Graph Wavelets for Spatial Traffic Analysis*

BUCS-TR-2002-020

Mark Crovella
Department of Computer Science
Boston University
Boston MA 02215

Eric Kolaczyk
Department of Math and Statistics
Boston University
Boston MA 02215

July 15, 2002

Abstract

A number of problems in network operations and engineering call for new methods of traffic analysis. While most existing traffic analysis methods are fundamentally temporal, there is a clear need for the analysis of traffic across multiple network links — that is, for *spatial* traffic analysis. In this paper we give examples of problems that can be addressed via spatial traffic analysis. We then propose a formal approach to spatial traffic analysis based on the wavelet transform. Our approach (*graph wavelets*) generalizes the traditional wavelet transform so that it can be applied to data elements connected via an arbitrary graph topology. We explore the necessary and desirable properties of this approach and consider some of its possible realizations. We then apply graph wavelets to measurements from an operating network. Our results show that graph wavelets are very useful for our motivating problems; for example, they can be used to form highly summarized views of an entire network's traffic load, to gain insight into a network's global traffic response to a link failure, and to localize the extent of a failure event within the network.

*This work was supported in part by NSF awards ANI-9986397 and ANI-0095988, and by ONR award N00014-99-1-0219.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
15 JUL 2002		**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Graph Wavelets for Spatial Traffic Analysis	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Office of Naval Research,One Liberty Center,875 North Randolph Street Suite 1425,Arlington,VA,22203-1995**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**18**	

1 Introduction

To date, the traffic analysis tools developed in the research community and the traffic analysis needs of network engineers and operators have been somewhat disconnected. Most research on traffic analysis has focused on the properties of traffic flowing over individual links, treated as a timeseries [1]. However, network engineers and operators are very often more concerned with the properties of traffic over multiple links, or whole networks.

In fact, there are many network engineering challenges that could be aided by better tools for traffic analysis. For example, traffic properties play a central role in 1) provisioning and capacity planning; 2) network configuration and traffic engineering; 3) failure detection and diagnosis; and 4) attack detection and prevention. However, traffic analysis tools and methods focused on these kinds of problems are generally not well developed. As a result, many network operators and engineers are forced to address these problems manually or via ad-hoc tools.

A common thread running through these problems is the importance of comparison and analysis of traffic patterns across multiple, or all, network links simultaneously. We call this *spatial* traffic analysis.[1] Spatial traffic analysis seeks to answer questions about traffic patterns in and between "regions" — topologically localized sets of links — of a network. For example, traffic engineering can be aided by summarizations of average traffic in different network regions; and failure and attack detection can be aided by comparisons of traffic across different network regions.

Providing useful, practical tools for spatial traffic analysis is difficult. Two problems arise: first, the large quantity and high dimensionality of the data involved is unmanageable without methods for efficient and flexible data summarization. Second, algorithms must be developed that correctly and intelligently make use of such summaries for the solution of network engineering problems. Given the many degrees of freedom introduced by the wealth of data, such algorithms are not immediately obvious. Good solutions for these two problems are interrelated, because each influences the other.

In this paper we present new techniques for spatial traffic analysis. These techniques are based on explicit consideration of network topology; we believe that effective network engineering must consider both the traffic properties on the network's links *and* the manner in which the links are connected. Thus our approach is intended to support a *whole-network* view of data traffic.

To enable this view, we develop a new set of formal tools based on wavelet analysis. Our principal insight in this thrust is that traditional wavelet analysis can be generalized for use on data elements connected via an arbitrary graph topology, leading to discrete-space analogues of the well-known wavelet transform. That is, in contrast to the traditional use of wavelets in traffic analysis, we apply wavelets to the spatial domain rather than the temporal domain. In this paper we show one way to accomplish this, and we develop a formal framework for what we call *graph wavelets*. Graph wavelets are quite general and flexible, and we explore some of the variations that are possible.

Using measurements taken from an operating network (Abilene [2]) we show that graph wavelets can provide considerable leverage on whole-network traffic analysis. We show how graph wavelets can be used to form highly summarized views of an entire network's traffic load; how they can be

[1] More accurately, we might instead write *topological* to distinguish between this context and that in which the actual spatial geography of the network is incorporated into the analysis. However we use *spatial* to emphasize the similarity in spirit of our methods to those in the spatial analysis domain.

used to gain insight into a network's global traffic response to a link failure; and how they can help localize the extent of a failure event.

The examples in this paper use link counts available from routers via SNMP. However the methods are general enough to apply to other kinds of measurements associated with the network graph: for example, to study spatial correlation patterns in packet loss. Furthermore, the methods apply equally well to measurements associated with a graph's edges (links) or vertices (routers).

The remainder of the paper is organized as follows. In Section 2 we review related work. Then in Section 3 we describe example motivating problems, and we present an informal introduction to graph wavelets as tools for addressing those problems. Section 4 lays out the formal aspects of graph wavelets: their definition and certain basic properties. Section 5 then presents detailed examples of how graph wavelets shed light on the nature of measurements taken from the Abilene network. Finally, in Section 6 we conclude.

2 Background and Related Work

The vast majority of research into network traffic analysis has focused on the behavior of traffic on individual links over time. This approach has yielded a number of insights. Most salient among these are observations about the time scaling behavior of traffic: *self-similarity* and *long-range dependence*, [3, 4]; and *multifractality* and *local scaling* [5, 6]. Many of the key results in traffic time scaling analysis have been facilitated by the use of powerful tools, in particular the techniques of wavelet analysis (e.g., [7]).

These temporal traffic analyses have been quite successful in illustrating the presence of surprising patterns in the traffic flowing over individual links. However remarkably little research has sought to investigate whether traffic patterns are discernable across multiple links.

A similar trend has taken place in the development of network anomaly detection systems. The authors in [8] propose that an anomaly detection system should:

1. summarize the nature of typical network conditions in some compact set of metrics,
2. continually compare current conditions to the typical metrics, and
3. draw operator attention to deviations from typical conditions in as precise and informative manner as possible.

In a style similar to time scaling analysis, work in anomaly detection has generally approached the summarization task in step 1 from a single-link, temporal analysis standpoint – for example, [9].

These timeseries-based approaches to anomaly detection have also made use of wavelet analysis. An example is [10], which explores the diagnostic utility of the traffic energy function; this function is easily obtained using multiresolution analysis. Another approach applying wavelet analysis to anomaly detection is [11], which focuses on analysis of traffic flow measurements.

In contrast, the approach we take to anomaly detection — and traffic analysis generally — focuses on the spatial domain: that is, the relationship between traffic on topologically related links. In that sense our work bears a relation to [11], which shows that comparing traffic flows in the incoming and outgoing directions of an access link is useful for identifying anomalies such as denial of service attacks and flash crowd behavior.

Generally speaking, in fields ranging from image processing to geography, experience has found that *scale* is a concept of fundamental importance to the analysis of spatially indexed data. We will argue in Section III (and throughout this paper) that many of the spatial challenges faced by network engineers similarly involve scale in some essential fashion. And therefore our emphasis here is on methods for the *multiscale* analysis[2] of spatial network data.

The image processing literature arguably has to date the most well known and technically developed body of multiscale analysis techniques for spatial data. Modern representatives from this body perhaps can be said to begin with early work on Laplacian pyramid filtering [13] (which itself formalized still earlier ideas in image processing and computer vision), which was soon followed and replaced by the current paradigm based on two-dimensional wavelets and their extensions. Wavelet-based tools have had a fundamental impact on a variety of standard tasks in image processing, including compression (witness their key role in the JPEG2000 standard), denoising, segmentation, and classification.

However, wavelet-based methods for images are not immediately applicable to the analysis of spatial network data, for the simple reason that the former are designed for standard L_2 topologies and not arbitrary network topologies. On the other hand, there has been recent success in extending the basic wavelet framework to non-standard topologies (e.g., [14, 15, 16]), although none of this work so far is relevant to network analysis. What is needed is a notion of wavelets on graphs, which we develop in detail in this paper.

3 Motivation and Approach

In this section we provide a more detailed motivation for the notion of spatial traffic analysis, and complement that with an informal illustration of our approach to the problem.

3.1 Spatial Traffic Analysis: Motivation

A number of example problems in network engineering and operations will serve to highlight the need for a whole-network approach to traffic analysis.

1. *Identifying Spatial Aspects of Typical Operating Conditions*

   An important problem facing network operators and engineers is the identification and interpretation of typical operating conditions. This is fundamentally a whole-network problem.

   As an example, spatial analysis plays a role in traffic engineering. The goal of traffic engineering is to assign traffic flows to network paths in a way that meets some design criterion. One commonly sought criterion can be load balancing across the network. Engineers may seek to balance load so as to minimize the effects of any single link failure, or to minimize the utilization of the busiest links.

[2] Multiscale analysis, as used in the various literatures, refers simply to the analysis of a given object(s) at multiple scales. While the term "multiresolution analysis" sometimes is used interchangeably with "multiscale analysis," the former has a specific technical meaning in the mathematics and engineering communities (referring to a sequence of successive approximation spaces, as developed originally by Mallat and Meyer – e.g., see [12]). The latter more accurately describes the perspective and contributions in this paper and will be used throughout.

A valuable precursor to load balancing is an understanding of which network regions are carrying the most load, and which regions are relatively less utilized. Armed with this knowledge, network engineers can make high-level, qualitative decisions about the intended outcome of route changes when performing load balancing. Summary information about traffic loads over varying network locations and region sizes provides help in making these kinds of traffic engineering decisions.

2. *Understanding Shifts in Traffic Patterns*

   A related goal involves understanding the shifts in traffic patterns as a result of traffic engineering decisions or network equipment failures.

   Some networks are engineered with sufficient bandwidth for "protection," *i.e.*, so that traffic shifts due to equipment failures can be absorbed without manual intervention in the routing system. In contrast, some networks are provisioned with the expectation that equipment failures will be addressed through explicit traffic engineering actions. In each case, it is essential to have a whole-network view of how traffic patterns shift when equipment fails or traffic is manually re-routed. This whole-network view must provide quantitative information about which regions of the network experienced increased load and which experienced decreased load as a result of the network event.

3. *Identifying Regions Exhibiting Traffic Anomalies*

   When traffic exhibits unusual characteristics, an immediate and fundamental question concerns the size and extent of the region over which the anomaly occurs.

   For example, if observed traffic load increases to an unusual level, this may be due to a number of factors. Traffic throughout the network may have risen, due to some external driver of increased demand such as a breaking news story. Alternatively, traffic in a localized network region may be increased due to a flash crowd effect (publication of a popular video or report that drives traffic to a single location). Finally, traffic load may be due to a particular pair of hosts engaging in abnormally high traffic. These three scenarios are primarily distinguished by the size of the "neighborhood" over which the anomalously high traffic is observed, and they each demand a different response from network operators.

   As another example, rapidly detecting denial of service (DoS) attacks is crucial for responsive network management. Unfortunately, increased traffic on a single link is not a good indicator of the presence or nature of a DoS attack. Most DoS attacks are distributed, with flooding packets arriving from multiple sources along multiple paths. Accurate identification of a distributed DoS attack using traffic counts requires the simultaneous assessment of traffic on multiple links of the network.

3.2 Spatial Traffic Analysis: Approach

The problems just described all concern questions about one or more "regions" or "neighborhoods" within a given network. To place our discussion of network neighborhoods in a formal setting we consider the graph that is isomorphic to the network as follows: routers or switches in the network correspond to vertices in the graph; and communication links in the network correspond to edges in the graph. We will call the collection of routers and links the *network* and the corresponding

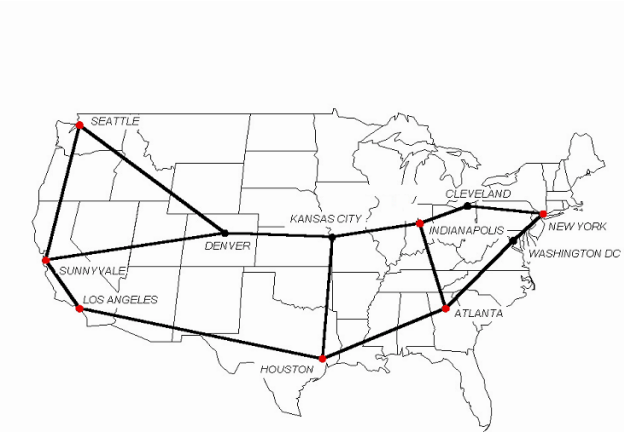


Figure 1: Example Network: Abilene

graph the *network graph*. Furthermore, we will reserve the terms "links" and "routers" for the network elements and the terms "edges" and "vertices" for the graph elements.[3]

Furthermore, another graph will be important: the *network line graph*. For any given graph $G = (V, E)$, its corresponding line graph $L(G) = (V', E')$ is defined such that $|V'| = |E|$ and there is an edge in $L(G)$ for each *pair* of edges in G that share a common endpoint; *i.e.*, $\{(i, j) \in E'\} \leftrightarrow \{\exists k | (k, i) \in E, (k, j) \in E\}$. The network line graph is the line graph of the network graph.

The two kinds of graphs are both useful because in a network, certain measurements are associated with the routers, and certain measurements are associated with the links. When we are concerned with comparing measurements associated with routers, then we will be concerned with the adjacency relationships of routers, and so with the network graph. However when we are concerned with measurements associated with links (as will be the case in all of the examples in this paper) we will be concerned with the adjacency relationships of links, and so our analyses will involve the network line graph. All of the numerical results that follow in this paper will be based on network line graphs. However, our graph wavelets are defined for arbitrary connected graphs.

Our examples in this paper use the network shown in Figure 1. This is a map of the Abilene network, which is the network supporting Internet2 (this image is from [2]).[4] This network has 11 routers and 14 links. The corresponding line graph (not shown) has 14 vertices and 23 edges.

Armed with these tools, we can begin to explore methods for analyzing measurements with respect to network neighborhoods. In the remainder of this section we present an intuitive view of our proposed approach. A formal, rigorous development is deferred to the next section.

For purposes of discussion here, let us define the zero-hop neighborhood of a link as the link itself. The one-hop neighborhood is the link, and the set of all links that can be reached in one hop; that is, by traversing a single edge in the network line graph. The two-hop neighborhood of a link is its one-hop neighborhood and all the links that can be reached from any link in that neighborhood in one additional hop, and so on.

[3]We will consider edges in the graph to be undirected. This is a simplification, and we discuss some implications of this simplification below.

[4]We have omitted one link from this figure for which we have no data. Internet2 is a project developing new network applications and technologies; it has built and uses the Abilene network for testing and deploying these experimental systems. All of the links shown are OC-48, running at 2.48 Gbps.

Consider the NYC-Cleveland link. Its one-hop neighborhood consists of the three links from Indianapolis to Washington DC, and its two-hop neighborhood consists of those three links plus the Indianapolis-Kansas City, Indianapolis-Atlanta, and Washington DC-Atlanta links.

The central idea in our approach to spatial network data analysis is the comparison of metrics between neighborhoods. For example, for any given link and metric, we might define a comparison at level ℓ (where, for convenience, ℓ is an even number) as the average of that metric over all links in the $(\ell/2)$-hop neighborhood, minus the average over all links that are in the ℓ-hop but not the $(\ell/2)$-hop neighborhood. That is, we compare the average measurement in a "disk" around the link to the average measurement in the corresponding "ring" around the disk; if the metric is larger on average closer to the link, the comparison will be positive. So the level 2 comparison for the NYC-Cleveland link might consist of the average measurements on the Indianapolis-Cleveland and NYC-Washington links, minus the measurements on the NYC-Cleveland link.[5]

A number of considerations motivate this general style of data analysis. First, data traffic on nearby links may often be highly correlated; this will occur for a number of reasons, including the fact that each link carries data flows which themselves likely pass over multiple links. Thus it may often be reasonable to summarize the traffic on many links in a neighborhood with a single value. Second, differences between neighborhoods are important, as can be seen from the example problems in this section; we wish to draw attention to such differences in our analysis. Third, different effects will be expected to occur at different spatial scales in the network; hence, we define comparisons at varying levels so as to flexibly detect a wide range of phenomena.

The general notion of summarization and comparison over varying locations and scales is the underlying principle of wavelet analysis. Indeed, the example problems and approach described in this section bear strong similarity to problems in signal and image processing, domains in which wavelet analysis has provided considerable leverage. However, traditional wavelet analysis has restricted itself primarily to regular spaces, *e.g.*, L_2 [12]. Therefore, in pursuit of a formal basis for attacking the problems described here, it is necessary and appropriate to extend the notion of wavelets to certain graph topologies, which we do in the next section.

4 Graph Wavelets

In this section we approach the topic of whole-network wavelet analysis in a more formal fashion. After reviewing some basic concepts and principles from traditional wavelet analysis we develop a framework for a class of graph-based wavelets.

4.1 Traditional Wavelets and Multiscale Analysis

At the most basic level, a multiscale analysis (based on wavelets or otherwise) is simply a coordinated way of examining local differences in a set of measurements, across a range of scales. Multiscale analyses based on wavelets, of course, are now the most common and well-known example of this approach. Although there are currently a host of wavelet functions available, possessed of a variety of different properties and characteristics, at their most basic these functions share the two fundamental properties that they (i) are localized (having either finite or essentially

[5]This description in terms of simple averages over neighborhoods of different levels is a simplification for purposes of illustration.

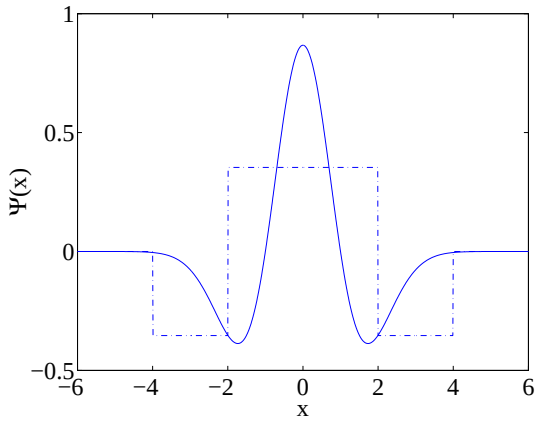


Figure 2: Haar (dot-dashed) and Mexican Hat (solid) Wavelet Functions on the Real Line.

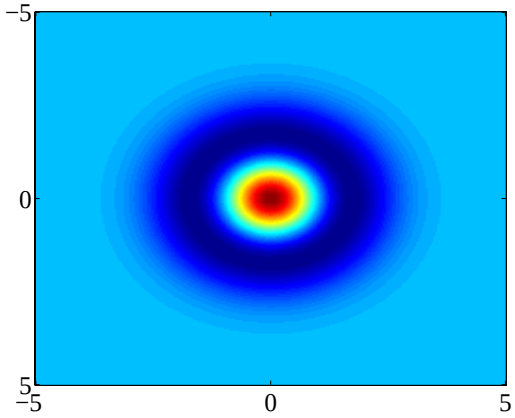


Figure 3: Mexican Hat Wavelet Function on the Plane. The central disk is strongly positive, and the surrounding ring is strongly negative.

finite support) and (ii) have zero integral (and hence, excluding the trivial case, they must oscillate positive and negative). By virtue of this locally oscillating behavior, the inner product of a wavelet, say Ψ, with another function, say f, yields coefficients $\langle f, \Psi \rangle = \int f(x)\Psi(x)\,dx$ that are essentially the weighted difference of information in f on the regions of positive and negative support of Ψ. Any other properties or characterisics of Ψ, such as compact support or smoothness, are the result of using additional "degrees of freedom" in the overall design process.

Figure 2 shows two examples of wavelet functions Ψ on the real line $\mathbb{R}$. The first is a symmetric variant of the well-known Haar wavelet, piecewise constant and of finite support. The second is the one-dimensional analogue of the so-called "mexican hat" wavelet, $\Psi(x) \sim (1 - x^2)\exp(-x^2/2)$, which is infinitely differentiable and of infinite support. Both have zero integral and unit L_2 norm. When these two functions are rotated about their point of symmetry, the results are radially symmetric analogues in the plane $\mathbb{R}^2$. The analogue of the latter is the mexican hat wavelet, whose relative shape and magnitude are shown in Figure 3 in the form of an image.

Traditionally, a wavelet analysis of f, for functions f defined on some subset of a Euclidean space (e.g., $\mathbb{R}$ or $\mathbb{R}^2$) is based on the collection of coefficients $\{\langle f, \Psi_{a,b} \rangle\}$, where $\Psi_{a,b}(x) = a^{-1/2}\Psi\left((x - b)/a\right)$ is a shifted and dilated version of Ψ by b and $a > 0$, respectively. For $a, b \in \mathbb{R}$,

the mapping $f \rightarrow \{\langle f, \Psi_{a,b} \rangle\}$ is known as a *continuous wavelet transform.* If on the other hand $a = 2^j$ and $b = k$, for $j, k \in \mathbf{Z}$, this mapping is known as a *discrete wavelet transform.* And in the latter case, when the function Ψ is chosen appropriately, it is possible to create a system $\{\Psi_{j,k}\}$ of wavelet functions that are orthonormal. Within each of these classes of wavelet transforms (continuous, discrete-redundant, and discrete-orthogonal) there are numerous examples to be found, customized to have various additional properties felt to be useful for a particular application(s). See [12], for example.

Regardless of the specifics, the end result of a wavelet transform is an alternative representation of the information in f with respect to an indexing in scale and location. Which particular class of wavelet transform is preferable (as well as the choice of wavelet function Ψ within class), if any, typically depends on the uses to which one intends to put such a representation. For example, the continuous wavelet transform has been quite popular in astronomy, particularly for the detection of point sources and anomolies in satellite image data (e.g., [17]). Alternatively, the discrete wavelet transform and its extensions have proven especially useful for the tasks of denoising and compression (e.g., [18]). Our own contribution in this paper can be said to more closely resemble the traditional continuous wavelet transform in spirit.

More recently, there has been much work on so-called "second generation" wavelets (*e.g.,* [19]). Systems of such wavelets are not necessarily composed of either shifts or dilations of some single function Ψ. Nevertheless, the members are localized and indexed across a range of scales and locations within scales, have zero integral, and share some common characteristic(s) in their definition. Examples include piecewise constant wavelets defined on general measure spaces [14], wavelets on triangular meshes of arbitrary topology [15], and wavelets on the sphere [20, 16]. The wavelets we develop in this paper, in extending the traditional framework described above to the context of network graphs, are a new contribution to this second generation of wavelets.

4.2 A Class of Wavelets for Graphs

Let $G = (V, E)$ be a connected graph, of degree $n = |G|$, corresponding to a network of interest. Without loss of generality, we assume that measurements correspond to vertices v_k, $k = 1, \ldots, n$. That is, G is either the network graph itself or the network line graph (as defined in Section 3.2) depending on whether the actual measurements are taken at routers or on links of the underlying network.

The vertex set V is a (finite) metric space when equipped with the shortest path distance metric (in units of "hops") defined with respect to G. In fact, it is a measure space when equipped with simple counting measure as well, say $\mu(\cdot)$. In extending the notion of wavelets to graphs, we seek a collection of functions $\Psi_{j,k} : V \rightarrow \mathbb{R}$, localized with respect to a range of scale/location indices (j, k), such that at a minimum we have $\int_V \Psi_{j,k}(v)\mu(dv) = 0$. Additional properties are built in by choice.

As foreshadowed by our discussion in Section 3.2, the construction of our graph wavelets centers on the notion of comparing network measurements within a given region(s) (e.g., a simple "disk") to those in a surrounding region(s) (e.g., a simple "ring"), with both sets of regions centered on a particular vertex v_k and calibrated to a scale j. The notion of regions will be made explicit through the concept of hop-neighborhoods. Specifically, we define the h-hop neighborhood $N_h(v_k)$ about v_k, $h \geq 0$, to be the set of vertices $v \in V$ that are less than or equal to h hops from v_k. The zero-hop neighborhood of v_k will simply be v_k itself i.e., $N_0(v_k) = \{v_k\}$. Similarly, we let

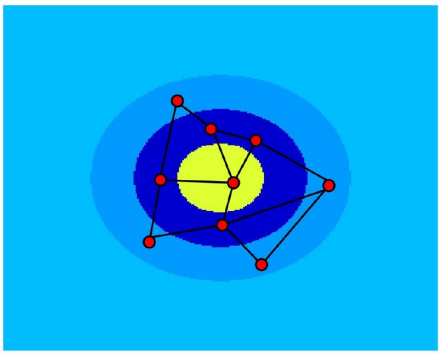


Figure 4: Schematic Illustration of Graph Wavelet Weighting Scheme: Weights obtained from analogue of mexican hat wavelet.

$N'_h(v_k) \equiv N_h(v_k) \setminus N_{h-1}(v_k)$ be the set of vertices exactly h hops away from v_k. We call $N'_h(v_k)$ the h-hop ring about v_k.

In addition to the condition of having zero integral, we will require that each function $\Psi_{j,k}$ be constant within hop rings $N'_h(v_k)$ and zero on hop rings outside $N_j(v_k)$. These constraints have the effect of imposing a type of symmetry on $\Psi_{j,k}$ and a scaling of the underlying support. Figure 4 shows an illustration of this effect, based on the construction given below, which may be compared to Figure 3 for example.

Let J_k denote the largest h for which the hop ring $N'_h(v_k)$ is non-empty. Given the nature of the graph topology, in contrast to that of Euclidean space, this is a necessary and well-defined parameter in our construction. Within this framework there is still a good deal of freedom in choosing the form of our wavelet functions. Specifically, we note that our symmetry condition implies that definition of $\Psi_{j,k}$ can be reduced to that of a set of constants $c_{j,k,h}$ on rings $N'_h(v_k)$, for $h = 0, \ldots, j$. Note then that

$$\begin{aligned} 0 &= \int_V \Psi_{j,k}(v)\mu(dv) \\ &= \sum_{h=0}^{j} c_{j,k,h} \int_{N'_h(v_k)} \mu(dv) \\ &= \sum_{h=0}^{j} c_{j,k,h} |N'_h(v_k)| \quad . \end{aligned} \tag{1}$$

With the choice of $c_{j,k,h} \equiv \psi_{j,h}/|N'_h(v_k)|$, we reduce the problem further to that of finding an appropriate set of constants $\psi_{j,h}$ that depend only on scale j and hop distances $h \leq j$. But it can be seen from (1) that, for each location $k = 1, \ldots, n$ and scale $j \leq J_k$, $\int_V \Psi_{j,k}(v)\mu(dv) = 0$ if and only if $\sum_{h=0}^{j} \psi_{j,h} = 0$.

We therefore have the following result regarding wavelets on graphs.

Theorem 1 *Let $G = (V, E)$ be as above. For each $k = 1, \ldots, n$ and $j \leq J_k$, define $\Psi_{j,k} : V \to \mathbb{R}$ as*

$$\Psi_{j,k}(v) = \mathcal{C}_{j,k} \sum_{h=0}^{j} \frac{\psi_{j,h}}{|N'_h(v_k)|} \mathcal{I}_{k,h}(v) \quad , \tag{2}$$

where $\sum_{h=0}^{j} \psi_{j,h} = 0$ *(excluding the trivial case* $\psi_{j,h} \equiv 0$*),*

$$\mathcal{I}_{h,k}(v) = \begin{cases} 1, & \textit{if } v \in N'_h(v_k) \\ 0, & \textit{otherwise} \end{cases} \quad ,$$

and $\mathcal{C}_{j,k} = \left(\sum_{h=0}^{j} \psi_{j,h}^2 / |N'_h(v_k)| \right)^{-1/2}$ *is a normalizing constant.*

Then the system of functions $\{\Psi_{j,k}\}$ *satisfies*

$$\int_V \Psi_{j,k}(v)\mu(dv) = 0 \qquad \textit{and} \qquad \int_V \Psi_{j,k}^2(v)\mu(dv) = 1 \ ,$$

and for functions $f : V \to I\!R$

$$\langle f, \Psi_{j,k} \rangle = \mathcal{C}_{j,k} \sum_{h=0}^{j} \psi_{j,h} \, \textit{Ave}_{N'_h(v_k)}(f) \quad . \tag{3}$$

It can be seen that our system of graph wavelets most closely resembles the traditional continuous wavelet transform in spirit, as opposed to the discrete wavelet transform, in that the set of wavelet coefficients produced under our transform are indexed at all scales and all locations within scale allowed by the graph topology. Note, however, that these wavelets are not defined as translations of some single simple function Ψ. Nor, while there is an explicit scaling in their support, does the actual "shape" of the wavelets have to necessarily scale with j in any obvious manner. The lack of translation-invariance is unavoidable, due to the underlying network topology. However, a sense of scaling of shape is often desireable for the purpose of comparing wavelet coefficients across scales or between two different types of wavelets, and may be accomplished by linking the choice of $\psi_{j,h}$'s to some common mechanism.

For example, a straightforward approach is to begin with a function $\psi(x)$ supported (perhaps after appropriate re-scaling) on the unit interval $[0, 1)$, for which $\int_0^1 \psi(x)dx = 0$, and the shape of which is appealing for analysis on the real line. Then, letting the $\psi_{j,h}$ be averages of ψ on equal-length subintervals $I_{j,h} \equiv [h/(j+1), (h+1)/(j+1))$ i.e., $\psi_{j,h} \equiv (j+1) \int_{I_{j,h}} \psi(x)dx$, it follows automatically that the $\psi_{j,h}$ sum to zero and that their "shape" scales with j.

We illustrate this approach with two choices of ψ, taken from the functions shown in Figure 2. Each function is restricted to the positive real line, truncated at $x = 4$, re-scaled so as to live on $[0, 1)$, and re-normalized so as to maintain the properties of zero integral and unit norm. In the case of the Haar wavelet function, for j odd, this results in the first $(j+1)/2$ of the $\psi_{j,h}$ being equal and positive, the second $(j+1)/2$ being equal and negative, and all being equal in absolute value. The wavelet coefficient of $\Psi_{j,k}$ is then essentially the difference of averages of measurements on the disc $N_{(j-1)/2}(v_k)$ and the ring $N_j(v_k) \setminus N_{(j-1)/2}(v_k)$. However, when j is even, we will have $\psi_{j,h} = 0$ for $h = j/2$. The wavelet coefficients are then still proportional to the difference of averages on a disc and surrounding ring of equal 'radius', but now the disc and ring are separated by a thin ring whose measurements are ignored (due to their ring corresponding to this zero coefficient). If one instead derives the $\psi_{j,h}$ from the other function shown in Figure 2 (*i.e.*, the analogue of the mexican hat wavelet), there is no such phenomenon for j either even or odd, though the coefficients now yield a more general weighted average of measurements on the rings in $N_j(v_k)$.

Figure 5 shows the coefficients resulting in the case of $j = 2$, for both of these choices of $\psi(x)$. The two wavelet functions have been modified to sit on the unit interval, as described above. The

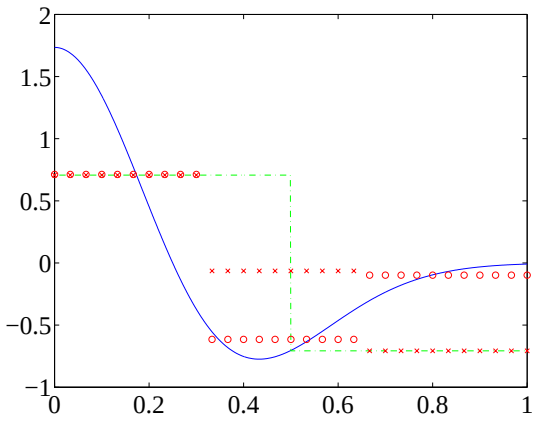


Figure 5: Average Sampling of Haar (dot-dashed) and Mexican Hat (solid) Wavelets, $j = 2$: Wavelet functions in Figure 2 are adjusted to the unit interval. Weights obtained by average sampling are shown using piecewise constant segments of appropriate height and length ($\times$: Haar; $\circ$: Mexican Hat).

values of the weights $\psi_{j,h}$, for $h = 0, 1,$ and 2, obtained by average sampling of the ψ functions, are shown on the same plot through the use of piecewise constant segments of appropriate height and length. For this particular choice of j the values of the $\psi_{j,0}$ weights are almost identical, thus indicating a common treatment of the values of f at the central vertex v_k for each wavelet coefficient $\langle f, \Psi_{j,k} \rangle$ under either choice of wavelet. However, the treatment of values at vertices of $h = 1$ or 2 hops from v_k receive almost opposite treatment by the two wavelet functions, although whereas $\psi_{j,1} = 0$ exactly for the Haar wavelet $\psi_{j,2}$ is only close to zero for the mexican hat wavelet. Of course, for larger choice of j the overall pattern ("shape") in the $\psi_{j,h}$ becomes more rich for the mexican hat wavelet, while that for the Haar wavelet remains unchanged (other than the fluctuations with odd and even j, as described above).

Figure 4 provides a schematic illustration of how the $\psi_{j,h}$ relate to a hypothetical network graph local to a vertex v_k, for $j = 2$. Weights from Figure 5 for the mexican hat wavelet have been used. The result may be compared to the image representation of the actual two-dimensional mexican hat wavelet in Figure 3.

On a final note, we mention that numerical computation of our graph wavelet coefficients is straightforward. A direct implementation of the formulas underlying Theorem 1, using the $n \times n$ adjacency matrix of the graph V, yields an algorithm that produces all coefficients at a single scale j in time $O(n^2)$. We have used this direct algorithm to develop the results shown next in Section V.

5 Graph Wavelets in Practice

To illustrate the utility and features of traffic analysis based on graph wavelets we apply it to the most fundamental measure of network performance: bytes carried per unit time. We use measurements from the Abilene network, taken by polling RMON2 MIBs on core routers, which includes counts of bytes and packets flowing over each of the links shown in Figure 1. We aggregate the bytes flowing over each link, placing the measurements into 30 second intervals on a common

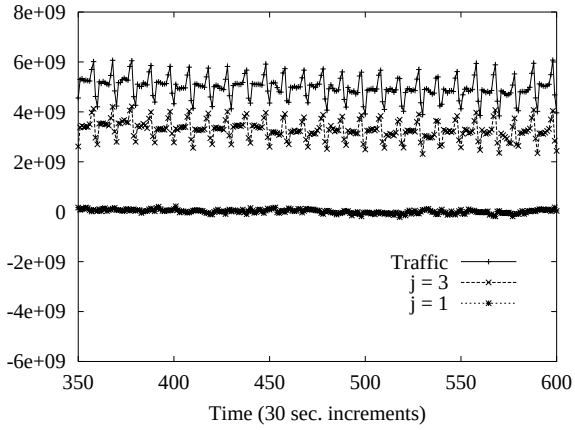


Figure 6: Isolating Periodicity on the NYC-Cleveland Link

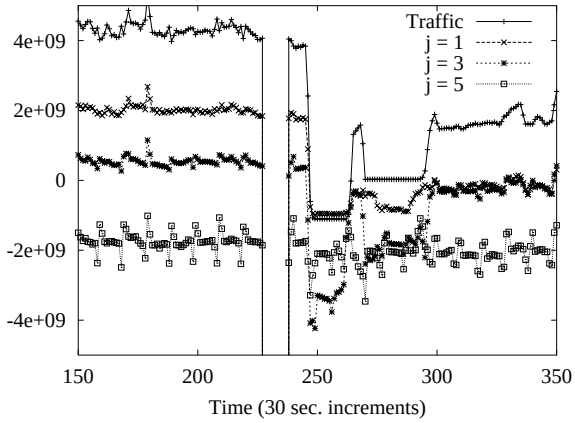


Figure 7: Haar coefficents for the Denver-Kansas City Link during Failure Event

timebase. For simplicity in this example, we sum the bytes flowing in both directions over any link; however our analysis can be extended to the case in which traffic in different directions is considered separately. The measurement period we focus on is Wednesday, April 24, 2002 starting at approximately 11:00pm UTC.

For simplicity, and to integrate the material of Section IV with the motivation in Section III-B in a more immediate fashion, our first few examples will use the Haar-style wavelet for our analysis; later we contrast the features of the Haar and the mexican hat wavelets as applied to our data. Also, as described in Section 4.2, the Haar-style wavelet involves zero weightings for certain rings when j is even; therefore, to simplify interpretation of the analysis results further, we will restrict our attention initially to analyses where j is odd, and address the case where j is even later.

Our first example is shown in Figure 6. This is a plot of traffic on the NYC-Cleveland link during a typical portion of the measurement period (upper line), along with the values for coefficients when $j = 1$ (lower line) and $j = 3$ (middle line). Note the periodic traffic pattern on the link. Understanding the nature of this anomalous behavior effect falls under the category of "identifying regions exhibiting traffic anomalies" as discussed in Section 3.1. This periodicity may be introduced by a dominant application, by interactions between applications or flows, or by an

Figure 8: Network-Wide View of Wavelet Coefficients

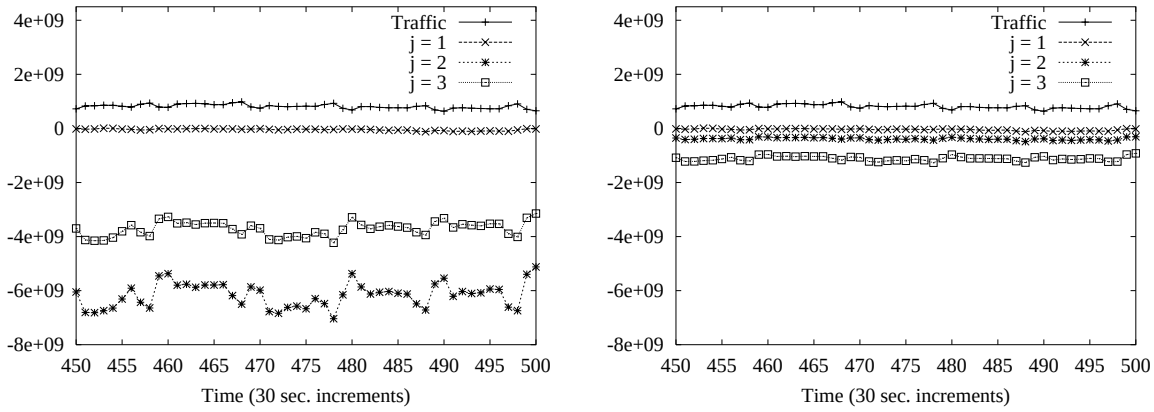
incipient fault in a network element such as a router. An important question therefore is where else on the network this perodicity is visible; that is, over what spatial neighborhood is the periodicity occurring?

In this case we can immediately infer the extent of the periodicity by inspecting the $j = 1$ and $j = 3$ coefficients. The $j = 1$ comparison shows little evidence of periodicity, indicating that the link itself is behaving very similarly to its one-hop neighbors (NYC-Washington and Cleveland-Indianapolis). On the other hand, the $j = 3$ comparison does show evidence of periodicity, indicating that the periodic traffic pattern on these three links (Washington-NYC-Cleveland-Indianapolis) is different from the behavior of the wider neighborhood – that is, the effect is localized to the one-hop neighborhood.

Our next example is shown in Figure 7. This is a plot of traffic on the Denver-Kansas City link spanning the period of a service outage on this link. This event falls under our discussion of "identifying regions exhibiting traffic anomalies" in Section 3.1, *i.e.*, given that there is a change in the traffic on a link, how widespread through the network is this change? We can look at wavelet coefficients before and after the event to answer this question. As shown in the Figure, traffic drops to approximately half its previous level as a result of the failure; coefficients at $j = 1$ and 3 show drops as well, but the drop is greater at $j = 1$ than 3; and $j = 5$ shows virtually no change. This tells us that this traffic change is a local one, but not a network-wide event (since the largest-scale coefficient is unchanged).

In fact, we can also learn about network-wide behavior during this service outage event. Figure 8 shows two sets of wavelet coefficients for each link in the network (data for the Indianapolis-Atlanta link is not shown to lessen clutter). In each inset chart, traffic is shown in the top bar, and coefficients for $j = 1, 3,$ and 5 are shown in successive bars underneath. The left hand bars show

coefficients before the service outage event, and the right hand bars show coefficients during the service outage event.



(a) Haar Wavelet (b) Mexican Hat Wavelet

Figure 9: Comparison of Haar and Mexican Hat Wavelets; Atlanta-Houston Link

Looking at the $j = 3$ and 5 coefficients (lowest set of bars) before the event versus during the event, it is clear that the network's response to the service outage was to shift traffic from its northern path to its southern path. That is, links along the path NYC-Cleveland-Indiapolis-Denver-Sunnyvale had generally negative (or small) coefficients during the event, while links along the path NYC-WashDC-Atlanta-Houston-Los Angeles-Sunnyvale had generally positive coefficients. This suggests that the two intercoastal paths are principle paths through the network, shedding light on typical traffic source-destination paths. These comparisons highlight the utility of wavelet-based analysis to capture system-wide effects.

In contrast to this north-south effect, wavelet analysis can inform us about traffic levels on a east-west axis as well. Looking at the $j = 5$ coefficients across the network before the outage (left hand set of bars), it is clear that they are largest (most positive) in the east, and smallest (most negative) in the west. This tells us that in the period before the failure, the network was generally more heavily loaded in its eastern region as compared to its western region.

Finally, we consider how the choice of wavelet function ψ affects the properties of the transform. To do so, we examine the regions around the Atlanta-Houston link. Analyzing using the Haar wavelet yields the results shown in Figure 9(a). This plot shows that the $j = 1$ coefficients are close to zero, while the $j = 2$ coefficients are strongly negative. To interpret this recall that even numbered Haar coefficients include a ring with zero weight, which in the $j = 2$ case is the one-hop ring (as shown in Figure 5). So these coefficients show that the traffic on this link is similar to its immediate neighbors, but on average is much less than traffic in the ring that is two hops away. The $j = 3$ coefficients are not as strongly negative, which shows that the average of the link and its one-hop neighbors is comparatively more similar to the two- and three-hop rings.

This abrupt change in values from $j = 1$ to $j = 2$ to $j = 3$ is an indirect effect of the lack of smoothness of the underlying ψ function used in the Haar case (see Figure 2). This lack of smoothness places a sharp emphasis on the difference between the inner disk and outer ring.

Replacing the Haar ψ with the mexican hat ψ yields more gradually changing coefficients. For $j = 1, 2,$ and 3, the mexican hat ψ has positive weight only at $h = 0$, that is, on the central link.

Furthermore, the most negative values move out more slowly than in the Haar case; for example, when $j = 2$ the Haar emphasizes the two-hop ring while the mexican hat emphasizes the one-hop ring (see Figure 5). This effect can be seen in Figure 9(b). The figure shows that, as the traffic comparison moves outward by rings, the difference between the traffic on the Atlanta-Houston link and its weighted neighbors changes gradually and smoothly.

The difference gives insight into the relative utility of the different ψ functions. The Haar wavelet is useful for precise distinctions between rings (as in Figure 6); however, it is quite aggressive, with each new level j averaging in another ring in full measure. The mexican hat wavelet is more gradual and better suited for finer study of the successive neighborhoods of a particular node.

6 Conclusion

As discussed in Section 1, in contrast to the traditional use of wavelets in traffic analysis, we apply wavelets in the spatial domain rather than the temporal domain. However an even higher goal is to build tools that can fuse these two views (spatial and temporal analysis) into a single framework that allows network operators and engineers to ask questions such as: "Is this region of the network experiencing unusual conditions? How long has this been going on?" Achieving this combined spatio-temporal view is a considerable challenge; what we present here is one step toward that goal.

More generally, we hope to move toward study of system-wide effects in large scale networks. The interaction of network topology and traffic characteristics is a potentially important but largely unexplored area, in part due to lack of good tools. We believe that it will be easier to understand how whole networks behave as spatial traffic analysis tools like those we describe here mature.

Lastly, we note that for some network analysis problems, graph wavelets analogous to the continuous wavelet transform (as described in this paper) might be less appropriate than an approach analogous to the discrete-orthogonal wavelet transform. The discrete transform differs most fundamentally from the continuous transform in the sense that the discrete transform involves wavelets constructed to sit on a nested series of coarsenings of the underlying data space. While the concept of coarsening is straightforward in Euclidean space, it is much less so on arbitrary graph topologies (*e.g.*, see [21]). However, just as the development of the continuous wavelet transform was a precursor to that of the discrete-orthogonal wavelet transform in the context of traditional wavelets, the approach we describe here may serve as a precursor for an analogue of the discrete wavelet transform for graphs.

Acknowledgments

We gratefully acknowledge the help of Paul Barford and Joel Sommers at U. Wisconsin, and Guy Almes, Matt Zekauskas and Steve Corbato at Internet2 for providing us with access to Abilene traffic measurements.

References

[1] Walter Willinger, Murad Taqqu, and Ashok Erramilli, "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks," *Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series*, vol. 4, 1996.

[2] "Internet2," Available at *http://www.internet2.edu*.

[3] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic," in *Proceedings of SIGCOMM '93*, September 1993, pp. 183–193.

[4] Walter Willinger, Murad S. Taqqu, Robert Sherman, and Daniel V. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," in *Proceedings of ACM SIGCOMM '95*, 1995, pp. 100–113.

[5] J. Levy-Vehel and R. Riedi, "Fractional Brownian motion and data traffic modeling: the other end of the spectrum," in *Fractals in Engineering*, pp. 185–202. Springer-Verlag, Berlin, 1997.

[6] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic," in *Proceedings of SIGCOMM '98*, October 1998, pp. 42–55.

[7] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Transactions on Information Theory*, vol. 44, pp. 2–15, 1998.

[8] Frank Feather, Daniel P. Siewiorek, and Roy A. Maxion, "Fault detection in an ethernet network using anomaly signature matching," in *Proceedings of ACM SIGCOMM '93*, 1993, pp. 279–288.

[9] Joseph L. Hellerstein, Fan Zhang, and Perwez Shahabuddin, "A statistical approach to predictive detection," *Computer Networks*, January 2000.

[10] Polly Huang, Anja Feldmann, and Walter Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *Procedings of the 2001 ACM Internet Measurement Workshop*, 2001.

[11] Paul Barford and Dave Plonka, "Characteristics of network traffic flow anomalies," in *Proceedings of ACM SIGCOMM Internet Measurement Workshop*, July 2001.

[12] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Pennsylvania, 1992.

[13] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31:4, pp. 532–540, 1983.

[14] M. Girardi and W. Sweldens, "A new class of unbalanced Haar wavelets that form an unconditional basis for l_p on general measure spaces," *Journal of Fourier Analysis and Applications*, vol. 3:4, pp. 457–474, 1997.

[15] I. Guskov, W. Sweldens, and P. Schröder, "Multiresolution signal processing for meshes," *Computer Graphics Proceedings (SIGGRAPH 99)*, pp. 325–334, 1999.

[16] T.H. Li, "Multiscale representation and analysis of spherical data by spherical wavelets," *SIAM Journal of Scientific Computing*, vol. 21, pp. 924–953, 1999.

[17] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, Cambridge, 1998.

[18] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1:2, pp. 205–220, 1992.

[19] W. Sweldens, "The lifting scheme: a construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29:2, pp. 511–546, 1997.

[20] Peter Schröder and Wim Sweldens, "Spherical wavelets: Efficiently representing functions on the sphere," *Computer Graphics Proceedings (SIGGRAPH 95)*, pp. 161–172, 1995.

[21] S-H. Teng, "Coarsening, samping, and smoothing: elements of the multilevel method," in *Algorithms for Parallel Processing*, M.T. Heath, A. Ranade, and R.S. Schreiber, Eds., IMA Volumes in Mathematics and its Applications. Springer-Verlag, New York, 1999.