] P. Erdős and A. Rényi, "On random graphs, I," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.
- [28] P. Erdős and A. Rényi, *On the Evolution of Random Graphs*. Budapest, Hungary: Akad. Kiadó, 1960.
- [29] D. Klein and M. Randić, "Resistance distance," *J. Math. Chem.*, vol. 12, no. 1, pp. 81–95, 1993.
- [30] F. R. K. Chung, *Spectral Graph Theory*. Providence, RI, USA: Amer. Math. Soc., 1997.
- [31] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: With Applications*, 2nd ed. New York, NY, USA: Academic, 1985.
- [32] R. S. Strichartz, *The Way of Analysis*. Burlington, VT, USA: Jones & Bartlett Learning, 2000.
- [33] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [34] M. Newman, *Networks: An Introduction*, 1st ed. New York, NY, USA: Oxford Univ. Press, 2010.
- [35] R. Albert and A. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [36] V. Blondel, J. Guillaume, J. Hendrickx, and R. Jungers, "Distance distribution in random graphs and application to network exploration," *Phys. Rev. E*, vol. 76, no. 6, pp. 066101-1–066101-8, 2007.
- [37] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci.*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [38] M. Penrose, *Random Geometric Graphs*. Oxford, U.K.: Oxford Univ. Press, 2003.
- [39] G. Strang and G. J. Fix, *An Analysis of the Finite Element Method*, 2nd ed. Cambridge, MA, USA: Wellesley Cambridge, 2008.
- [40] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [41] A. Barvinok, "A remark on the rank of positive semidefinite matrices subject to affine constraints," *Discr. Comput. Geom.*, vol. 25, no. 1, pp. 23–31, 2001.

**Ameya Agaskar** (S'10) received the B.S. degree in engineering physics and the M.Eng. degree in electrical and computer engineering from Cornell University in 2007 and 2008, respectively. In 2008 he joined the Advanced Sensor Techniques group at MIT Lincoln Laboratory, where he worked on adaptive signal processing for radar and communication applications. In 2010 he was awarded a Lincoln Doctoral Scholarship and began doctoral studies at the Harvard School of Engineering and Applied Sciences, Cambridge, MA, where he is a member of the Signals, Information, and Networks Group (SING). His research interests are in high-dimensional signal processing and statistical inference.

**Yue M. Lu** (S'04–M'07–SM'12) received the M.Sc. degree in mathematics and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, both in 2007.

He was a Research Assistant at the University of Illinois at Urbana-Champaign, and was with Microsoft Research Asia, Beijing, China and Siemens Corporate Research, Princeton, NJ. From September 2007 to September 2010, he was a postdoctoral researcher at the Audiovisual Communications Laboratory at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is currently an Assistant Professor of electrical engineering at Harvard University, Cambridge, MA, directing the Signals, Information, and Networks Group (SING) at the School of Engineering and Applied Sciences. His research interests are in the general areas of signal processing, statistical inference and imaging.

He received the Most Innovative Paper Award of IEEE International Conference on Image Processing (ICIP) in 2006 for his paper (with Minh N. Do) on the construction of directional multiresolution image representations, and the Best Student Paper Award of IEEE ICIP in 2007. He also coauthored a paper (with Ivan Dokmanić and Martin Vetterli) that won the Best Student Paper Award of IEEE International Conference on Acoustics, Speech and Signal Processing in 2011.