Interpretable Structural Analysis of Traffic Matrix

Awnish Kumar [*1] **Vijaya V. Saradhi** [*1] **T. Venkatesh** [*1]

Abstract

Structural analysis is concerned with the decomposition of traffic matrix into basis vectors, which corresponds to temporal patterns. In general, the effectiveness of basis vectors is determined by the extent to which it approximates the current week as well as subsequent consecutive week traffic matrix, i.e., the basis vectors should be temporally stable. Principal component analysis (PCA) is the most commonly employed matrix decomposition method in literature. Unfortunately being the linear combination of up to all OD flows, the basis vectors of PCA are i) notoriously difficult to interpret in terms of PoP pairs generating it, and ii) are obtained with the assumption that the variables in question are continuous random variables. To overcome these issues, we propose CUR decomposition for decomposition of traffic matrices. Experimental results shows that basis vectors obtained using CUR decomposition i) are temporally more stable, ii)are 100% interpretable in terms of PoP pairs generating it, and iii) provides an improved classification of temporal patterns into periodic, spikes and noise pattern.

1. Introduction

Traffic matrix, which is an abstract representation of traffic volume flowing between set of point of presence (PoP) pairs, is considered as a more direct and fundamental input to network-wide applications including network tomography(Zhang et al., 2003; Soule et al., 2004; Papagiannaki et al., 2004; Soule et al., 2005) for accurately estimating traffic matrix, network anomography(Zhang et al., 2005; Lakhina et al., 2004a; Huang et al., 2006) for inferring anomalies in the network traffic, compressive sensing (Zhang et al., 2009) for estimating the missing entries in the traffic matrix. These applications require a deeper understanding of the components and structure of the traffic matrix, which falls in the realm of structural analysis of traffic matrix.

However, the high dimensional multivariate structure of traffic matrix is a prime source of difficulty in the structural analysis of traffic matrices. In practice, when presented with the need to analyze a high-dimensional structure, a commonly-employed and powerful approach is to seek an alternate lower-dimensional approximation to the structure that preserves its important properties. Matrix decomposition techniques aim to achieve a lower-dimensional approximation by decomposing the traffic matrix into a fewer number of basis vectors. Decomposition techniques have been extensively utilized for addressing the network-wide applications listed above. Most prevalent of the decomposition methods being singular valued decomposition (SVD). SVD is the backbone for various matrix decomposition methods such as principle component analysis (PCA), non-negative matrix factorization (NMF) and co-clustering methods. PCA is extensively used in all the above-mentioned applications.

Amongst all, PCA was first exercised by Lakhina et al.(Lakhina et al., 2004b) for decomposition of traffic matrix as a product of three orthogonal matrices using SVD as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{S}$ is a diagonal matrix containing singular values, and the columns of $\mathbf{U}$ and $\mathbf{V}$ are basis vectors, more specifically, left and right singular vectors respectively, arranged in the order of decreasing significance. They demonstrated that traffic matrix is inherently low dimensional, i.e., their structure can be well captured using remarkably few (generally between 5 and 10) significant basis vectors. They have also introduced the notion of eigenflows, to refer left singular vectors, which is a time series that captures a particular source of a temporal pattern (a "structure") in the traffic matrix. Traffic flow across a PoP pair (say P1), also termed as an origin-destination (OD) flow corresponding to P1, can be expressed as a weighted sum of these eigenflows, which fall into three natural classes: periodic, spikes and noise. They have also examined the temporal stability of basis vectors by demonstrating that the first week's basis vectors appear to remain good choices for forming a low-dimensional representation

[*]Equal contribution [1]Department of CSE, IIT Guwahati, Assam, India.. Correspondence to: Awnish Kumar <awnish@iitg.ernet.in>, Vijaya V. Saradhi <saradhi@iitg.ernet.in>, T. Venkatesh <t.venkat@iitg.ernet.in>.

ICML 2017 Time Series Workshop, Sydney, Australia, 2017.

of the subsequent consecutive week.

Later, Wang et al.(Wang et al., 2012) demonstrated the inefficiency of eigenflow classification approach for structural analysis because of the presence of indeterminate (which fall into more than one class) and nondeterminate (which do not fall in any of the classes) temporal patterns. They have utilized a robust variant of PCA for structural analysis of traffic matrix by decomposing it as a sum of three submatrices, with each corresponding to three classes as discussed in (Lakhina et al., 2004b). Despite the simplicity, PCA-based approach suffers from two fundamental limitations. First, interpretability of obtained basis vectors. The objective of PCA is to find directions along which maximum variance is achieved by projecting the data. The resulting directions, namely eigenflows, do not preserve meaning in terms of original dimensions (PoP pairs) as they stand for a weighted linear combination of the OD flows. It has been argued in the literature that recovering original input space from the projected space is an inherently challenging problem(Ringberg et al., 2007). This is problematic when one is interested in obtaining insights from the output of matrix decomposition. Consider, for example, the structural analysis of traffic matrix using eigenflows. Unfortunately, being linear combinations of up to all the OD flows, these eigenflows (basis vectors) are notoriously difficult to interpret in terms of the original PoP pairs generating the eigenflow. Second, data associated with every variable is assumed to be continuous and randomly drawn from the multivariate normal distribution. Sample PCA too follows this assumption. Variables in the traffic matrix, namely PoP pairs, are discrete in nature. This assumption has a direct impact on the covariance matrix computation which is subject to matrix decomposition. Violation of this fundamental assumption results in excessive skewness and kurtosis of the obtained matrix factors (Kolenikov et al., 2004).

In general, the basis vectors obtained after decomposition of a traffic matrix (say **X01**) should provide a low rank approximation of i) the current week traffic matrix (**X01**), as well as ii) subsequent consecutive week traffic matrices. This can be achieved through temporal stability of basis vectors, which can be beneficial for online applications. The choice of the decomposition method does influence the application outcome.

In order to alleviate the identified limitation of existing decomposition methods and meet the desired objectives of traffic matrix decomposition, we propose to use **CUR decomposition**, which is a low-rank matrix decomposition technique with basis vectors consisting of a small number of actual columns and actual rows of the traffic matrix. Because they are constructed from actual data elements, the basis vectors are 100% interpretable by the practitioners of the field from which the data are drawn (to the extent that the original data are). By preferentially choosing columns/rows as basis vectors that exert a disproportionately large influence on the best low-rank approximation, CUR decomposition ensures that the error in reconstruction of original traffic matrix is always less as compared to SVD. In addition to this, there are no distributional assumptions involved in the selection process.

1.1. Main Contributions

In the present work, we employ CUR decomposition for structural analysis of traffic matrix using the approach of eigenflow classification(Lakhina et al., 2004b) because of its simplicity. Using CUR decomposition, we are able to alleviate the limitations of eigenflow flow classification approach identified in(Wang et al., 2012) as well as limitations of PCA-based approaches, and obtain: i) **temporally stable basis vectors** in terms of low error incurred in low rank approximation of current as well as subsequent consecutive week traffic matrix, ii) **100% Interpretability of basis vectors** in terms of PoP pairs generating it, and iii) **Improved classification of temporal patterns** in terms of reduction in the number of unclassified temporal patterns.

1.2. Roadmap of the Report

This subsection presents the roadmap of the paper. Section 2 discusses the technique proposed for decomposition of traffic matrices, i.e., CUR decomposition. Section 3 presents the details of dataset collected from Abilene network. In the present work, we demonstrate the effectiveness of CUR decomposition for structural analysis of traffic matrices. The results obtained after evaluation are presented in section 4. Finally, we provide the summary in section 5

2. Proposed Method : CUR decomposition

CUR decomposition is a low-rank matrix decomposition technique that is explicitly expressed in terms of a small number of *original* columns and/or *original* rows of the data (traffic) matrix. Variants of CUR algorithm(Drineas & Kannan, 2003; Drineas et al., 2006; 2008) has been proposed in the literature which compete on the reconstruction error bounds, with the most improved one presented in (Mahoney & Drineas, 2009). According to (Mahoney & Drineas, 2009), given an $m \times n$ matrix **X** and a rank parameter 'k', decompose **X** as a product of 3 matrices, **C**, **U**, and **R**, where **C** consists of a small number c (= O(k log k/ϵ^2)) of actual columns of **X**, **R** consists of a small number r (= O(k log k/ϵ^2)) of actual rows of **X**, and **U** is a small carefully constructed matrix that guarantees the error bound of the following form

$$\|\mathbf{X} - \mathbf{CUR}\|_{\mathbf{F}}^{\mathbf{2}} \leq (\mathbf{2} + \epsilon)\|\mathbf{X} - \mathbf{X_k}\|_{\mathbf{F}}^{\mathbf{2}} \quad (1)$$

where $\mathbf{X_k}$ is best rank-k approximation of **X** obtained using SVD, ϵ is the error parameter, and $\|.\|_F$ denote the frobenius norm.

Several things should be noted about this definition. **First**, to construct **C** (similarly **R**), an importance score is computed, which depends on the euclidean norm of rows of top 'k' right singulars obtained after SVD of **X**. Using this importance score as sampling probability distribution, CUR decomposition randomly sample 'c' columns of **X**, which exert a disproportionately large influence on the best low-rank fit of **X**. In fact, the product **C*U*R** will be nearly as good as the best low-rank approximation to **X** that is obtained by truncating the SVD. **Second**, the construction of **C** and **R** involves truncated SVD, i.e., computation of top k basis vectors, which has a time complexity of O(mnk), which is an improvement over full SVD with the complexity of O(min{$\mathrm{m^2n}$, $\mathrm{mn^2}$}). **Third**, the matrix **C** and **R** can be used in place of left and right singular vectors, but since they consist of actual data elements they will be interpretable in terms of the original columns and rows of **X**. **Fourth**, a CUR approximation approximately expresses all of the columns of **X** in terms of a linear combination of a small number 'c' of original columns of **X**. This will provide an aid to structural analysis of traffic matrix in terms of interpretable basis vectors. **Fifth**, CUR matrix decomposition has structural properties that are auspicious for its use as a tool in the analysis of large data sets. For example, if the data matrix **X** is large and sparse but well-approximated by a low-rank matrix, then **C** and **R** (consisting of actual columns and rows) are sparse, whereas the matrices consisting of the top left and right singular vectors will not, in general, be sparse. **Sixth**, there are no distributional assumptions involved in the selection of **C** and **R**.

3. Dataset

Traffic matrices utilized in the present work is collected from Abilene network(abi). In the experimentation traffic matrices numbered from **X10** to **X24** of size (2016×121) is utilized because of no missing periods, with each traffic matrix containing traffic flow volume measured over a week.

4. Application: Structural Analysis

Structural analysis aims to decompose an OD flow into its constituent temporal patterns. CUR decomposition of traffic matrix expresses each OD flow as a weighted linear combination of few significant basis vectors (i.e., columns of **C**), which corresponds to temporal patterns of traffic matrix. Before presenting the results of structural analysis, we demonstrate the temporal stability of basis vectors in obtaining a low rank approximation of current as well as subsequent consecutive week traffic matrix.

4.1. Temporal Stability of Basis Vectors

The question we are concerned with in this section is whether the basis vectors obtained after decomposition of a given traffic matrix is useful for analyzing subsequent week traffic matrix that was not part of the input. In general, we envision applications that may benefit from using matrix decomposition in an online manner as follows. Given traffic matrix observed over some time period [t0, t1), obtain the basis vectors. Subsequently, at some time t2 > t1, use previously derived basis vectors to decompose a new set of traffic matrix into temporal patterns. Does the subsequent decomposition still have relatively low effective dimensionality?

To answer this question, we proceed as follows. One way to assess whether a traffic matrix has low effective dimension is to measure the error resulting from approximating the matrix using a small number of basis vectors. Following the approach in(Lakhina et al., 2004b), given rank parameter 2, and two consecutive weeks of traffic matrices **X10** and **X11**, we obtain basis vectors of **X10** and approximate each OD flow of **X10** and **X11** using CUR, PCA and NMF, yielding $\mathbf{X10'}$ and $\mathbf{X11'}$. The error incurred in both approximation is quantified using sum of squared error (SSE) per OD flow, which is given as $SSE1_j = \|X10(:,j) - X10'(:,j)\|$ and $SSE2_j = \|X11(:,j) - X11'(:,j)\|$, where $SSE1_j$ and $SSE1_j$ denote the sum of squared error incurred in approximating j-th OD flow of **X10** and **X11** respectively.

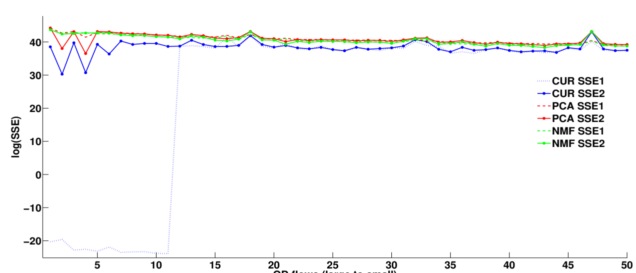

Figure 1. Exploring the temporal stability of basis vectors obtained using CUR and PCA

Figure 1 shows the plot of logarithmic of SSE1 and SSE2, obtained using CUR decomposition, PCA, and NMF of top 50 OD flows ordered by decreasing mean rate from left to right on the x-axis. It can be observed from the figure that the error incurred in approximation of top 11 OD flows of **X10** using CUR decomposition is significantly small (close to 0) as compared to PCA and NMF. In addition to this, the basis vectors of **X10** obtained using CUR decomposition provides a better low rank approximation of **X11** in terms of low SSE per OD flow as compared to PCA and NMF. This suggests that the basis vectors of CUR decomposition are temporally more stable than PCA and NMF, hence should be preferred for online applications of traffic matrix.

4.2. Interpretability of Basis Vectors

One of main reasons of preferring CUR decomposition over PCA-based decomposition methods lies in the interpretability of eigenflows. Using CUR decomposition, we propose to decompose an OD flow into its constituent temporal patterns, that are interpretable in terms of PoP pairs generating it. That is,

$$\mathbf{x}_j = \sum_{i=1}^{c} \mathbf{c}_i (ur)_{ij} \quad (2)$$

where $\mathbf{x}_j$ is the time-series of the j-th OD flow and ur_{ij} is the (i,j)-th element of the matrix $(\mathbf{U} * \mathbf{R})$. Equation 2 makes clear that j-th OD flow $\mathbf{x}_j$ is in turn a linear combination of a few significant OD flows (i.e., columns of **C**), with associated weights given by j-th column of the matrix (**U*R**). These significant OD flows, i.e., columns of **C**, stand for the temporal patterns (or "constituents" or "structures") of traffic matrix, which have an edge over the ones identified in prior works(Lakhina et al., 2004b; Wang et al., 2012) in terms of interpretability. Experimentation shows that top two significant temporal patterns ("structures") of **X10** captured using PCA and CUR decomposition show spikes behaviour during Wednesday, Thursday and Friday. However, with CUR decomposition, we can say that these spikes pattern are generated across Chicago-Los Angeles (C-L) and Los Angeles-Chicago (L-C) pairs. This type of interpretability about temporal patterns is inherently absent in PCA-based approaches and NMF.

4.3. Classification of Temporal Patterns

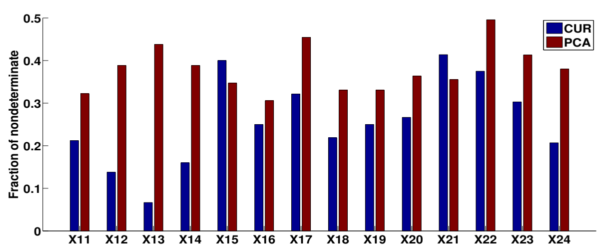

Figure 2. Bar plot of fraction of nondeterminate temporal patterns captured using PCA (red) and CUR decomposition (blue).

Later, to obtain significant insight into the whole-network properties of data traffic, these temporal patterns are classified into three classes as defined in (Lakhina et al., 2004b; Wang et al., 2012) using a more apt procedure as compared to(Lakhina et al., 2004b). The common periodic trend is captured using autocorrelation if the traffic flow exhibits periodicity at 12 or 24 hours. Short-lived burst is captured if at any time the traffic flow volume exceeds 3- standard deviation from the mean. Chi-square goodness of fit test is used to decide whether a flow is a noise or not. This classification approach has helped in identification of more temporal pattern which remained unclassified using classification approach in (Lakhina et al., 2004b). The take-away point of CUR decomposition is that we have to classify only a small number c of temporal patterns for structural analysis of traffic matrices as compared to classification of a large number n of eigenflows using PCA and NMF, where c≪n. This reduced number of temporal patterns will ease the burden during classification into three classes. Experimentation shows a significant reduction in computational time in case of CUR decomposition (4.308 seconds) as compared to PCA (13.382 seconds) and NMF (18.98 seconds) for **X10** of size 2016×121, using Matlab on a 3.20 GHz Windows machine.

The temporal patterns of CUR decomposition, i.e., significant OD flows, arises from the superposition of periodic, spike and noise component(Lakhina et al., 2004b); thus can fall into more than one class. Hence, the presence of indeterminate temporal patterns is justified to a greater extent for structural analysis of traffic matrices. Figure 2 shows the bar plot of the fraction of nondeterminate temporal patterns captured using CUR decomposition and PCA for 14 traffic matrices (**X11** to **X24**). CUR decomposition achieves a drop in the number of unclassified temporal patterns as compared to PCA for 12 out of 14 matrices. NMF, on the other hand, lags behind CUR decomposition in terms of percentage of indeterminate temporal patterns captured. The difference is as low as 0.05 for **X16** and as high as 0.37 for **X13**. This improvement in classification of temporal patterns is achieved because of adopting apt and heuristic mechanism as used in the present work.

5. Summary

In the present work, we perform CUR decomposition of traffic matrices for structural analysis of traffic matrices, which provides basis vectors in terms of original PoP pairs and time intervals, and is computationally less expensive. These basis vectors provide a more accurate low rank approximation of current as well as subsequent consecutive week traffic matrix. Using CUR decomposition, along with identification of significant temporal patterns, we are also able to identify PoP pairs generating it. The classification accuracy is also improved in terms of reduced number of unclassified temporal patterns.

References

Drineas, Petros and Kannan, Ravi. Pass efficient algorithms for approximating large matrices. In *SODA*, volume 3, pp. 223–232, 2003.

Drineas, Petros, Kannan, Ravi, and Mahoney, Michael W. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. *SIAM Journal on Computing*, 36(1):184–206, 2006.

Drineas, Petros, Mahoney, Michael W, and Muthukrishnan, S. Relative-error CUR matrix decompositions. *SIAM Journal on Matrix Analysis and Applications*, 30(2): 844–881, 2008.

Huang, Ling, Jordan, Michael I., Joseph, Anthony, Garofalakis, Minos, and Taft, Nina. In-network PCA and anomaly detection. In *NIPS*, pp. 617–624. MIT Press, 2006.

Kolenikov, Stanislav, Angeles, Gustavo, et al. The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices. *Chapel Hill: Carolina Population Center, University of North Carolina*, pp. 1–59, 2004.

Lakhina, Anukool, Crovella, Mark, and Diot, Christophe. Diagnosing network-wide traffic anomalies. In *ACM SIGCOMM Computer Communication Review*, volume 34, pp. 219–230. ACM, 2004a.

Lakhina, Anukool, Papagiannaki, Konstantina, Crovella, Mark, Diot, Christophe, Kolaczyk, Eric D, and Taft, Nina. Structural analysis of network traffic flows. In *ACM SIGMETRICS Performance evaluation review*, volume 32, pp. 61–72. ACM, 2004b.

Mahoney, Michael W and Drineas, Petros. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

Papagiannaki, Konstantina, Taft, Nina, and Lakhina, Anukool. A distributed approach to measure IP traffic matrices. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, IMC ’04, pp. 161–174, New York, NY, USA, 2004. ACM. ISBN 1-58113-821-0.

Ringberg, Haakon, Soule, Augustin, Rexford, Jennifer, and Diot, Christophe. Sensitivity of PCA for traffic anomaly detection. *ACM SIGMETRICS Performance Evaluation Review*, 35(1):109–120, 2007.

Soule, Augustin, Nucci, Antonio, Cruz, Rene, Leonardi, Emilio, and Taft, Nina. How to identify and estimate the largest traffic matrix elements in a dynamic environment. In *IN ACM SIGMETRICS*, 2004.

Soule, Augustin, Lakhina, Anukool, Taft, Nina, Papagiannaki, Konstantina, Salamatian, Kave, Nucci, Antonio, Crovella, Mark, and Diot, Christophe. Traffic matrices: balancing measurements, inference and modeling. In *In Proceedings of the ACM SIGMETRICS*, pp. 362–373. Press, 2005.

Wang, Zhe, Hu, Kai, Xu, Ke, Yin, Baolin, and Dong, Xiaowen. Structural analysis of network traffic matrix via relaxed principal component pursuit. *Comput. Netw.*, 56 (7):2049–2067, May 2012. ISSN 1389-1286.

Zhang, Yin, Roughan, Matthew, Lund, Carsten, and Donoho, David. An information-theoretic approach to traffic matrix estimation. In *In Proc. ACM SIGCOMM*, pp. 301–312, 2003.

Zhang, Yin, Ge, Zihui, Greenberg, Albert, and Roughan, Matthew. Network anomography. In *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, IMC ’05, pp. 30–30, Berkeley, CA, USA, 2005. USENIX Association. URL http://dl.acm.org/citation.cfm?id=1251086.1251116.

Zhang, Yin, Roughan, Matthew, Willinger, Walter, and Qiu, Lili. Spatio-temporal compressive sensing and internet traffic matrices. In *ACM SIGCOMM Computer Communication Review*, volume 39, pp. 267–278. ACM, 2009.