An Uncertainty Principle for Functions Defined on Graphs

Ameya Agaskar and Yue M. Lu
Harvard School of Engineering and Applied Sciences
33 Oxford Street, Cambridge, MA 02138

ABSTRACT

The classical uncertainty principle provides a fundamental tradeoff in the localization of a function in the time and frequency domains. In this paper we extend this classical result to functions defined on graphs. We justify the use of the graph Laplacian's eigenbasis as a surrogate for the Fourier basis for graphs, and define the notions of "spread" in the graph and spectral domains. We then establish an analogous uncertainty principle relating the two quantities, showing the degree to which a function can be simultaneously localized in the graph and spectral domains.

1. INTRODUCTION

The *uncertainty principle* is a cornerstone result in time-frequency signal processing and harmonic analysis. It limits the degree to which a function can be simultaneously localized in time and frequency. More precisely, let $x(t) \in L^2(\mathbb{R})$ be a real-valued function with norm $||x||$ and Fourier transform $\widehat{x}(\omega)$. We use[1]

$$\Delta_t^2 \stackrel{\text{def}}{=} \frac{1}{||x||^2} \int_{-\infty}^{\infty} (t-t_0)^2 |x(t)|^2 dt,$$
$$\Delta_\omega^2 \stackrel{\text{def}}{=} \frac{1}{||x||^2} \int_{-\infty}^{\infty} \omega^2 |\hat{x}(\omega)|^2 \frac{d\omega}{2\pi}$$

to measure the "spreads" of $x(t)$ in time and frequency, respectively, where $t_0 \stackrel{\text{def}}{=} \frac{1}{||x||^2} \int_{-\infty}^{\infty} t|x(t)|^2 dt$. The uncertainty principle states that

$$\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}, \tag{1}$$

meaning that localizing a function in one domain must be done at the cost of increased spread in the other domain.

The history of uncertainty principles dates back to Heisenberg, who proved a result that Weyl (giving credit to Pauli) later showed was equivalent to (1).[2] Much later, several authors showed that a similar results hold for discrete-time signals as well.[3, 4] Donoho and Stark developed a new concept of uncertainty, showing that the maximum number of consecutive zeros in a discrete signal and in its DFT cannot simultaneously be large.[5] In this paper, we establish an uncertainty principle analogous to Heisenberg's for functions defined on graphs.

In recent years, there has been rapidly growing interest in extending traditional signal processing theory from standard domains (e.g., $\ell^2(\mathbb{Z}^d)$) to non-standard domains such as Riemannian manifolds and graphs.

Early work in this direction focused on multiscale representations of meshes for computer graphics applications.[6, 7] Another line of work examined the use of graph approximations to manifolds, with Belkin and later Giné and Koltchinskii examining the role of the graph Laplacian.[8–10] Meanwhile, several authors began considering multiscale wavelet-like transforms on graphs.[11–16] In parallel, Pesenson studied sampling theorems for "bandlimited" functions on graphs, results which may be useful in constructing critically sampled transforms.[17–19]

The rest of the paper is organized as follows. We start by briefly reviewing notations related to graphs and describe the graph Laplacian in Section 2. Fundamental to developing signal processing techniques on graphs is the notion of a Fourier transform, and the associated notion of "frequency." The field of spectral graph theory involves studying the eigenvalues of certain operators on vectors defined on a graph. When the Laplacian

Email: {aagaskar, yuelu}@seas.harvard.edu. Ameya Agaskar is also with MIT Lincoln Laboratory.

operator (an analogy to the standard, continuous-time Laplacian operator) is used, the resulting eigenvectors can be considered a Fourier basis, with "frequencies" defined by the associated eigenvalues. This is a standard approach in the literature, and we will provide some justification in this section. In section 3 we define a notion of "spread" for functions defined on graphs and prove an analogous uncertainty principle. We conclude in section 4.

2. MATHEMATICAL FORMULATION

We begin with a simple, undirected graph $G = (V(G), E(G))$, where $V(G) = \{v_1, v_2, \ldots, v_N\}$ is the set of N vertices and $E(G) = \{e_1, e_2, \ldots, e_M\}$ is the set of edges. Each edge is of the form $e = (u, v), u, v \in V$, an *unordered* pair of *different* vertices. The graph is simple in that is has no loops, or edges connecting a vertex to itself; the graph is undirected because the edges have no orientation. We will use the notation $u \sim v$ to indicate that u and v are connected by an edge. The graph is uniquely determined by its adjacency matrix $A = [a_{ij}]_{ij}$, where $a_{ij} = 1$ if there is a link between v_i and v_j, and $a_{ij} = 0$ otherwise. The diagonal of A is zero because no loops are allowed, and A is symmetric because the graph is undirected. A simple generalization is a weighted graph, where each edge has a "weight," and the off-diagonal entries of the adjacency matrix are replaced by the weights of the corresponding edges.

The degree of a vertex $\delta(v), v \in V$ is the number of edges incident upon that vertex. It is equal to the sum of the entries in the row or column corresponding to that vertex. We define Δ as the diagonal matrix that has the vertex degrees on the diagonal.

Spectral graph theory relates the properties of graphs to the eigenvalues of certain linear operators related to the graph.[20] These operators transform a signal on the graph to a different signal on the graph. Consider a finite-energy signal $\mathbf{x} \in \ell^2(G)$. We can think of $\mathbf{x}$ as an N-dimensional vector with norm $||\mathbf{x}|| = \sqrt{\mathbf{x}^T\mathbf{x}}$. A linear operator on $\ell^2(G)$ can be represented by an $N \times N$ matrix. The simplest operator considered in spectral graph theory is the adjacency matrix A. However, for reasons that will be described shortly, it is more common to use the spectrum of the Laplacian matrix.

The unnormalized Laplacian matrix is given by

$$L = \Delta - A. \tag{2}$$

The normalized Laplacian matrix is defined as

$$\begin{aligned} \mathcal{L} &= \Delta^{-1/2} L \Delta^{-1/2} \\ &= I - \Delta^{-1/2} A \Delta^{-1/2} \end{aligned} \tag{3}$$

The elements of $\mathcal{L}$ can be written as

$$\mathcal{L}_{ij} = \begin{cases} 1 & i = j \\ -\frac{1}{\sqrt{\delta(v_i)\delta(v_j)}} & i \neq j, v_i \sim v_j \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Both version of the Laplacian are symmetric, positive semidefinite matrices. There are several advantages to working with the normalized version. The eigenvalues are all between 0 and 2; a connected graph has the eigenvalue 0 with multiplicity 1, and a connected, bipartite graph (one in which the vertices can be divided into two sets and each vertex is only connected to vertices in the *other* set) has the eigenvalue 2 with multiplicity 1. In the sequel, we will refer to the normalized Laplacian at "the" Laplacian.

Since the Laplacian matrix is symmetric, we can diagonalize it as

$$\mathcal{L} = F \Lambda F^T, \tag{5}$$

where F is the matrix whose columns are the eigenvectors of $\mathcal{L}$, and Λ is the diagonal matrix of $\mathcal{L}$'s eigenvalues. Given a vector $\mathbf{x}$, we might like to find its representation in terms of the eigenvectors of $\mathcal{L}$. This can be computed by taking

$$\hat{\mathbf{x}} = F^T \mathbf{x}, \tag{6}$$

where we call $\hat{\mathbf{x}}$ the graph Fourier transform of $\mathbf{x}$. The matrix F^T is the Fourier transform operator. Since the Laplacian is symmetric, F is orthogonal, so $FF^T = F^T F = I$. So it is easy to see that we can invert the Fourier transform:

$$\mathbf{x} = F\hat{\mathbf{x}} \tag{7}$$

This is a standard analogy in the literature that several authors have used to define signal processing operations on graphs.[12,18,19,21,22] It is not immediately obvious that it is a fair analogy, but there is some justification.

First, the continuous-time Laplacian operator on the real line is $-\frac{d^2}{dx^2}$, an operator whose eigenvectors are of the form $e^{j\omega t}$ and $e^{-j\omega t}$, with eigenvalues ω^2. The Fourier transform for real signals then, finds the representation of a signal as a linear combination of the eigenvectors of the Laplacian operator. When applied to a graph formed by sampling the real line and connecting adjacent samples, the Laplacian matrix provides the standard stencil approximation for the second derivative (and the same holds for higher-dimensional lattices.) This is only a heuristic argument for many reasons. One is that the functions $e^{\pm j\omega t}$ are not in $\mathcal{L}^2$, or even any $\mathcal{L}^p$ space for $p < \infty$. Another is that it is still not clear that the Laplacian matrix measures what we want it to measure. We would like the Fourier transform to split a signal into its component frequencies. But what is a frequency on a graph?

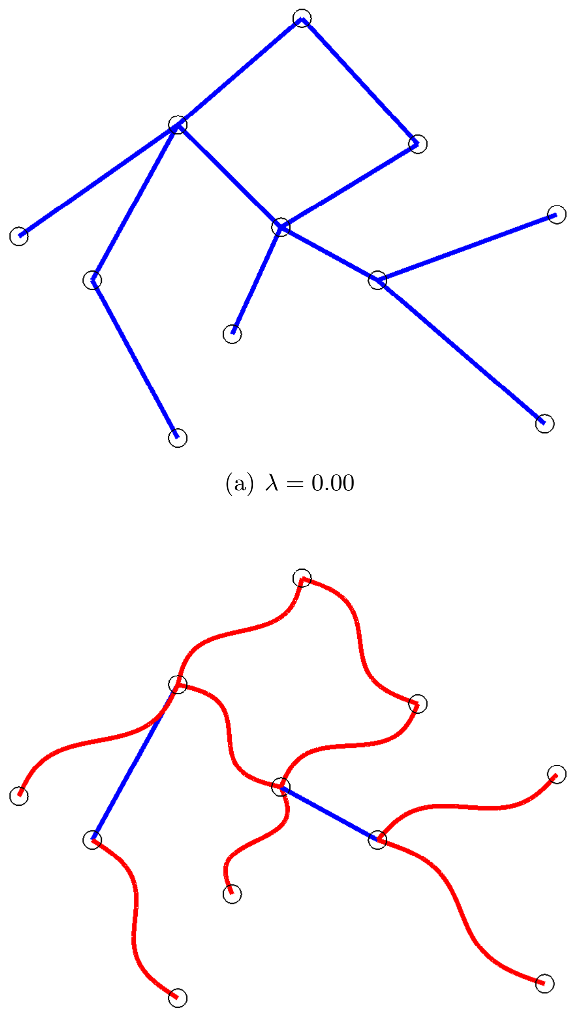
Another argument is that the Laplacian matrix for a graph formed by sampling a manifold becomes, as the density of samples becomes large, a good approximation for the Laplace-Beltrami operator (a differential geometric analogy to the second derivative) on the manifold.[9] The eigenvectors of the Laplace-Beltrami operator on simple manifolds provides well-accepted harmonic functions (such as the spherical harmonics) that allow us to pursue harmonic analysis on those manifolds. So it stands to reason that an operator that approximates the Laplace-Beltrami operator provides some kind of frequency information.

Furthermore, we would like to say that a signal has lots of high frequency components if its value changes quickly from one vertex to its neighbors, and that it has mostly low frequency components if its value varies very little from a vertex to its neighbors. To this end, we construct the $N \times M$ normalized incidence matrix D, where each column corresponds to an edge $e = (u, v)$ and has exactly two nonzero values: $+\frac{1}{\delta(u)}$ in the row corresponding to vertex u, and $-\frac{1}{\delta(v)}$ 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.) This choice constitutes a choice if *orientation*—an assignment of a direction for each edge in an undirected graph. Then the vector $\mathbf{y} = D^T\mathbf{x}$, for $\mathbf{x} \in \ell^2(G)$, is in some sense a derivative of $\mathbf{x}$. The vector $\mathbf{y}$ is a signal on the edges of the graph, where each edge has the difference between the normalized values on its endpoint vertices. Then our measure of relative variation of $\mathbf{x}$ is

$$\begin{aligned}\text{RelativeVariation}(\mathbf{x})^2 &= \frac{1}{||x||^2}||D^T\mathbf{x}||^2 \\ &= \frac{1}{||x||^2}\mathbf{x}^T DD^T\mathbf{x} \\ &= \frac{1}{||x||^2}\mathbf{x}^T\mathcal{L}\mathbf{x},\end{aligned} \tag{8}$$

where the last equality is well-known and easy to verify, and is completely independent of the orientation chosen.[23] So it seems that the components of $\mathbf{x}$ corresponding to the higher eigenvalues of $\mathcal{L}$ are the high-variation components, and the lower eigenvalues correspond to low-variation components. Thus, we are in some sense justified in calling this frequency.

Another argument is by analogy. Consider the cycle graph, illustrated in Figure 2. A signal on this graph is akin to a discrete, periodic signal. Because the Laplacian matrix of such a graph is a circulant matrix, it is diagonalized by the standard DFT matrix. The eigenvalue corresponding to a frequency of $\omega_k = \frac{2\pi k}{N}$ is $1 - \cos(\omega_k)$, which is an increasing function of $|\omega|$ on $[0, \pi]$ (and is equal to ω^2 for small values of k/N.) In a way, signal processing on this graph gives us equivalent results to signal processing of discrete, periodic signals.



(a) $\lambda = 0.00$ (b) $\lambda = 0.69$

(c) $\lambda = 1.68$ (d) $\lambda = 2.00$

Figure 1. An illustration of the 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, larger eigenvalues correspond to more sign changes and faster variation.

3. SPECTRAL GRAPH UNCERTAINTY PRINCIPLE

There are many reasons we might want to define notions of "spread" on the graph and graph spectral domains. One is simply to further justify the formal analogy between standard signal processing and graph signal processing—if we can prove an uncertainty principle akin to the uncertainty principles for standard signal processing, then we will have further evidence that our "Fourier" basis on graphs is a reasonable analogy to the standard Fourier transforms. Secondly, it will help define the tradeoffs in wavelet bases on graphs—basis vectors will either be tightly localized around a particular node, or have a sharply-defined frequency.

It will be helpful to first consider the related definitions of spread in a standard signal processing setting. Consider a continuous-time signal $x \in \mathcal{L}^2(\mathbb{R})$. Its Fourier transform is given by

$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt, \tag{9}$$

and the Fourier transform can be inverted by

$$x(t) = \int_{-\infty}^{\infty} \hat{x}(\omega) e^{+j\omega t} \frac{d\omega}{2\pi} \tag{10}$$

The time spread of the function is defined as[1]

$$\Delta_t^2 = \frac{1}{||x||^2} \int_{-\infty}^{\infty} (t - t_0)^2 |x(t)|^2 dt, \tag{11}$$

where the argument $t_0 = \frac{1}{||x||^2} \int_{-\infty}^{\infty} t|x(t)|^2 dt$ is called the time center of the function.

For a real signal, the Fourier transform is symmetric, so the frequency center is 0. This gives us a simpler definition of the frequency spread:

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

Simple manipulation and an application of the Cauchy-Schwarz inequality gives rise to the standard Uncertainty Principle for functions $x \in \mathcal{L}^2$, which we would like to emulate:

$$\Delta_t^2 \Delta_\omega^2 \geq \frac{1}{4}, \tag{13}$$

with equality achieved for Gaussian-shaped functions.

A similar result can be obtained for the discrete time case, with the added restriction that the DTFT evaluated at π is $X(e^{j\pi}) = 0$. This restriction prevents, among other things, the degenerate case of an impulse at a single sample, which would have a time spread of 0. This is not a problem in the continuous time case because the Dirac delta is not an element of $\mathcal{L}^2$.

In order to construct analogous definitions in the graph case, we need a definition of distance on the graph. We will let $d(u, v)$, the distance between vertex u and vertex v, be the smallest number of edges that need to be traversed to get from one to the other. Since we are dealing with connected graphs, this will always be finite. We briefly note that $d(\cdot, \cdot)$ satisfies all of the conditions of a metric—it is nonnegative, it is 0 if and only if the vertices are equal, and the triangle inequality holds. (It is clear that $d(u, w) \leq d(u, v) + d(v, w)$ since at worse we could get from u to v in $d(u, v)$ steps and then from v to w in $d(v, w)$ steps, though we might be able to do better.)

In both the standard continuous and discrete-time cases, the center of a function was allowed to attain any real value. It is hard to see how this would be possible on the graph, and furthermore it is hard to see how to even define the center of a function on a graph at all, since it is not a vector space like the real line. Instead, we note that the time spread is equivalent to the more general definition

$$\Delta_t^2 = \min_{t_0} \frac{1}{||x||^2} \int_{-\infty}^{\infty} (t - t_0)^2 |x(t)|^2 dt, \tag{14}$$

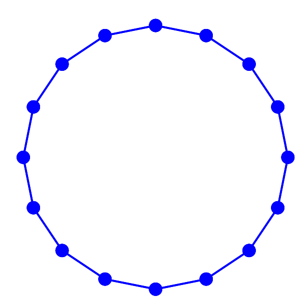


Figure 2. A cycle graph with 16 vertices. Graph Laplacian-based signal processing on this graph yields results similar to standard signal-processing on discrete-time, periodic signals.

where the argument that minimizes the expression is the time center of the signal. Now we can define by analogy that the graph spread of a vector $\mathbf{x} \in \ell^2(G)$ is

$$\Delta_g^2 \stackrel{\text{def}}{=} \frac{1}{||x||^2} \min_{u_0 \in V} \sum_{u \in V} d(u, u_0)^2 |x(u)|^2. \tag{15}$$

Note that if the graph is approximating a discrete-time lattice, the graph spread is at least the time spread, since the graph spread requires that the center be located on a vertex, while the time spread allows a more general center.

The spectral spread requires more thought. The eigenvalues λ of the Laplacian are more closely related to ω^2 than to ω. First, they are non-negative. Furthermore, the eigenvalues of the continuous-time Laplacian $-\frac{d^2}{dx^2}$ are ω^2 (corresponding to eigenvectors $e^{\pm j\omega t}$.) Since we are only dealing with real signals, and we only know the squares of the "frequencies," we will again take the frequency center to be zero, and define the spectral spread to be

$$\begin{aligned}
\Delta_s^2 &\stackrel{\text{def}}{=} \frac{1}{||x||^2} \sum_{i=1}^{N} \lambda_i |\hat{\mathbf{x}}(i)|^2 \\
&= \frac{1}{||x||^2} \hat{\mathbf{x}}^T \Lambda \hat{\mathbf{x}} \\
&= \frac{1}{||x||^2} \mathbf{x}^T F \Lambda F^T \mathbf{x} \\
&= \frac{1}{||x||^2} \mathbf{x}^T \mathcal{L} \mathbf{x}. \qquad (16)
\end{aligned}$$

It will be helpful to find a more explicit formula for $\mathcal{L}$ as a quadratic form:

$$\begin{aligned}
\mathbf{x}^T \mathcal{L} \mathbf{x} &= \sum_{u \in V} \left(x(u)^2 - \sum_{v \sim u} \frac{x(u)}{\sqrt{\delta(u)}} \frac{x(v)}{\sqrt{\delta(v)}} \right) \\
&= \sum_{u \in V} \sum_{v \sim u} \left(\frac{x(u)^2}{\delta(u)} - \frac{x(u)}{\sqrt{\delta(u)}} \frac{x(v)}{\sqrt{\delta(v)}} \right) \qquad (17) \\
&= \sum_{(u,v) \in E} \left(\frac{x(u)^2}{\delta(u)} - 2 \frac{x(u)}{\sqrt{\delta(u)}} \frac{x(v)}{\sqrt{\delta(v)}} + \frac{x(v)^2}{\delta(v)} \right) \\
&= \sum_{(u,v) \in E} \left(\frac{x(u)}{\sqrt{\delta(u)}} - \frac{x(v)}{\sqrt{\delta(v)}} \right)^2 \qquad (18)
\end{aligned}$$

where (17) follows because there are $\delta(u)$ vertices v neighboring each u, so the $\frac{x(u)^2}{\delta(u)}$ term is counted $\delta(u)$ times.

Given these definitions, we can prove a graph uncertainty principle. The theorem proved in this paper only applies to acyclic graphs. An acyclic graph is a graph that has no cycles, or closed loops of consecutive edges—this means that there is exactly one path from each node to any other node. A connected, acyclic graph is also known as a tree. Two extra conditions are needed. In the infinite cases of continuous and discrete time, the signal is required to have finite energy, which means that it must decay. This is in a sense a requirement that the function vanish at the boundaries of the index space (where in the infinite case the boundaries are at $\pm\infty$.) In the graph case, we will consider the boundary to be the leaves of the tree, or vertices with degree 1. Another problem is that there could be a delta function—a function with zero everywhere except for one vertex. This would have a graph spread of 0, preventing any kind of uncertainty bound. In fact, we could make the graph spread arbitrarily low by placing most of a signal's energy on one node. To prevent this, we must impose a condition on the value of the center node u_0. Two different conditions lead to two theorems; the second theorem requires a stronger condition on u_0, but improves the bound by a factor of $1/4$.

THEOREM 3.1. *Suppose we have a connected, acyclic graph G, and a function $x \in \ell^2(G)$ with center vertex u_0, such that $|x(u_0)| \leq \min_{u \sim u_0} |x(u)|$, and $x(v) = 0$ if $\delta(v) = 1$. Then the graph and spectral spreads are bounded by*

$$\Delta_g^2 \Delta_s^2 \geq \frac{1}{32} \tag{19}$$

Proof. We begin by converting the graph spread from a sum over vertices into a sum over edges. When we do this, each vertex u will be counted $\delta(u)$ times, since that is the number of undirected edges incident upon it. So we have

$$\Delta_g^2 = \frac{1}{||x||^2} \sum_{(u,v)\in E} \left(\frac{d(u,u_0)^2 |x(u)|^2}{\delta(u)} + \frac{d(v,u_0)^2 |x(v)|^2}{\delta(v)} \right)$$

Now we consider each term in this sum. Since for any real numbers a and b, $a^2 + b^2 \geq \frac{1}{2}(a+b)^2$, we get

$$\left(d(u,u_0)^2 \frac{|x(u)|^2}{\delta(u)} + d(v,u_0)^2 \frac{|x(v)|^2}{\delta(v)} \right) \geq$$
$$\frac{1}{2} \left(d(u,u_0) \frac{|x(u)|}{\sqrt{\delta(u)}} + d(v,u_0) \frac{|x(v)|}{\sqrt{\delta(v)}} \right)^2. \tag{20}$$

Define $d_{\min}((u,v),u_0) = \min(d(u,u_0), d(v,u_0))$ and $d_{\max}((u,v),u_0) = \max(d(u,u_0), d(v,u_0))$. Now, suppose that $d_{\min}((u,v),u_0) \geq 1$ (this is equivalent to $u \neq u_0$ and $v \neq u_0$.) Then

$$\frac{1}{2} \left(d(u,u_0) \frac{|x(u)|}{\sqrt{\delta(u)}} + d(v,u_0) \frac{|x(v)|}{\sqrt{\delta(v)}} \right)^2$$
$$\geq \frac{1}{2} \left(d_{\min}((u,v),u_0) \left(\frac{|x(u)|}{\sqrt{\delta(u)}} + \frac{|x(v)|}{\sqrt{\delta(v)}} \right) \right)^2 \tag{21}$$
$$\geq \frac{1}{2} \left(\frac{1}{2} d_{\max}((u,v),u_0) \left(\frac{|x(u)|}{\sqrt{\delta(u)}} + \frac{|x(v)|}{\sqrt{\delta(v)}} \right) \right)^2 \tag{22}$$

where (22) follows because the graph is a tree, so $d_{\max}((u,v),u_0) = 1 + d_{\min}((u,v),u_0)$, and we have stipulated that $1 \leq d_{\min}((u,v),u_0)$. The only edges for which $d_{\min}((u,v),u_0) = 0$ are the edges incident upon u_0. The condition on $x(u_0)$ in the statement of the theorem thus guarantees that

$$\frac{1}{2} \left(d(u,u_0) \frac{|x(u)|}{\sqrt{\delta(u)}} + d(v,u_0) \frac{|x(v)|}{\sqrt{\delta(v)}} \right)^2$$
$$\geq \frac{1}{8} \left(d_{\max}((u,v),u_0) \left(\frac{|x(u)|}{\sqrt{\delta(u)}} + \frac{|x(v)|}{\sqrt{\delta(v)}} \right) \right)^2$$
$$\geq \frac{1}{8} \left(d_{\max}((u,v),u_0) \left(\frac{x(u)}{\sqrt{\delta(u)}} + \frac{x(v)}{\sqrt{\delta(v)}} \right) \right)^2$$

holds for all edges (u,v). So we have

$$\Delta_g^2 \geq$$
$$\frac{1}{8} \frac{1}{||x||^2} \sum_{(u,v)\in E} \left(d_{\max}((u,v),u_0) \left(x(u) + x(v) \right) \right)^2 \tag{23}$$

The spectral spread can be written as

$$\Delta_s^2 = \frac{1}{||x||^2} x^T \mathcal{L} x$$
$$= \frac{1}{||x||^2} \sum_{(u,v)\in E} \left(\frac{x(u)}{\sqrt{\delta(u)}} - \frac{x(v)}{\sqrt{\delta(v)}} \right)^2$$

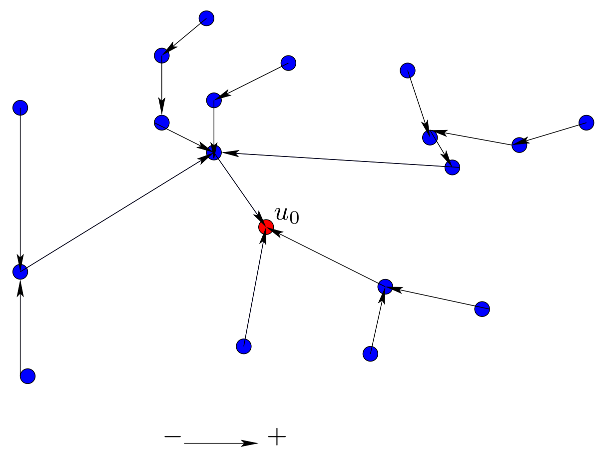


Figure 3. The chosen orientation in the proof of the Theorems. The edges are assumed to point toward the vertex closer to u_0. In the sum resulting from the Cauchy-Schwarz inequality, the closer vertex contributes a positive term, and the further one contributes a negative term.

This value, of course, does not depend on the chosen orientation O of the graph (a choice that determines which vertex is the (+) term and which is the (-) term.) However, the orientation will matter when we apply the Cauchy-Schwarz inequality later in the proof. We will choose the orientation illustrated in Figure 3: $u \leftarrow v$ if $d(u, u_0) < d(v, u_0)$. (Two neighbors in an acyclic graph cannot have the same distance to u_0, or else there would be a cycle composed of (u, v) and portions of each path from u to u_0 and v to v_0.)

So we rewrite the spectral spread

$$\begin{aligned}\Delta_s^2 &= \frac{1}{||x||^2} x^T \mathcal{L} x \\ &= \frac{1}{||x||^2} \sum_{u \leftarrow v \in O} \left(\frac{x(u)}{\sqrt{\delta(u)}} - \frac{x(v)}{\sqrt{\delta(v)}} \right)^2\end{aligned}$$

Using the Cauchy-Schwarz inequality to combine our results for graph and spectral spreads, we obtain

$$\Delta_g^2 \Delta_s^2 \tag{24}$$

$$\geq \frac{1}{||x||^4} \frac{1}{8} \left(\sum_{u \leftarrow v \in O} \begin{array}{c} d_{\max}((u,v), u_0) \cdot \left(\frac{x(u)}{\sqrt{\delta(u)}} + \frac{x(v)}{\sqrt{\delta(v)}} \right) \\ \cdot \left(\frac{x(u)}{\sqrt{\delta(u)}} - \frac{x(v)}{\sqrt{\delta(v)}} \right) \end{array} \right)^2$$

$$= \frac{1}{||x||^4} \frac{1}{8\delta_{\max}} \left(\sum_{u \leftarrow v \in O} d_{\max}((u,v), u_0) \left(\frac{x(u)^2}{\delta(u)} - \frac{x(v)^2}{\delta(v)} \right) \right)^2 . \tag{25}$$

Now we make use of the fact that G is acyclic and connected. This means that for every vertex u (not counting u_0 itself,) there can be only one neighbor closer to u_0 than it; all of its remaining neighbors must be further away, with distance $d(u) + 1$. (If this were not the case, then there would be multiple paths from u to u_0, and there would therefore be a cycle somewhere in the graph.) So in (25), each vertex $u \neq u_0$ contributes one negative term with distance $d(u, u_0)$ and $\delta(u) - 1$ positive terms with distance $d(u, u_0) + 1$. So we can rewrite the sum in terms of the vertices as

$$\Delta_g^2\Delta_s^2 \geq$$
$$\frac{1}{||x||^4}\frac{1}{8}\left(\sum_{u\in V}\begin{matrix}(\delta(u)-1)\cdot(1+d(u,u_0))\frac{x(u)^2}{\delta(u)} \\ -1\cdot d(u,u_0)\frac{x(u)^2}{\delta(u)}\end{matrix}\right)^2$$
$$= \frac{1}{||x||^4}\frac{1}{8}\left(\sum_{u\in V} x(u)^2\left[d(u,u_0)\left(1-\frac{2}{\delta(u)}\right)+1-\frac{\delta(u)}{}\right]\right)^2, \tag{26}$$

where we can check that the inequality holds for the special case term of $u = u_0$, since $d(u_0, u_0) = 0$. The quantity being squared in (26) is always nonnegative, so minimizing it is equivalent to minimizing (26) itself. This quantity will be smallest for a path graph (see Figure 4) where $\delta(u) = 2$ for every node except the endpoints, which have degree one and at which therefore x is zero. Although making the degree of the vertices small will increase the typical distances, ensuring that every (contributing) vertex has degree two eliminates the distance-dependent term entirely, and minimizes the other term. Finally, we have that

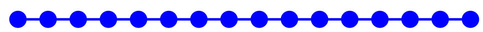


Figure 4. A path graph's vertices each have degree 2, except for the endpoints, which have degree 1.

$$\Delta_g^2\Delta_s^2 \geq \frac{1}{||x||^4}\frac{1}{8}\left(\sum_{u\in V}\frac{1}{2}x(u)^2\right)^2$$
$$= \frac{1}{32}. \tag{27}$$

□

By requiring stronger conditions on the signal x, we can get the tighter bound described in the following theorem

Theorem 3.2. *Suppose we have a connected, acyclic graph G, and a function $x \in \ell^2(G)$ with center vertex u_0, such that $x(u_0) = 0$ and $x(v) = 0$ if $\delta(v) = 1$. Then the graph and spectral spreads are bounded by*

$$\Delta_g^2\Delta_s^2 \geq \frac{1}{8} \tag{28}$$

Proof. The proof mostly follows that of Theorem 3.1. When bounding the graph spread, we stop at (21) rather than proceeding to (22), which gives us

$$\Delta_g^2 \geq$$
$$\frac{1}{2}\frac{1}{||x||^2}\sum_{(u,v)\in E}\left(d_{\min}((u,v),u_0)\left(\frac{x(u)}{\sqrt{\delta(u)}}+\frac{x(v)}{\sqrt{\delta(v)}}\right)\right)^2, \tag{29}$$

where an extra factor of $\frac{1}{4}$ is present because we do not convert from $d_{\min}$ to the $d_{\max}$. Using Cauchy-Schwarz and the new inequality, we get

$$\Delta_g^2\Delta_s^2 \tag{30}$$
$$\geq \frac{1}{||x||^4}\frac{1}{2}\left(\sum_{u\leftarrow v\in O} d_{\min}((u,v),u_0)\left(\frac{x(u)^2}{\delta(u)}-\frac{x(v)^2}{\delta(v)}\right)\right)^2. \tag{31}$$

Once again, we use the fact that in an acyclic graph, every vertex $u \neq u_0$ has exactly one neighbor closer to u_0, and $\delta(u) - 1$ neighbors that are further away. So there are $\delta(u) - 1$ positive terms with distance $d(u, u_0)$

and 1 negative term with distance $d(u, u_0) - 1$. The case of u_0 does not matter—since $x(u_0) = 0$, it contributes nothing to either sum. Therefore,

$$\Delta_g^2 \Delta_s^2 \geq$$
$$\frac{1}{||x||^4} \frac{1}{2} \left(\sum_{u \in V} \begin{array}{l} (\delta(u) - 1) \cdot d(u, u_0) \frac{x(u)^2}{\delta(u)} \\ -1 \cdot (d(u, u_0) - 1) \frac{x(u)^2}{\delta(u)} \end{array} \right)^2$$
$$= \frac{1}{||x||^4} \frac{1}{2} \left(\sum_{u \in V} x(u)^2 \left(d(u, u_0) \left(1 - \frac{2}{\delta} \right) + \frac{1}{\delta(u)} \right) \right)^2 \tag{32}$$

Every term inside the sum is nonnegative, and all nonzero terms have $d(u, u_0) \geq 1$, so we can lower bound it by substituting 1 for $d(u, u_0)$ for every term, and finally get

$$\Delta_g^2 \Delta_s^2 \geq \frac{1}{||x||^4} \frac{1}{2} \left(\sum_{u \in V} x(u)^2 \left(1 - \frac{1}{\delta(u)} \right) \right)^2, \tag{33}$$

This quantity is minimized by making the vertex degrees as small as possible, which is once again accomplished by taking the path graph, where $\delta(u) = 2$ for every vertex except the endpoints. This gives us the final bound

$$\Delta_g^2 \Delta_s^2 \geq \frac{1}{8} \tag{34}$$

□

4. CONCLUSION

Graph signal processing is an emerging field with many potential applications. Wavelet frames defined on graphs, with frame vectors suitably localized in the graph and spectral domains. To define this localization, graph and spectral spreads were defined that are analogous to the classical time and frequency spreads. Finally, an uncertainly principle analogous to the Heisenberg-Weyl uncertainty principle was proved.

REFERENCES

[1] Vetterli, M. and Kovačević, J., [*Wavelets and Subband Coding*], Prentice Hall, Englewood Cliffs, NJ (1995).

[2] Folland, G. B. and Sitaram, A., "The uncertainty principle: A mathematical survey," *The Journal of Fourier Analysis and Applications* **3**, 207–238 (May 1997).

[3] Ishii, R. and Furukawa, K., "The uncertainty principle in discrete signals," *IEEE Transactions on Circuits and Systems* **33**, 1032–1034 (Oct. 1986).

[4] Calves, L. C. and Vilbe, P., "On the uncertainty principle in discrete signals," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* **39**, 394–395 (June 1992).

[5] Donoho, D. L. and Stark, P. B., "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics* **49**(3), 906–931 (1989).

[6] Lounsbery, J. M., *Multiresolution Analysis for Surfaces of Arbitrary Topological Type*, PhD thesis, University of Washington, Seattle, Washington (1994).

[7] Hoppe, H., "Progressive meshes," in [*Proceedings of the 23rd Annual Conference on Computer Graphics and Iinteractive Techniques (SIGGRAPH)*], 99–108 (1996).

[8] Bernstein, M., De Silva, V., Langford, J. C., and Tenenbaum, J. B., "Graph approximations to geodesics on embedded manifolds," technical report, Stanford (2000).

[9] Belkin, M., *Problems of learning on manifolds*, PhD thesis, University of Chicago, Chicago, Illinois (2003).

[10] Giné, E. and Koltchinskii, V., "Empirical graph laplacian approximation of Laplace-Beltrami operators: Large sample results," *IMS Lecture Notes-Monograph Series* **51**, 238–259 (2006).

[11] Crovella, M. and Kolaczyk, E., "Graph wavelets for spatial traffic analysis," in [*IEEE INFOCOM*], (2002).

[12] Coifman, R. R. and Maggioni, M., "Diffusion wavelets," *Applied and Computational Harmonic Analysis* **21**, 53–94 (July 2006).

[13] Wang, W. and Ramchandran, K., "Random multiresolution representations for arbitrary sensor network graphs," in [*Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*], **4**, IV (2006).

[14] Hammond, D. K., Vandergheynst, P., and Gribonval, R., "Wavelets on graphs via spectral graph theory," arXiv Preprint (Dec. 2009).

[15] Narang, S. and Ortega, A., "Local two-channel critically sampled filter-banks on graphs," in [*Image Processing (ICIP), 2010 17th IEEE International Conference on*], 333–336 (2010).

[16] Wagner, R., *Distributed multi-scale data processing for sensor networks*, PhD thesis, Rice University, Houston, Texas (2007). Ph.D.

[17] Pesenson, I., "Band limited functions on quantum graphs," *Proceedings of the American Mathematical Society* **133**(12), 3647–3656 (2005).

[18] Pesenson, I., "Sampling in Paley-Wiener spaces on combinatorial graphs," *American Mathematical Society* **360**(10), 5603–5627 (2008).

[19] Pesenson, I. Z. and Pesenson, M. Z., "Sampling, filtering and sparse approximations on combinatorial graphs," *Journal of Fourier Analysis and Applications* (Jan. 2010).

[20] Chung, F. R. K., [*Spectral graph theory*], American Mathematical Society (1997).

[21] Pesenson, I., "Variational splines and Paley–Wiener spaces on combinatorial graphs," *Constructive Approximation* **29**, 1–21 (Jan. 2008).

[22] Pesenson, I., "Removable sets and approximation of eigenvalues and eigenfunctions on combinatorial graphs," *Applied and Computational Harmonic Analysis* **29**(2), 123–133 (2010).

[23] Biggs, N., [*Algebraic Graph Theory*], Cambridge University Press, 2 ed. (1994).