Towards a Meaningful MRA of Traffic Matrices

David Rincón
Univ. Politècnica de Catalunya
Av. Canal Olímpic,15
Castelldefels 08860
Barcelona, Spain
drincon@entel.upc.edu

Matthew Roughan
School of Mathematical Sciences
University of Adelaide
SA 5005 Australia
matthew.roughan@adelaide.edu.au

Walter Willinger
AT&T Labs - Research
180 Park Avenue
Florham Park, NJ 07932, USA
walter@research.att.com

ABSTRACT

Most research on traffic matrices (TM) has focused on finding models that help with inference, but not with other important tasks such as synthesis of TMs, traffic prediction, or anomaly detection. In this paper we approach the problem of a general model for traffic matrices, and argue that such a model must be sparse, i.e., have a small number of parameters in comparison to the size of the TM. A Multi-Resolution Analysis (MRA) of TMs can provide such a sparse representation. The Diffusion Wavelet (DW) transform is a good choice as a MRA tool here, because it inherently adapts to the structure of the underlying network. The paper describes our construction of the two-dimensional version of the DW transform and shows how to use it for our proposed MRA of TMs. The results obtained with operational networks confirm the sparseness of the DW-based TM analysis approach and its applicability to other TM-related tasks.

Categories and Subject Descriptors

C.2.5 [**Computer Communications**]: Local and Wide Area Networks—*Internet*; C.4 [**Performance of Systems**]: Modeling Techniques.

General Terms

Algorithms, Measurement.

Keywords

Traffic Characterization, Traffic Matrices, Diffusion Wavelets, Multi-Resolution Analysis.

1. INTRODUCTION

Internet Traffic Matrices (TMs), giving traffic volumes from ingress to egress nodes in a network, have drawn considerable interest (see [1] and the references therein). TMs are a basic input to many network engineering tasks, but they are non-trivial to measure, and so much work has gone into measurement [8] or their indirect inference [2, 14, 19, 20, 21] from readily available link load measurements. There have been a number of practical outcomes as a result [17].

The papers about *inference* of traffic matrices have found utility because they use easily obtained measurements. However, the link load measurements commonly collected from operational networks do not provide enough information to form a well-posed inference problem, with the consequence that some type of side information (usually in the form of a traffic matrix model) is needed to perform the inference.

But modeling TMs is not just about inference. There are other tasks such as TM synthesis [9, 15], giving one the control needed to generate many samples with precisely controlled parameters, in order to undertake performance analysis of, for instance, new traffic engineering algorithms. It is critical in such synthesis that the important properties of real TMs are accurately represented in the synthetic TMs, otherwise incorrect results may arise. However, the properties of real-world TMs have not been exhaustively studied as yet, and it is not currently known which are most important. This problem is magnified by a lack of publicly available traffic data from commercial ISPs.

Network engineering with TMs also needs models. For instance, it is common to use TMs as part of the design of the layout of a new or re-designed network [7]. As such, we need to *predict* a traffic matrix at some time in the future. In addition, we may wish to detect anomalous traffic behaviour, and a simple approach to such anomaly detection is to look for large deviations from predicted behaviour.

Most research on traffic matrices has focused almost exclusively on inference, but not the other tasks. The problem of finding a "good" model for TMs is problem dependent, and so quite open. In this paper we suggest one key criterion for a TM model and point out a tool for constructing and estimating such a model.

The criterion we focus on here is that the traffic matrix model should be *sparse*. A traffic matrix for a network with N nodes has N^2 terms, and since N can be in the thousands, the number of terms in the traffic matrix can become very large. A sparse model has a number of parameters $M \ll N^2$. There are good reasons to search for a sparse model:

- In general, there is a tradeoff between model fidelity and the model's predictive power. For instance, by having a large number of parameters we may have a model that works well for one set of data, but does not provide good predictions because it is too specific.
- If the model has few parameters then we have more hope of attaching physical meaning to these parame-

IMC'08, October 20–22, 2008, Vouliagmeni, Greece.

ters, with the result that the "magic" of tuning them appropriately for new network settings can be replaced by engineering insight.

- The inference problem is ill-posed when we have $K = O(N)$ link-load measurements, but N^2 parameters that we need to estimate. If the TM model had $M \leq K$ parameters, then the problem might become well posed.

Often, approaches to TM inference have sought some kind of sparse model for the matrices, with a view that this will bring the problem back towards being well-posed. The gravity model [20] is a good example, with only $2N$ parameters. However, in this case the model itself is not a particularly accurate representation of a TM, it simply forms a *prior* used in a regularization approach for inference. We will not discuss all of the possible models, but note that the search for sparsity was not the *explicit* criterion for previous models. The majority of existing models have generally been argued from background engineering knowledge of the Internet.

Our approach for finding such a sparse model is to use Multi-Resolution Analysis (MRA). Wavelet-based MRA techniques for "denoising" seek a sparse model for a signal by soft-thresholding the transformed signal. Such approaches have a number of advantages: long-range correlations are approximately decorrelated in the wavelet domain, reducing dependencies between potential model parameters; fast algorithms exist for wavelet transforms; and standard wavelet transforms are linear, leading to a number of desirable properties. Most importantly, many of the "perceptible" features of signals are preserved in relatively few wavelet coefficients. The sparseness of the coefficients is the leading reason why modern compression techniques often use wavelets.

Standard wavelet-based MRA analysis is not appropriate for traffic matrices. In a TM the spatial relationships between elements are more complicated than in an image, which is after all a simple rectangular grid sampling of a two-dimensional field. We exploit the new Diffusion Wavelets (DW) approach [5] and perform multi-resolution analysis of functions defined on graphs. The graph represents the underlying network (over which our traffic matrix is routed), and reflects the natural spatial relationships in the TM. For instance, two traffic matrix elements originating from locations close together in the network may share characteristics such as their diurnal traffic pattern.

Our main contribution is to generalize Diffusion Wavelets to 2D and apply them to modelling traffic matrices, which we see as two-dimensional functions of the nodes. We find that in the Diffusion Wavelets domain traffic matrices are sparse and (surprisingly) stable across time. Indeed, when this stability is broken, it is a hint of the presence of anomalies in the traffic matrix. The results described in this paper were obtained from real data from operational networks and can be considered a "proof of concept", i.e. a first step towards creating viable sparse models of TMs for use in the various tasks mentioned above: *inference, synthesis, and prediction*. Further studies in each of these areas are needed to demonstrate the full potential of the proposed approach.

2. BACKGROUND AND RELATED WORK

An IP network can be abstractly thought of as a graph, whose nodes are routers or Points-of-Presence (PoPs), and whose edges are links between these. A traffic matrix describes the volumes of traffic traversing a network from the point at which it enters the network to the exit point, measured over some time period. Such a matrix is useful in capacity planning, traffic engineering, network reliability analysis, and many other network engineering tasks. It is possible to measure such a matrix using measurement technologies such as flow level traffic collection, but typically these are hard to implement across a large network [20]. On the other hand SNMP data is easy to collect, and almost ubiquitous. However, SNMP data only provides link load measurements, not TM measurements. The link measurements $\mathbf{y}$ are related to the TM, which is written as a column vector $\mathbf{x}$, by the relationship $\mathbf{y} = A\mathbf{x}$ where A is called the routing matrix [19]. The resulting problem of inferring the TM from link measurements is a classic underconstrained, linear-inverse problem which needs some sort of side information. Examples of such additional information used in the context of Internet TMs are a Poisson model [19], a Gaussian model [2], a logit-choice model [14], or a gravity model [20].

Further efforts on modeling the relationships between TM elements have been described in [11] and successfully exploited for anomaly detection in [10]. These papers focused on Principal Components Analysis (PCA) of the traffic matrices as times series. PCA exploits the correlations between TM elements to separate the periodic components of the traffic (see [16]) from random fluctuations and anomalous events. However, it is unclear how the structures described within [11] would lead to a simple model for use in synthesis. On the other hand, the gravity model [20] is so simple that it has already seen extensive use as a model for TMs.

A first approach to apply graph-based MRA to the study of traffic matrices was carried out by Crovella and Kolaczyk [6], where "graph wavelets" were introduced as an extension of the 2D wavelet transform. Graph wavelets provide a way of computing the load differences between links separated by a certain number of nodes (i.e., the concept of scale is replaced by that of hop distance between links). The authors also show how the tool can be used for anomaly detection. The main drawbacks of this approach are the lack of a fast computational algorithm and its non-orthogonality. Graph wavelets do not provide a sparse representation of traffic data but rather an overcomplete decomposition similar to that of the Continuous Wavelet Transform (CWT).

Diffusion Wavelets [5] can be understood as a generalization of graph wavelets, allowing more freedom for choosing the underlying kernel function (or "mother wavelet"), along with generalized distances in the graph, an orthonormal basis, and a fast computation algorithm. To the best of our knowledge, the only relevant application of Diffusion Wavelets in the context of computer networks is [4], where Coates *et al.* address the problem of assessing the number of measurement points required to monitoring end-to-end path metrics, such as delay at the IP level or bit-error-rate performance at the physical level of an optical network. The authors construct a diffusion operator on an alternative graph where the nodes are the routes of the original network, and the links are a similarity measure of the routes (the fraction of shared links, according to the routing data), and apply the (1D) DW transform. The implicit sparsity of the analyzed data in the transformed domain, together with the use of sparse inference techniques, allows an efficient monitoring with a reduced number of devices (for example, network mean end-to-end delay can be measured with high precision by monitoring only 7% of the routes).

3. WAVELETS & DIFFUSION WAVELETS

Wavelet methods have been used in signal and image processing for denoising and compressing, among other applications. The Discrete Wavelet Transform (DWT) analyzes signals by computing its scalar product with dilated (by powers of 2) and translated versions of the *mother wavelet* function, thus analyzing the input signal at time scales $t = 2^j$. Provided that the mother wavelet satisfies certain conditions [13], the resulting transform is orthonormal and can be efficiently implemented by a bank of quadrature-mirror low-pass and high-pass filters ($h(n)$ and $g(n)$, respectively) followed by downsampling, as illustrated in Figure 1.

The low-pass filters perform successive approximation on the signal at coarser and coarser scales. Intuitively, we might think of this as successively "blurring" the original signal. Implicitly, there exists a *mother scaling function* from which we could derive the blurring functions using the same dilations and translations as with the mother wavelet. The wavelet *details* $d_x(j,k)$ capture the difference between the approximation $a_x(j-1,k)$ at some scale $j-1$, and a coarser level of approximation $a_x(j,k)$. In mathematical terms, we obtain a set of nested approximation (*scaling*) subspaces V_j, $V_1 \supset V_2 \supset \ldots \supset V_J$ and their orthogonal complements, the high-frequency detail (*wavelet*) subspaces $W_j = V_{j-1} - V_j$.

In the frequency domain the DWT results in a decomposition in subbands whose spectra are halved at each step. This gives rise to a multiresolution analysis in which the original signal is decomposed into a low frequency approximation at the largest time scale $t = 2^J$, $a_x(J,k)$ and a set of high-frequency details $d_x(j,k)$ (the *wavelet coefficients*) for each time scale $t = 2^j$, $j = 1 \ldots J$. The transform can be generalized to 2D images, as we will see in Section 4.

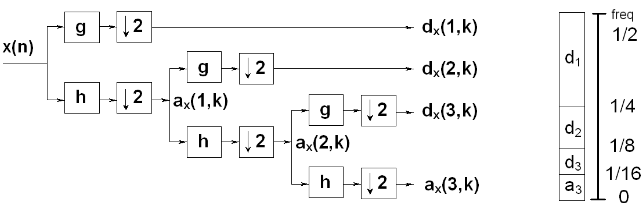

Figure 1: Left: the 1D DWT filter bank for J=3, with approximation $a_x(3,k)$ and details $d_x(j,k), j = 1 \ldots 3$. Right: the associated spectrum subbands.

The aforementioned *classical* time- and space-based wavelet transforms operate on signals defined on uniformly sampled grids on $\mathbb{R}$ and $\mathbb{R}^2$, respectively. However, a TM is not defined on a regular lattice — it is defined across a computer network, which can be represented by a graph. Diffusion Wavelets [5] are a generalization of the wavelet transform in which the MRA can be performed on structures such as manifolds or graphs. In our case the underlying structure is a graph $G\{V,E\}$ (where V and E are the vertex and edges sets, respectively). We wish to analyze a function $f : V \to \mathbb{R}$, i.e., we have a function $f(i)$, which maps each vertex i to a real number.

The approach is to create a *diffusion operator* that plays the role of the mother scaling function. Application of the diffusion operator "blurs" the original function, but in a way that is adapted to the underlying graph. Locations that are close together in the graph will be blurred into each other, while locations that are far apart will remain separated. The use of the underlying graph for the diffusion makes the DW intrinsically adapted to the "topology" over which the functions $f(i)$ are defined.

Mathematically we represent the diffusion operator by a linear transform Tf. Just as there are many possible mother wavelet and scaling functions, there are many choices we could make for T. Simple examples include a heat-like diffusion (hence the name) across the graph, or a stochastic matrix representing a random walk on the matrix. The latter seems a natural choice since it models an approximation of the distance between nodes in a graph, but instead we follow [12] and choose the $I - \mathcal{L}$ operator, where I is the identity matrix and $\mathcal{L}$ is the normalized Laplacian [3] of A, the (weighted) adjacency matrix of the graph[1]. This operator is closely related to the random walk [3, 12] and it has the same eigenvalues but, unlike the random walk, the $I - \mathcal{L}$ operator is symmetric, among other desirable properties. The operator is later scaled in order to be doubly-stochastic.

The dilation operator used to construct subsequent scaling functions is simply to take powers of the matrix T. Intuitively, if a diffusion continues over n time steps, we would apply the linear transform n times, i.e., $T^n f$. This results in successive blurring of the function, as required. In the random walk interpretation, assume f represents an initial distribution of states, then $T^n f$ represents the state distribution after n time steps, which we know will tend to blur (for an irreducible Markov chain) towards the equilibrium distribution. Analogously to the standard wavelet transform, DW progresses in powers of 2, i.e., we consider $T^{2^j} f$.

For graphs, the natural equivalent to the frequency-based decomposition resulting from the DWT is spectral graph theory, i.e., the study of the eigenvalues and eigenfunctions of linear operators [3]. The Spectral Theorem results in a simple representation of the linear operator

$$T = \sum_{i=1} \lambda_i \nu_i^T \nu_i,$$

where λ_i are the eigenvalues of T and ν_i are their associated eigenvectors. If T is a doubly stochastic matrix, $|\lambda_i| \leq 1$.

The principle that underlies PCA (Principal Components Analysis) is that it is common that many eigenvalues of such an operator will be near zero, and thus we may approximate the matrix T through a partial sum. This has been exploited in the direct analysis of traffic matrices [11]. Here, the concept is applied to the graph-diffusion operator T, but we will apply it very conservatively. In our approximation we will ignore eigenvalues $|\lambda_i| \leq \epsilon$, where ϵ is a tunable parameter with small value (ranging from 10^{-3} to 10^{-10} in our experiments). Few (if any) eigenvalues are removed in the first round, but things change when we consider powers of T.

The eigenvalues of T^n are λ_i^n, and the eigenvectors remain invariant with respect to n. As $n \to \infty$ all of the eigenvalues $|\lambda_i| < 1$ will tend to zero, and eventually they will fall below the threshold ϵ. As such, the successive application of the (now approximated) diffusion operator will break the graph spectrum into subbands, much as the classical wavelet transform does.

[1] In a network-related scenario we may wish to make a transition across a "long" link less likely. Weights can (for example) be inversely proportional to the routing weights, obtaining a diffusion operator that is somehow related to load-balancing routing. In the unweighted adjacency matrix case, a constant value (e.g. 1) can be used as weight.

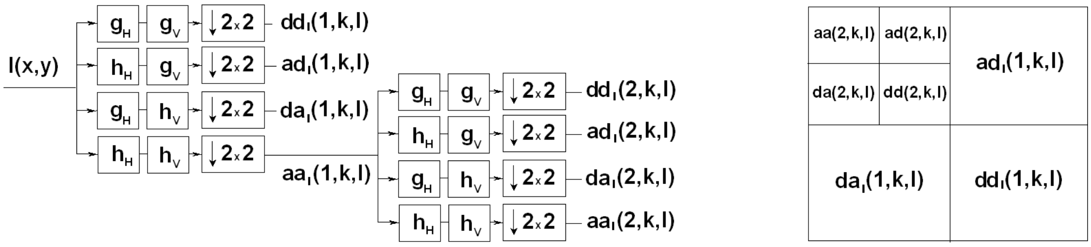

Figure 2: Left: filter bank associated with the 2D wavelet transform, for J=2 scales. Right: the associated subband decomposition.

The MRA can then be defined as follows: for a given scale j, the eigenvectors associated with $|\lambda_i^{2^j}| \geq \epsilon$ span the low frequency approximation subspace V_j, while the eigenvectors associated with the eigenvalues discarded at step j, i.e., such that $|\lambda_i^{2^j}| < \epsilon$ and $|\lambda_i^{2^{j-1}}| \geq \epsilon$ span the high frequency or detail subspace W_j. At each step j the surviving eigenvectors are appropriately reorthonormalized (for example, with a Gram-Schmidt-style algorithm).

Coiffman and Maggioni [5] present a fast algorithm for performing the aforementioned computations, and obtaining the approximation and detail coefficients at level j (denoted as C_{V_j} and C_{W_j}, respectively) by projecting the function under study onto the V_j and W_j subspaces.

4. 2D DW TRANSFORM

Traffic matrices can be represented as two-dimensional functions $F(v_1, v_2)$ of pairs of vertices where v_1 is the ingress node, v_2 is the egress node, and $F(v_1, v_2)$ is the traffic volume from v_1 to v_2. Hence, we need to extend DWs to 2D.

Classical 1D wavelets can be used to construct a separable basis in 2D by combining the application of the low- and high-pass filter banks in the horizontal and vertical dimensions of the input image $I(x, y)$, thus generating 4 subbands at each scale: the low-pass/low-pass approximation $aa_I(j, k, l)$, and the details $ad_I(j, k, l)$, $da_I(j, k, l)$ and $dd_I(j, k, l)$ from the other filter combinations. After appropriately downsampling the outputs, the process is iterated on the approximation subband, as shown in Figure 2.

Analogously, our approach to 2D DWs transforms the function $F(v_1, v_2)$ by projecting it twice onto the approximation and detail subspace bases defined by the 1D DW diffusion operator, once along each "direction". The details of the algorithm vary because we no longer have simple filter bank implementations of the high-pass and low-pass operations, but intuitively the process is similar. We denote by C_{VV_j}, C_{VW_j}, C_{WV_j} and C_{WW_j} the transform coefficients corresponding to the low-pass approximation subspace VV_j and the high-frequency detail subspaces VW_j, WV_j and WW_j, respectively. For the 1D DW transform described here, the resulting 2D transform retains the highly desirable properties of orthonormality, invertibility (i.e. we can exactly reconstruct the original function from its coefficients) and separability.

Figure 3 presents a simple illustrative example on a 10-node graph. The 1D DW transform decomposes the graph spectrum in 3 subbands with 2, 3 and 5 eigenvalues for the W_1, W_2 and V_2 subbands respectively, while the 2D DW divides the graph spectrum into the approximation VV_2 and the details VW_2, WV_2, WW_2, VW_1, WV_1 and WW_1, each one including a set of $n \times m$ eigenvalues/eigenvectors.

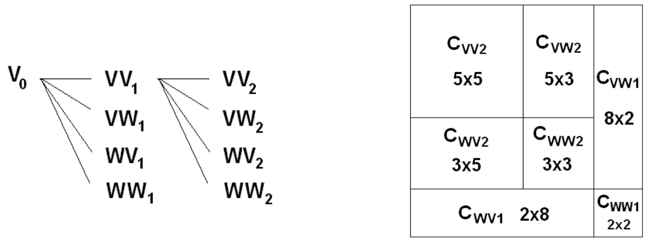

Figure 3: Example of a decomposition generated by the 2D diffusion wavelet transform, for J=2.

5. MRA OF TRAFFIC MATRICES

In our first experiments with the 2D DW tool we have studied over 20000 traffic matrices belonging to two datasets from the Abilene and GEANT networks, with 12 and 23 PoPs, respectively. The granularity of the TMs is 5 minutes in Abilene and 15 minutes for GEANT. For more details about the datasets refer to [18].

The TMs were analyzed with the 2D Diffusion Wavelet transform, with two goals in mind. First, we wanted to visualize how the diffusion process affected a traffic matrix, in order to develop our intuition about the multi-resolution decomposition and check the invertibility and perfect reconstruction properties. Secondly, we wanted to assess the compressibility obtained with the 2D DW.

Figure 4 shows the results obtained with the DW transform of a representative TM from Abilene (March 2^{nd} 2004 from 12:00 to 12:05). The graphs show reconstructions of the TMs from the approximation coefficients (C_{VV_j}) and detail coefficients (C_{VW_j}, C_{WV_j} and C_{WW_j}) at each scale j. The analysis has been performed with the unweighted and normalized random walk as diffusion operator and a precision $\epsilon = 10^{-7}$. Since no eigenvalues are discarded in the first and second detail subspaces, the associated approximation reconstructions are identical to the original traffic matrix and are not shown in the figure. We can see clearly the low-frequency effect of the successive blurring applied by the diffusion operator in the approximations, together with the high-frequency components of the details.

Regarding the compressibility of TMs we have performed several tests on fortnight-long and month-long series from both datasets in order to assess the extent to which the energy of the original matrices is compressed in a few coefficients. Table 1 shows the results for two representative, month-long traces. The results obtained with other traces are consistent with those shown in the table, and confirm the sparseness of the DW representation: on average, 15% of the coefficients retain more than 90% of the original TM energy. Figure 5 supports the results of this compressibility study by

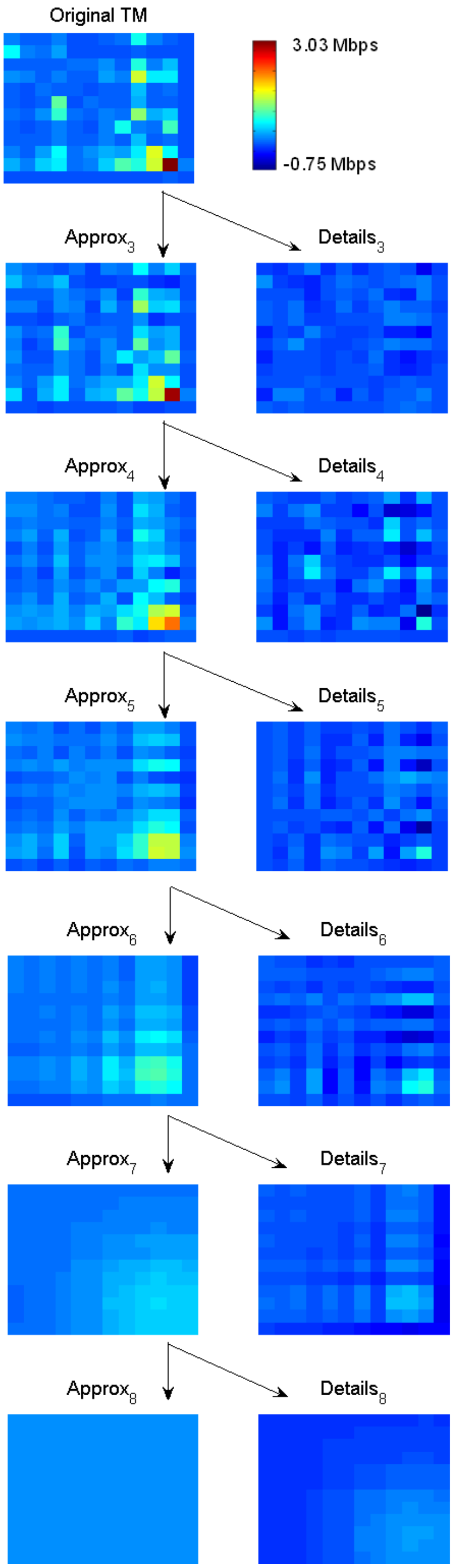

Figure 4: Approximations and details of the decomposition of the TM-2004-03-02-1200 Abilene TM.

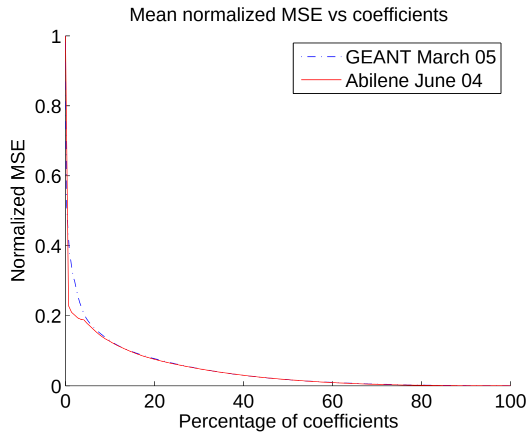

Figure 5: Mean normalized MSE vs percentage of coefficients for two month-long TM traces.

	Energy preserved	
Trace under study	80%	90%
Abilene June 04	4 coeffs (2.8%)	21 coeffs (14.6%)
GEANT March 05	24 coeffs (4.5%)	75 coeffs (14.2%)

Table 1: Percentage of coefficients needed to preserve certain fractions of the original TM energy.

plotting the Mean Square Error (MSE) of the reconstructed TMs versus the percentage of coefficients used in the reconstruction. Note that despite the major differences between Abilene and GEANT the results are almost identical, illustrating how the DW is able to capture the structure of the TM independently of the actual network topology.

Finally, we have found that there is a characteristic *signature* related to the rank of the coefficients (ordered in terms of their contribution to the TM's total energy), which is also consistent across time, except when there is some anomaly in the TM series (such as a sudden traffic volume change). Figure 6 shows an example of such a feature. The figure shows the rank of the biggest 20 coefficients of the 14-day long March 2004 Abilene trace. A change can be clearly seen around the 3500th TM (in the first hours of March 14th), which coincides in time with a notable structural change in the original traffic matrices (a big increase in the traffic directed to Houston from many sources). This suggests that structural TM changes can be detected by monitoring just a small number of coefficients (instead of all 144 paths) as long as they make up the bulk of the TM's total energy. This is still far from becoming a fully working MRA-based anomaly detection algorithm, but the DW transform seems potentially useful for such a task.

6. CONCLUSIONS

This paper presents a proof of concept, in the sense that we have shown that diffusion wavelets provide a sparse approximation of traffic matrices, and hence are a promising approach to modelling said TMs. However, more work has to be carried on (i) transforming the compressibility of the traffic matrix into a physical model; (ii) using this model to solve the various tasks at hand: inference (exploiting the dimensionality reduction), synthesis, prediction and anomaly detection; (iii) investigating how the sparse DW model is

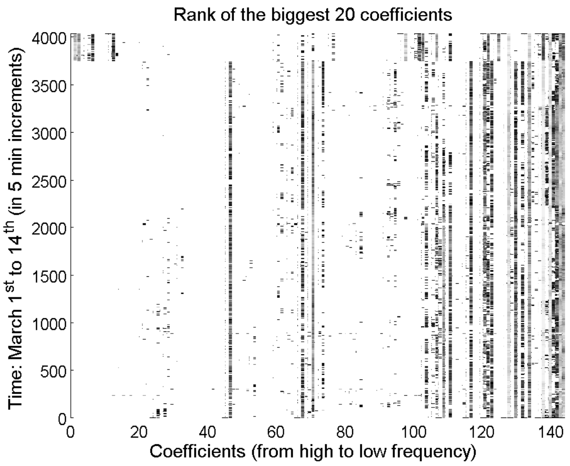

Figure 6: Coefficient signature of the month-long Abilene March 2004 trace.

related to other modelling efforts such as the gravity model or the PCA-based models, to mention some of them; (iv) exploring other diffusion operators (e.g. the weighted adjacency matrix case in the $I - \mathcal{L}$ operator); and (v) investigating how network topologies can be represented and studied with the MRA approach.

7. ACKNOWLEDGMENTS

This work was done while D. Rincón was visiting the University of Adelaide. D. Rincón was supported partially by UPC (Mobilitat 2007/08), the Spanish Government (TSI-2005-06092), and by EuroNF (INFSO-ICT-216366). M. Roughan was supported in part by ARC grant DP0665427.

The authors are grateful to Yin Zhang of the University of Texas for the Abilene data[2], and to S. Uhlig and the TOTEM project team [18] for the GEANT data[3]. The authors also wish to thank the anonymous reviewers for their insightful comments and suggestions.

8. REFERENCES

[1] D. L. Alderson, H. Chang, M. Roughan, S. Uhlig, and W. Willinger. The many facets of Internet topology and traffic. *Networks and Heterogeneous Media*, 1(4):569–600, December 2006.

[2] J. Cao, D. Davis, S. V. Wiel, and B. Yu. Time-varying network tomography. *Journal of the American Statistical Association*, 95(452):1063–1075, 2000.

[3] F. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, February 1997.

[4] M. Coates, Y. Pointurier, and M. Rabbat. Compressed network monitoring for IP and all-optical networks. In *Proceedings of ACM IMC 2007*, pages 241–252, 2007.

[5] R. R. Coifman and M. Maggioni. Diffusion Wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, July 2006.

[2] `http://www.cs.utexas.edu/ yzhang/research/AbileneTM/`

[3] `http://totem.run.montefiore.ulg.ac.be/datatools.html`

[6] M. Crovella and E. Kolaczyk. Graph wavelets for spatial traffic analysis. In *Proceedings of IEEE Infocom*, pages 1848–1857, Apr. 2003.

[7] J. C. Doyle, D. Alderson, L. Li, S. H. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. The "robust yet fragile" nature of the Internet. In *Procs. of the Natl. Academy of Sciences*, volume 102(40), pages 14123–14475, Oct. 2005.

[8] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True. Deriving traffic demands for operational IP networks: Methodology and experience. *IEEE/ACM Transactions on Networking*, pages 265–279, June 2001.

[9] B. Fortz, J. Rexford, and M. Thorup. Traffic engineering with traditional IP routing protocols. *IEEE Comm. Magazine*, 40(10):118–124, Oct. 2002.

[10] A. Lakhina, M. Crovella, and C. Diot. Diagnosing network-wide traffic anomalies. *Computer Communication Review*, 34(4):219–230, October 2004.

[11] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft. Structural analysis of network traffic flows. *SIGMETRICS Perf. Eval. Rev.*, 32(1):61–72, 2004.

[12] S. Mahadevan and M. Maggioni. Proto-value Functions: A Laplacian Framework for Learning Representation and Control in Markov Decision Processes. *The Journal of Machine Learning Research*, 8:2169–2231, 2007.

[13] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[14] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: existing techniques and new directions. *SIGCOMM Comput. Commun. Rev.*, 32(4):161–174, 2002.

[15] A. Nucci, A. Sridharan, and N. Taft. The problem of synthetically generating IP traffic matrices: Initial recommendations. *SIGCOMM Comput. Commun. Rev.*, 35(3), 2005.

[16] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang. Experience in measuring Internet backbone traffic variability: Models, metrics, measurements and meaning. In *Procs. of ITC-18*, pages 379–388, Berlin, Germany, 2003.

[17] M. Roughan, M. Thorup, and Y. Zhang. Traffic engineering with estimated traffic matrices. In *Proceedings of ACM IMC 2003*, pages 248–258, 2003.

[18] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon. Providing public intradomain traffic matrices to the research community. *SIGCOMM Computer Communication Review*, 36(1):83–86, 2006.

[19] Y. Vardi. Network tomography: estimating source-destination traffic intensities from link data. *J. of the Am. Statistical Association*, 91:365–377, 1996.

[20] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale IP traffic matrices from link loads. In *ACM SIGMETRICS*, pages 206–217, San Diego, California, June 2003.

[21] Y. Zhang, M. Roughan, C. Lund, and D. Donoho. An information-theoretic approach to traffic matrix estimation. In *ACM SIGCOMM*, pages 301–312, August 2003.