



# Seasonal Disorder in Urban Traffic Patterns: A Low Rank Analysis

Vaibhav Karve<sup>1</sup> · Derrek Yager<sup>2</sup> · Marzieh Abolhelm<sup>3</sup> · Daniel B. Work<sup>4</sup> · Richard B. Sowers<sup>5</sup> 

Received: 11 June 2020 / Revised: 7 December 2020 / Accepted: 4 January 2021  
© The Author(s) 2021, corrected publication 2021

## Abstract

This article proposes several advances to sparse nonnegative matrix factorization (SNMF) as a way to identify large-scale patterns in urban traffic data. The input to our model is traffic counts organized by time and location. Nonnegative matrix factorization additively decomposes this information, organized as a matrix, into a linear sum of temporal signatures. Penalty terms encourage this factorization to concentrate on only a few temporal signatures, with weights which are not too large. Our interest here is to quantify and compare the regularity of traffic behavior, particularly across different broad temporal windows. In addition to the rank and error, we adapt a measure introduced by Hoyer to quantify sparsity in the representation. Combining these, we construct several curves which quantify error as a function of rank (the number of possible signatures) and sparsity; as rank goes up and sparsity goes down, the approximation can be better and the error should decrease. Plots of several such curves corresponding to different time windows leads to a way to compare disorder/order at different time scale windows. In this paper, we apply our algorithms and procedures to study a taxi traffic dataset from New York City. In this dataset, we find weekly periodicity in the signatures, which allows us an extra framework for identifying outliers as significant deviations from weekly medians. We then apply our seasonal disorder analysis to the New York City traffic data and seasonal (spring, summer, winter, fall) time windows. We do find seasonal differences in traffic order.

**Keywords** Traffic · Normalization · Sparse nonnegative matrix Factorization

---

The authors acknowledge the Program for Interdisciplinary and Industrial Internships at Illinois (PI4) and the Illinois Geometry Laboratory (IGL). The many IGL students who have made invaluable contributions to this work are: Raghav Bakshi, James Kerns, Xinyi Li, Xinyu Liu, Yicheng Pu, Gabriel Shindnes, Haozhe Wang, Jing Wang, Ziyang Wang, Yu Wu, Zeyu Wu, Bin Xu, and Dajun Xu. The authors would also like to thank the Siebel Energy Institute for its support of this work. This material is based upon work supported by the National Science Foundation under Grant Numbers CMMI 1727785 and DMS 1345032. This work was also supported by a grant from the Siebel Energy Institute. The code for this work is at <https://gitlab.engr.illinois.edu/TrafficPatterns/CSNMF.git>

Sandia National Laboratories is a multimission laboratory operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration. Sandia Labs has major research and development responsibilities in nuclear deterrence, global security, defense, energy technologies and economic competitiveness, with main facilities in Albuquerque, New Mexico, and Livermore, California.

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

---

Extended author information available on the last page of the article

## Introduction

### Motivation

Traffic management is one of the persistent challenges of the modern industrialized world. It simultaneously reflects both a critical infrastructural necessity and a problem involving a wide range of scales and interactions. Over the past 2 decades, floating-car data (for example, from GPS-equipped taxis and other vehicles) has become an important source of real-time and large-scale city traffic information Deri and Moura (2015), Donovan and Work (2015). These data streams can be used to manage traffic signals, influence equilibrium traffic states Ban et al. (2011), Zheng and Liu (2017), Herman and Prigogine (1979), Mahmassani et al. (1984), Geroliminis and Daganzo (2008), Krichene et al. (2016), Deri and Moura (2015), Zhu et al. (2016), Zhan et al. (2014), Guan et al. (2016), Alonso-Mora et al. (2017), Ferreira et al. (2013), and optimally route traffic.

Our interest here is in part to quantify and compare regularity of traffic behavior, particularly across different broad temporal windows. Traffic behavior changes as

seasons change, and we would like to build a framework for comparing the regularity of these behaviors. With reliable methods which allow us to identify broad trends in traffic behavior, we can understand responses to urban planning decisions and singular events, and hopefully help cities be more responsive.

## Problem Statement and Contribution

We would like to find broad citywide patterns which can be used to describe macroscopic behavior of traffic counts; that is, the number of cars on a given link (directed edge) in a road network at a certain time. Two-way roads are, of course, represented as two links. The challenge is to find low-dimensional descriptions of this potentially large dataset, and to quantify disorder as the inability to fully describe traffic behavior in this low-dimensional way.

Our goal here is to decompose traffic counts on different links into different behavioral signatures. The weighted summation of such signatures will represent a low-rank approximation of the traffic behavior. In this additive decomposition method, a given behavioral signature can, for example, represent heavy traffic volumes during the morning rush hour, medium traffic for the rest of the workday, and light traffic in the evening. However, as a signature spans the entire year, this example suggests a segment of the overall pattern which regularly repeats (every workday). Both the behavioral signatures and the weights of links on the signatures should be non-negative; each signature describes a pattern of a non-negative number of vehicles on a link (e.g., activity near a public gathering place) and there are a non-negative number of people behaving according to that pattern; the mathematics of non-negative matrix factorization will play a key role.

We are interested in the *error* of this approximation; a large error, roughly, corresponds to disorder in traffic behavior. Our low-rank approximation depends on several parameters. By looking at this error across several seasonal windows and constructing a curve which captures the error of this approximation as it depends on these parameters, we can compare disorder in traffic patterns.

## Dataset

Our method uses traffic counts on links, broken up into time increments. In this paper, we will apply our calculations to New York City traffic data given in Donovan et al. (2016). See Section 4 for more information about the dataset (and its limitations). An added component of our analysis is that

it fills in some missing values via matrix completion; see [ZWFM], Xu et al. (2012).

## Related Works

Non-negative Matrix (and tensor) Factorization (NMF) has already been used for urban and network traffic analysis Ahmadi et al. (2015), Chen et al. (2019), Yufei and Fabien (2011), Han and Moutarde (2013), Han and Moutarde (2016), Hofleitner et al. (2012), Liu et al. (2017), Ma et al. (2018), Sun and Axhausen (2016), Xu et al. (2015); see also Lv et al. (2015). Furthermore, matrix factorization has also been used in studying train Gong et al. (2018), Ito et al. (2017), bicycle Cazabet et al. (2018), and risk Lee et al. (2016) data. Anomaly detection using related methods can be found in Djenouri et al. (2018), Zhang et al. (2016), Guo et al. (2015), Li et al. (2015), Wang et al. (2019).

More generally, NMF has been applied to a wide range of problems, like text data mining Chagoyen et al. (2006), Pauca et al. (2004), gene expression Brunet et al. (2004), Carmona-Saez et al. (2006), Gao and Church (2005), Kim and Tidor (2003), Maher et al. (2006), micro-array comparative genomics hybridization Carrasco et al. (2006), functional characterization of gene lists Pehkonen et al. (2005) and facial images Li et al. (2001), Hoyer (2004).

A focus of our work is sparsity in the matrix factorization. Various extensions of NMF have been made to impose sparsity, either on both factors, as in Dueck et al. (2005), Hoyer (2004), Pascual-Montano et al. (2006), Pauca et al. (2006), Kim and Park (2007), or on only one Gao and Church (2005), Pauca et al. (2004). There are furthermore ways of *measuring* sparsity Hoyer (2004) which are different than the algorithmic ways to encourage it; see Sect. 3.1.

Another somewhat novel focus of our work is using low rank factorization to quantify *disorder* (and in particular compare seasonal disorder); see Sect. 5. We carry out a range of low rank factorizations, finding that there are robust seasonal biases in the error.

Our work follows the ideas of Lee and Seung (2001) and Kim and Park (2007). Matrix factorization can be viewed as an approach to reduce the dimensionality or compress Asif et al. (2013) traffic data. Traffic data dimensionality techniques span more classical techniques building on principal component analysis Li et al. (2007), Yang and Qian (2019), Li et al. (2015), to more recent developments using variational autoencoders Boquet et al. (2020). Exploiting spatio-temporal patterns can both support data reduction and also prediction of future traffic states on the network, Yang and Qian (2019). For a recent review of approaches to exploit structure in traffic data for

prediction, we refer to the reviews Ermagun and Levinson (2018), Nagy and Simon (2018), Pavlyuk (2019).

### Outline

In Sect. 2, we set up the theory and background for the algorithm we use. In Sect. 3, we describe metrics and tools we use for analyzing error and sparsity associated with the approximation. In Sect. 4, we apply our analysis to the dataset of New York City traffic described in Sect. 1.3. We further analyze the error of our algorithm in relation to the New York City dataset in Sect. 5. We give some conclusions in Sect. 6. The appendix reviews the mathematics behind the iterative algorithm we develop in Sect. 2.3.

The supporting source code for this work is published at [KYA+].

## Setup and Theory

### Traffic Counts

We start by organizing traffic counts into a matrix  $D$ , where  $D_{t,\ell}$  is the traffic count on link  $\ell$  during time period  $t$ . The time index ranges through  $T \stackrel{\text{def}}{=} \{1, 2 \dots T\}$  and the link index ranges through a finite set  $\mathcal{L} \stackrel{\text{def}}{=} \{1, 2 \dots L\}$  of links. Our interest is seasonal fluctuations, so we assume that  $T$  represents a large time horizon. If the traffic count on link  $\ell$  at time  $t$  is unknown (i.e., missing), we set  $D_{t,\ell} \stackrel{\text{def}}{=} \text{NaN}$  and define

$$\mathcal{I} \stackrel{\text{def}}{=} \{(t, \ell) \in T \times \mathcal{L} : D_{t,\ell} \neq \text{NaN}\} \tag{1}$$

as the set of indices for which we have traffic estimates.

### Non-Negative Matrix Factorization

We want to decompose  $D$  as

$$\overbrace{D}^{T \times L} \approx \overbrace{W}^{T \times N} \overbrace{H}^{N \times L}, \tag{2}$$

where  $N$  is the rank of the factorization, with  $N \leq \min\{T, L\}$ , and where the entries of  $W$  and  $H$  are nonnegative, and where  $WH$  denotes the standard matrix product.

To quantify the error in (2), let us first define

$$([D]_{\mathcal{I}})_{t,\ell} \stackrel{\text{def}}{=} \begin{cases} D_{t,\ell} & \text{if } (t, \ell) \in \mathcal{I} \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

with  $\mathcal{I}$  as in (1). In other words, we replace NaNs with zeroes. Then, define

$$\mathcal{E}_o(W, H) \stackrel{\text{def}}{=} \|[D - WH]_{\mathcal{I}}\|_F^2 = \sum_{(t,\ell) \in \mathcal{I}} (D - WH)_{t,\ell}^2. \tag{4}$$

with  $\|\cdot\|_F$  being the Frobenius norm.

### Sparse Nonnegative Matrix Factorization

The rank  $N$  denotes the size of the “universe” of possible signatures. We can think of the columns of  $W$  as temporal signatures and the rows of  $H$  as weights. Within this collection of temporal signatures, we want to represent each column of  $D$  (i.e., the traffic count on each link) using as few of these signatures as possible.

For any matrix  $A \in \mathbb{R}^{M \times N}$ , let  $A_{\cdot,n}$  denote the  $n$ -th column of  $A$  and let  $A_{m,\cdot}$  denote the  $m$ -th row of  $A$ . Sparse Nonnegative Matrix Factorization (SNMF), as seen in Kim and Park (2008) (cf. Hoyer (2002)), fixes positive parameters  $\beta$  and  $\eta$  and minimizes

$$\mathcal{E}_{\beta,\eta}(W, H) \stackrel{\text{def}}{=} \mathcal{E}_o(W, H) + \beta \sum_{\ell \in \mathcal{L}} \|H_{\cdot,\ell}\|_1^2 + \eta \|W\|_F^2 \tag{5}$$

Sparsity in the columns of  $H$  is encouraged by the penalty term involving  $\beta$  (which involves an  $L_1$  LASSO-type penalty), while the term involving  $\eta$  is used to prevent the values in  $W$  from being too large. The choice of these parameters is problem-specific.

### Normalization

To provide a common reference for comparing behavior in different columns (i.e., links), we finally require that the columns of  $W$  sum to 1 (i.e., have  $L_1$ -norm of 1). This gives us a relative traffic count; allowing us to better understand how the traffic count in the signature is temporally broken down (i.e., by time, week, and season). It also explicitly resolves ambiguity in the multiplicative decomposition  $D \approx WH$ .

Due to this normalization, our  $H$  contains the scale-factor of  $D$ . In other words, scaling  $D$  by a constant factor leaves  $W$  unchanged but scales up  $H$  by that same factor. The choice for  $N$  and  $\beta$  therefore are independent of the scale of entries in  $D$ .

### Algorithm

Our iterative algorithm is based on Lee and Seung (2001) with sparsity modifications as laid out in Kim and Park (2008). Details of the calculation are in the appendix.

For positive integers  $R$  (number of rows) and  $C$  (number of columns), let  $\mathbf{1}_{R \times C}$  be the  $(R \times C)$ -matrix whose entries are all 1. For matrices  $A$  and  $B$  in  $\mathbb{R}^{R \times C}$ , let

$$(A \odot B)_{i,j} \stackrel{\text{def}}{=} A_{i,j} B_{i,j}$$

$$(A \oslash B)_{i,j} \stackrel{\text{def}}{=} \frac{A_{i,j}}{B_{i,j}} \quad \text{for } 1 \leq i \leq R, 1 \leq j \leq C$$

denote Hadamard (i.e., elementwise) multiplication and division.

Starting with any full-rank  $(W, H) \in \mathbb{R}_+^{T \times N} \times \mathbb{R}_+^{N \times L}$ , our iterative update is given by a sequence of recursive update rules:

$$\begin{aligned} W &\leftarrow W \odot ([D]_{\mathcal{I}} H^T) \oslash ([WH]_{\mathcal{I}} H^T + \eta W) \\ H &\leftarrow H \odot (W^T [D]_{\mathcal{I}}) \oslash (W^T [WH]_{\mathcal{I}} + \beta \mathbf{1}_{N \times N} H) \end{aligned} \tag{6}$$

followed by

$$(W_{\cdot, n}, H_{n, \cdot}) \leftarrow \left( W_{\cdot, n} / \|W_{\cdot, n}\|_1, H_{n, \cdot} \times \|W_{\cdot, n}^*\|_1 \right) \tag{7}$$

### Algorithm Termination

There are three natural indicators that the algorithm has converged:

- $\|W^{(m+1)} - W^{(m)}\|_F \approx 0$
- $\|H^{(m+1)} - H^{(m)}\|_F \approx 0$
- $|\mathcal{E}_{\beta, \eta}(W^{(m+1)}, H^{(m+1)}) - \mathcal{E}_{\beta, \eta}(W^{(m)}, H^{(m)})| \approx 0$ .

We use a combination of the first two, that is, we stop our algorithm when

$$\frac{\|W^{(m+1)} - W^{(m)}\|_F}{\|W^{(m)}\|_F} + \frac{\|H^{(m+1)} - H^{(m)}\|_F}{\|H_n\|_F} \leq \tau,$$

for a fixed positive threshold  $\tau$ .

### Error Analysis

#### Sparsity

As we increase the parameter  $\beta$  of (5), which increases the size of the LASSO-type penalty on the columns of  $H$ , we expect  $H$  to become more *sparse*; i.e. to have more small entries. We can measure this via the calculations of Hoyer (2004), which are based on the relationship between the  $L_1$  and  $L_2$  norms. For nonzero  $x \in \mathbb{R}^N$ , define

$$\text{Sparsity}(x) = \frac{\sqrt{N} - \|x\|_1 / \|x\|_2}{\sqrt{N} - 1}. \tag{8}$$

Then,  $0 \leq \text{Sparsity}(x) \leq 1$ , with  $\text{Sparsity}(x) = 1$  if and only if all except one of the entries of  $x$  is nonzero, and  $\text{Sparsity}(x) = 0$  if and only if all entries of  $x$  take

on a common (nonzero) value. We also note that  $\text{Sparsity}(\alpha x) = \text{Sparsity}(x)$  for nonzero  $\alpha \in \mathbb{R}$ ; i.e., Sparsity is scale-invariant.

For  $H \in \mathbb{R}^{N \times L}$  having no nonzero columns, let us similarly define

$$\text{Sparsity}(H) = \frac{1}{L} \sum_{\ell \in \mathcal{L}} \text{Sparsity}(H_{\cdot, \ell}); \tag{9}$$

$\text{Sparsity}(H)$  is the average of the  $\text{Sparsity}(H_{\cdot, \ell})$  over all  $\ell$ . We note that this measure of sparsity is unaffected by the rescaling of (7).

#### Tradeoff

Once we have found a reasonable value for  $\eta$ , a range of values of the sparsity penalty  $\beta$  and the rank  $N$  (see Sect. 4.1), we can carry out a perturbative analysis of error. For a given low-rank decomposition  $(W, H)$ , we can construct the error matrix

$$D - WH. \tag{10}$$

We are interested in the seasonal fluctuations of the error (10) as an indication of how “ordered” traffic behavior is at different times. A small error would mean that traffic behavior can, in fact, be additively decomposed into different signatures.

We first fix a reference  $\eta$  (see Sect. 4.1). Let’s also fix a set  $\mathcal{N}$  of possible values of the rank parameter  $N$ , and a set  $\mathcal{B}$  of possible values of the penalty parameter  $\beta$ . For each  $(N, \beta) \in \mathcal{N} \times \mathcal{B}$  (i.e., a grid of values for  $N$  and  $\beta$ ) let  $(W^*(N, \beta), H^*(N, \beta)) \in \mathbb{R}_+^{T \times N} \times \mathbb{R}_+^{N \times L}$  be the result of the iterative scheme of Sect. 2.5, with this  $\beta$  and  $N$ , terminated according to the criterion laid out in Sect. 2.6. If in some cases, the algorithm fails to converge, then we say that  $(W^*(N, \beta), H^*(N, \beta))$  is undefined.

Fix  $\mathcal{T} \subset \mathcal{T}$ , and define

$$\mathcal{E}^*(N, \beta, \mathcal{T}) \stackrel{\text{def}}{=} \frac{\|(D - W^*(N, \beta)H^*(N, \beta))_{\mathcal{T} \times \mathcal{L}}\|_F}{\|(D)_{\mathcal{T} \times \mathcal{L}}\|_F}$$

$$\text{Sparsity}^*(N, \beta) \stackrel{\text{def}}{=} \text{Sparsity}(H^*(N, \beta))$$

for  $(N, \beta) \in \mathcal{N} \times \mathcal{B}$  such that  $(W^*(N, \beta), H^*(N, \beta))$  is defined. Here in the definition of  $\mathcal{E}^*$ ,  $(A)_{\mathcal{T} \times \mathcal{L}}$  denotes the submatrix corresponding to  $(A_{t, \ell} : (t, \ell) \in \mathcal{T} \times \mathcal{L})$  for any  $A \in \mathbb{R}^{\mathcal{T} \times \mathcal{L}}$ . For each  $N \in \mathcal{N}$ , define the sets

$$\begin{aligned} \mathcal{B}(N) &\stackrel{\text{def}}{=} \{\beta \in \mathcal{B} : (W^*(N, \beta), H^*(N, \beta)) \text{ is defined}\} \\ \mathcal{S}(N) &\stackrel{\text{def}}{=} \{\text{Sparsity}^*(N, \beta) : \beta \in \mathcal{B}(N)\} \end{aligned} \tag{11}$$

We want to compare  $\mathcal{E}^*$  for different subsets  $\mathcal{T}'$  of  $\mathcal{T}$  (e.g. different  $\mathcal{T}'$ ’s corresponding to different seasons). Informally, we want to compare  $\mathcal{E}^*(N, \beta, \mathcal{T}')$  and  $\mathcal{E}^*(N, \beta, \mathcal{T})$  for the

same values of  $N$  and  $\beta$ . A larger value of  $\mathcal{E}^*$  corresponds to more disorder on the original data. Note that our matrix factorization does not depend on the subset  $\mathcal{T}$ ; we rather are evaluating the error over different subsets of time. This allows us to use as much data as possible in the factorization, and resolve some of the challenges in dealing with sparse data. Informally, we are using a common collection of  $H$  weights, while selecting seasonal parts of the  $W$  signatures.

**Plot 1** (Monotonicity of sparsity). For each  $N$ , we can plot  $\{\beta \text{ vs. Sparsity}^*(N, \beta) \mid \beta \in \mathcal{B}(N)\}$ . (12)

For a fixed  $N \in \mathcal{N}$ , we expect that  $\text{Sparsity}^*$  should be increasing in  $\beta$  (see Fig. 10 on page 13). If so, we can uniquely reparameterize effects of  $\beta$  as effects of  $\text{Sparsity}^*$  of (9).

Next, we can understand how  $\text{Sparsity}^*$  and  $\mathcal{E}^*$  vary for each given  $N \in \mathcal{N}$ .

**Plot 2** (Dependence of  $\mathcal{E}^*$  on rank and  $\beta$ ) For each  $N \in \mathcal{N}$ , we can parametrically plot

$$\{\text{Sparsity}^*(N, \beta) \text{ vs. } \mathcal{E}^*(N, \beta, \mathcal{T}) \mid \beta \in \mathcal{B}(N)\}.$$

We expect this to be nondecreasing; more sparsity reflects more restrictions on the factors in the matrix product, leading to more error (see Figs. 11–14 on pages 13–15).

We finally can approximately plot error as a function of  $N$  for fixed sparsity values. Fix  $N \in \mathcal{N}$  and let  $\mathcal{S}_N : \mathcal{B} \rightarrow [0, 1]$  be the piecewise linear function with knots at (12).

**Plot 3** (Dependence of  $\mathcal{E}^*$  on rank and sparsity). If  $\mathcal{S}_N$  is nondecreasing for each  $N \in \mathcal{N}$ , and  $s \in (0, 1)$  is in the range of all of the  $\mathcal{S}_n$ 's (for  $n \in \mathcal{N}$ ), we can plot

$$\{s \text{ vs. } \mathcal{E}^*(N, \mathcal{S}_N^{-1}(s), \mathcal{T}) \mid N \in \mathcal{N}\}.$$

(See Figs. 15, 16, 17, 18 on pages 15–16).

We expect that as sparsity increases, so does the error.

### Analysis of New York City Data: Matrix Factorization

Let's apply our analysis to the data of Sect. 1.3 (i.e., Donovan et al. (2016)), which reverse-engineers estimates of traffic counts from origin-destination pairs for taxi trips. Our taxi dataset is an illustrative (and perhaps scaled) proxy for true traffic counts, but we recognize that it is potentially biased. Our methodology could readily be modified when

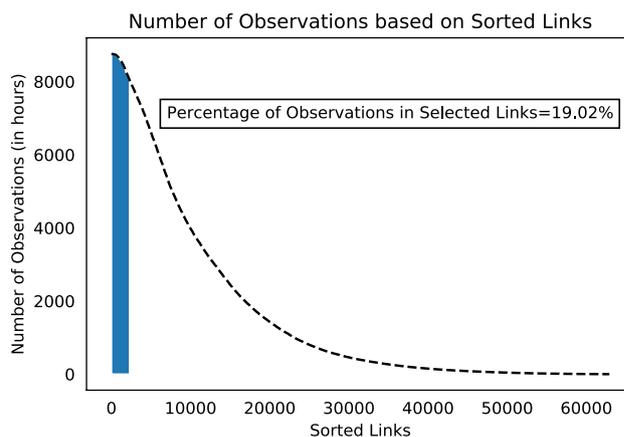


Fig. 1 Nonzero entries with respect to links sorted in decreasing order of traffic usage

Table 1 Statistic of link lengths (in meters)

Statistic	value for all 260, 855 links	value for 2, 302 links with < 30 missing hours
Mean	133.0	106.4
Median	95.6	80.0
Mode	79.3	78.5
Std.Dev	107.8	133.1
Minimum	2.8	40.0
Maximum	3937.1	2676.248

non-biased counts (possibly available from private companies or from dedicated traffic counting sensors installed by municipalities; see also [pem]) are available.

The dataset contains hourly traffic data in the interval [2011-01-01 00:00, 2012-01-01 00:00). (13)

We thus have

$$T = \underbrace{365}_{\text{days per year}} \times \underbrace{24}_{\text{hours per day}} = 8,760$$

time records. Arranging them in order, we get  $\mathcal{T}$  of Sect. 2.1.

The dataset contains estimates of taxi counts for  $L_o \stackrel{\text{def}}{=} 260,855$  one-directional links (roadways) in New York City; a two-directional road segment is represented as two one-directional links. We use OpenStreetMap labels. Table 1 gives a statistical summary of these links.

In total, the dataset of Donovan et al. (2016) thus has  $8,760 \times 260,855 \approx 2.3 \times 10^9$  entries. About 95% of these (2,181,923,208) are zero. Given that New York City is an urban environment and unlikely to

have empty roads at any time of the day, we reset these zero entries to NaN. To restrict our calculations to a subset for which we have somewhat reliable data, we will consider only those links for which there are at most a total of 720 missing hours (30 days worth) of data. We get  $\mathcal{L}$  with  $L \stackrel{\text{def}}{=} 2,302$ . Selecting more links comes with the benefit of incorporating more traffic data into our analysis while also introducing the cost of having a higher number of missing data entries. The distribution of entries is shown in Fig. 1. Multiplying these links by the time horizon of 8,760 hours gives over  $2 \times 10^7$  data points. Our goal is to understand how to appropriately identify and understand complex patterns in this data. Figure 1 shows that 2,302 of the most-traveled links captures 19% of the entries in  $D$ . Adding more (sparsely traveled) links would correspond to adding sparse columns to  $D$ , perhaps with diminishing returns for data interpretation.

### Constants, Initialization, and Computational Considerations

For our year-long taxi traffic dataset, on the links of  $\mathcal{L}$ , we use:

- $\eta = \max_{\substack{1 \leq t \leq T \\ 1 \leq \ell \leq L}} |D_{t,\ell}| = 4171$ ,
- rank:  $N = 50$ ,
- $\beta = 5000$ .

These values stem from a grid search of the results of the algorithm of Sect. 2.5. The initial entries of  $W$  and  $H$  (i.e., the initial condition of  $W$  and  $H$ ) are taken to be i.i.d. Uniform(0, 1) random variables (thus ensuring that  $W$  and  $H$  start with full rank). We disregard runs which lead to zero columns of  $W$ ; such zero columns are invariant under our multiplicative update rule and indicate sub-optimal use of rank. Note that, by (5) and (6), an iteration leading to a zero column of  $W$  will produce a zero row of  $H$  and vice versa. The algorithm is relatively insensitive to changes in  $\eta$ , and the above value of  $\eta$  leads to values of  $W$  which are not too large. Fixing the value of  $\eta$  (to the above value), we carry out a refined grid search on the  $(N, \beta)$  parameters. Table 2 gives specific numerical results for two pairs of  $(N, \beta)$  parameters.

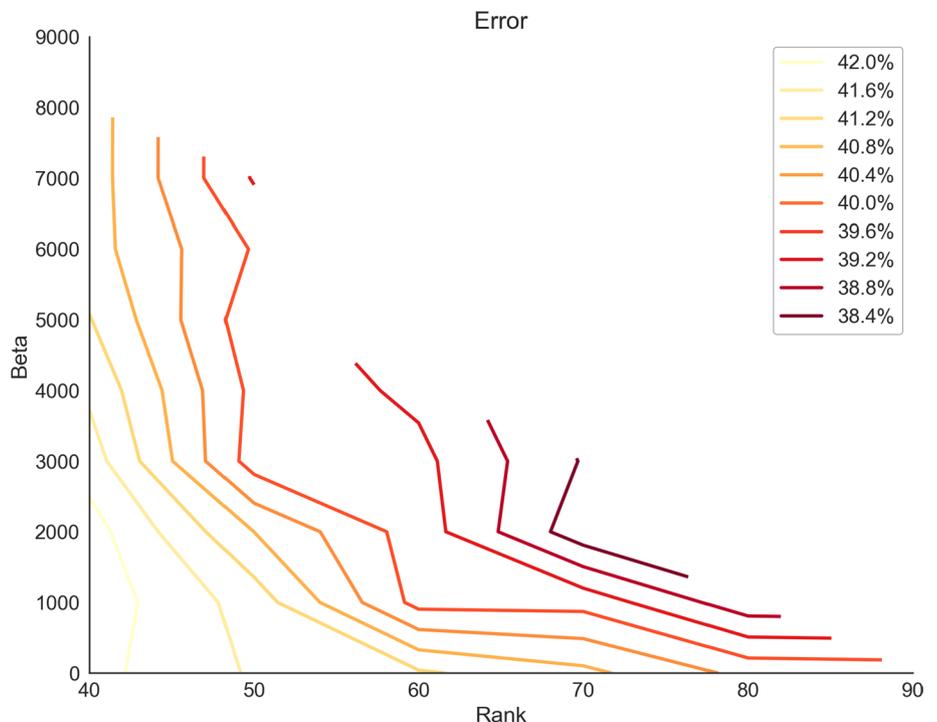
By thresholding  $H$ , we can forcibly express our approximation of each  $D$  as a linear combination of as few signatures as possible. For each column of  $H$ , we set all entries in each column below the 40th percentile of that column to zero. In practice, this reduces each link to a linear combination of at most eight signatures. Table 3 updates Table 2 once we have thresholded.

The results of the grid search in  $(N, \beta)$  are summarized in Figs. 2 and 3. These figures confirm a natural tradeoff:

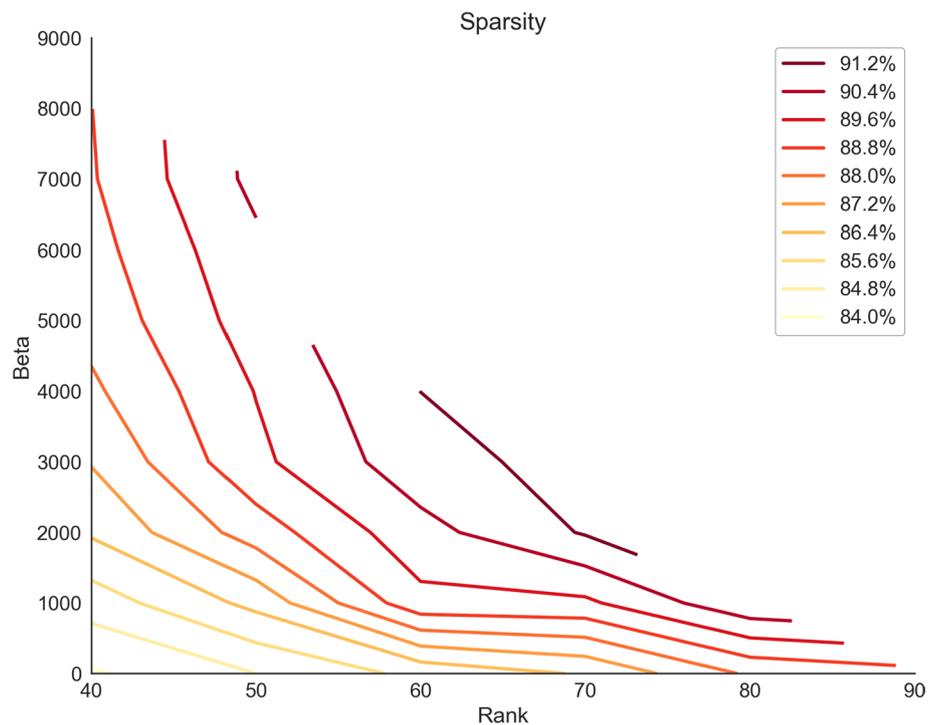
**Table 2** Comparison of two  $(N, \beta)$  choices without thresholding  $H$

	$N = 50,$ $\beta = 5000$	$N = 70,$ $\beta = 2000$
Error	25.8	23.7
Sparsity	0.789	0.814

**Fig. 2** Results of an  $(N, \beta)$  grid search: Relative error percentages (after thresholding  $H$ ) for various  $(N, \beta)$  pairs



**Fig. 3** Results of an  $(N, \beta)$  grid search: Sparsity of  $H$  (after thresholding  $H$ ) for various  $(N, \beta)$  pairs



**Table 3** Updated comparison of two  $(N, \beta)$  choices after thresholding  $H$

	$N = 50,$ $\beta = 5000$	$N = 70,$ $\beta = 2000$
Error before thresholding	25.8	23.7
Error after thresholding	39.3	38.2
Sparsity before thresholding	0.789	0.814
Sparsity after thresholding	0.900	0.913

higher  $N$  allows lower error and higher  $\beta$  leads to better sparsity, and these are competing objectives, however. We generally consider a sparsity value between 0.8 and 1.0 to be sufficient as it allows for only a handful of dominant signatures per link.

Informally, higher values of  $\beta$  are likely to mean sparser  $H$ , and higher values of  $N$  mean a larger universe of signatures. Hence, in cases where the algorithm produces a zero signature, we take this to mean that the matrix factorization needs a smaller rank.

To focus on the interpretation of the output of the algorithm, let us now fix  $N = 50$  (the smaller value of  $N$ ) and  $\beta = 5000$  for the rest of this section. The output consists of two matrices  $W$  and  $H$  of sizes  $8,760 \times 50$  and  $50 \times 2,302$ , respectively. The columns of  $W$  are  $L^1$ -normalized and the columns of  $H$  are sparse.

Recall that columns of  $W$  represent traffic signatures over time. Each signature is a time-series for the entire year and

hence need not be periodic. For example, the signatures can capture traffic anomalies during holidays, hurricanes, and blackouts.

Furthermore, the entries of a column of  $H$  are coefficients for the linear decomposition of a link into distinct signatures. For example, if the 4-th column of  $H$  is  $(0, 7, 2, 0, \dots, 0)^T$ , then the traffic in link 4 of  $\mathcal{L}$  can be written as seven times the second signature plus two times the third signature. This decomposition allows us to identify spatial patterns in traffic across the city. These matrices and the patterns derived from them can then aid in making specific observations about the large-scale behavior of traffic (as detailed in Sect. 4.4).

### Independence of Signatures

To quantify the linear independence of the signatures obtained from the algorithm, we can compute the condition number of  $WH$  (Hagen et al. (2000)). A high condition number (in the thousands) would indicate that the rank should be reduced. By performing several runs of our algorithm for rank  $N = 50$  with different  $(W, H)$ -initializations, we determine

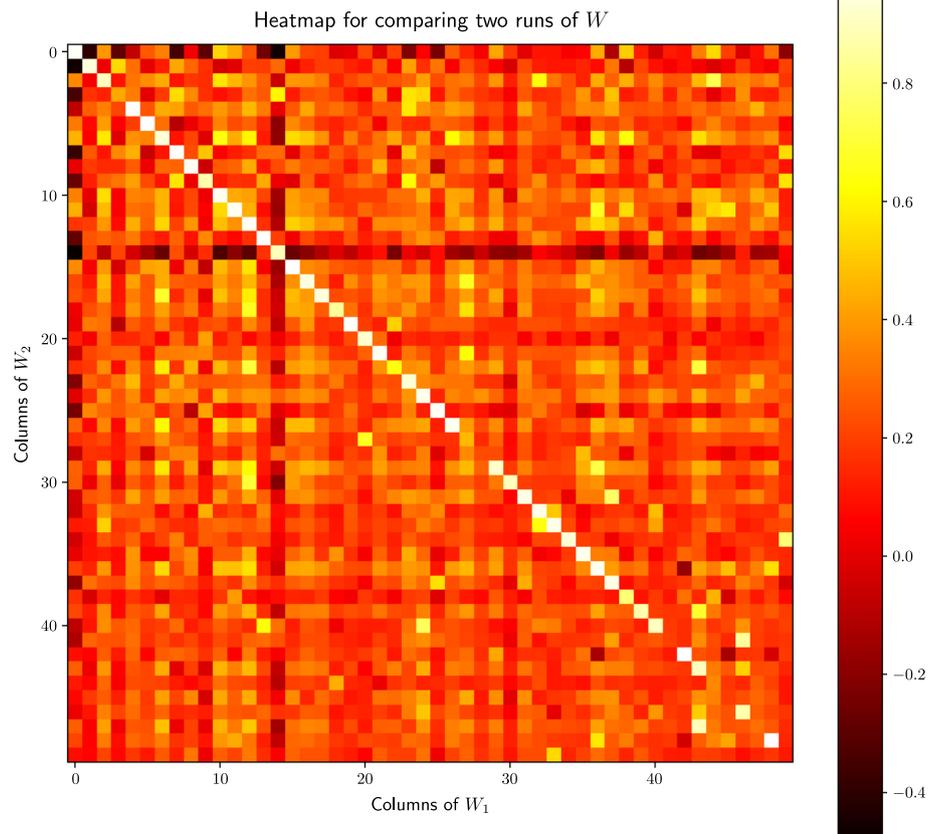
Condition Number for  $W = 24 \pm 2$

This number is low enough to give us confidence in the  $W$  returned by the algorithm., and further validates our choice of  $N = 50$ .

## Robustness of Algorithm

The factors produced by our algorithm are not unique and can differ by permutations. If  $W_1$  and  $W_2$  are produced by two runs of the algorithm (starting with different initial conditions), we can calculate the Pearson product-moment correlation coefficient between the columns of  $W_1$  and the columns of  $W_2$ . A coefficient value close to one implies that the signatures follow the same pattern up to a scale factor. We can construct a greedy algorithm for searching for a permutation which will maximize correlations by sequencing through the columns of  $W_1$ , finding the column of  $W_2$  which maximizes the correlation (with respect to the selected column of  $W_1$ ), and then removing that column of  $W_2$  from future computations. We can thus permute the columns of  $W_2$  to better match those of  $W_1$ . After doing this, we can construct a heatmap which shows the correlation between the columns of  $W_1$  and the columns of (the permuted)  $W_2$ . Figure 4 shows this heatmap for two runs of the Algorithm of Sect. 2.5 with different (random) initial conditions and with the prevailing values of  $\eta = 4171$ ,  $N = 50$  and  $\beta = 5000$  (see Sect. 4.1). 4 gives the correlation between the  $i$ -th column of  $W_1$  and the  $j$ -th column of the permuted  $W_2$ . We see high correlation on the diagonal, meaning that the original matrix  $W_2$  is very close to a permutation of  $W_1$ .

**Fig. 4** Heatmap of correlation coefficients of  $W$ -columns from two runs of the algorithm



We make observations about the low-rank decomposition are in Sect. 4.4.

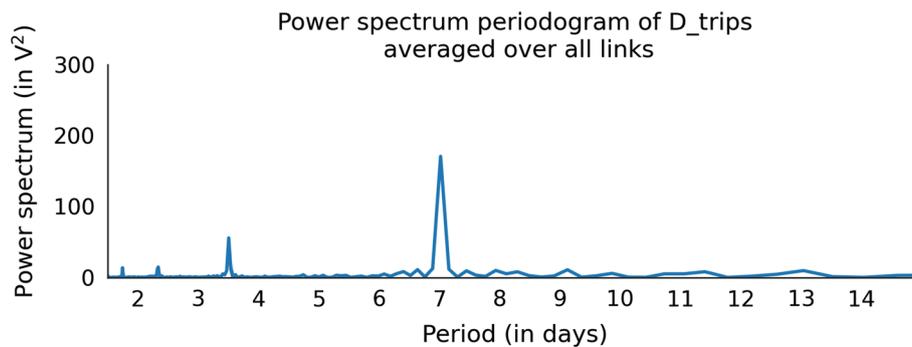
## Periodicity and Anomalous Observations

We note that the columns of  $D$  (and hence the signatures) are roughly periodic, with a period of 7 days. A power spectrum periodogram of  $D$  (averaged over all columns) is shown in Fig. 5.

In light of this periodicity, we can look at one day of the week across the entire year (e.g., all Mondays) and compute the hour-wise median traffic for that day and then identify anomalous behavior. See Figs. 6, 7a, 8a, and 8b (in dark red). In gray, we plot the relative taxi counts i.e., entries of the normalized signatures. We then determine which dates have relative taxi counts that differ significantly from the hour-wise median traffic in the sense that those (*signature, day*) pairs have the highest sum of absolute deviations from the median weekly traffic. In the subsections below, we identify some possible origins of anomalous behavior.

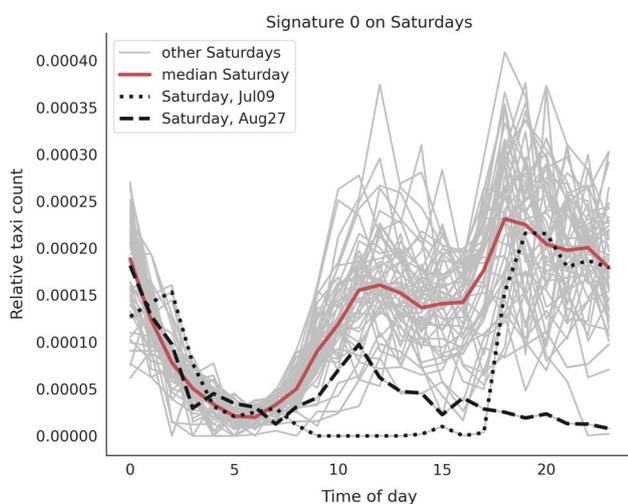
### Hurricane Irene

Figure 6 shows Signature 0, which captures a near-shut-down of taxi traffic on August 27, 2011. This may have



**Fig. 5** A power-spectrum periodogram showing 7-day periodicity in taxi count data averaged over all links. It is standard practice to measure y-axis values in periodograms in terms of Volt-squared. The peak

at 7 shows weekly periodicity. The peaks at 3.5 days (i.e. half-time-step), 2.33 days (i.e. a third-time-step), and 1.75 days (i.e. a quarter-time-step) can be seen as overtones of the 7-day periodicity



**Fig. 6** Signature 0 during Hurricane Irene (August 27). The other low traffic volume day marked by the dashed line is July 09 which saw a Yankees vs. Tampa Bay Rays baseball game in Bronx and was attended by more than 48 thousand people

been caused by Hurricane Irene hitting NYC. There was an early warning and all subways and buses were shut down at noon on Saturday, August 27. A zoned taxi system was implemented at 9 am and taxis were thereafter running flat fares instead of meters [wny]. All other signatures also show similar behavior on and around August 28.

### Wisconsin Labor Rally

Figure 7a shows the behavior of Signature 21 on February 26, 2011. The traffic deviates from the median Saturday trend. This may have been caused by a Labor Rally that took place near the New York City Town Hall [wis]. Figure 7b shows that Signature 21 is used by links near the Town Hall, which can be seen as further evidence connecting the rally to this traffic deviation.

### Christmas Day

Anomalous behavior was also observed on Christmas Day. This can be seen in Fig. 8a for Signature 0 and Fig. 8b for Signature 4. Note how the Christmas Day traffic is reduced by about half throughout the day.

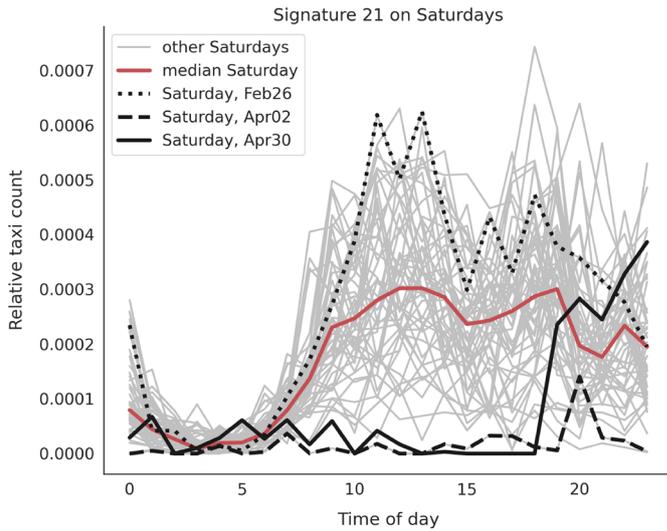
### Endemic Signatures

After thresholding as discussed in Sect. 4.1, we can show the support  $\{\ell \in \mathcal{L} : H_{j,\ell} > 0\}$  of signature  $n$  on a map, for any given  $n \in \{1, 2 \dots N\}$ . See Fig. 9. We note that of the 50 signatures, some tend to be geographically restricted (called *endemic*), while others are spread out over larger areas (called *dispersed*). For example, Signature 0, as seen in Fig. 9a is dispersed. Endemic signatures might sometimes explain traffic densities only on a single but long stretch of road. For example, Fig. 9b shows that Signature 10 is largely used by the the northbound 3rd Avenue and streets like Bowery, Lafayette St. and the southernmost part of Broadway that feed into 3rd Avenue. Similarly, Fig. 9d shows Signature 40 being used exclusively by a small section of the south-bound Broadway traffic near Central Park.

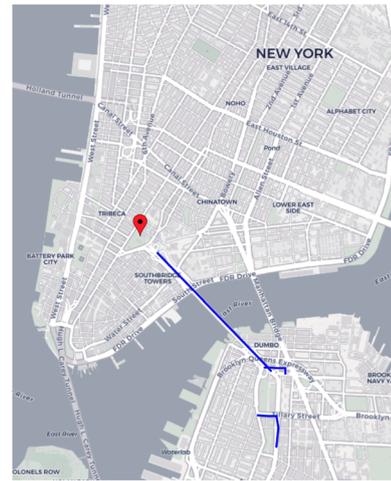
In some other cases, signatures can be seen as having a lateral sphere of influence in that they affect not only one street but also others feeding into or out of the street transversally—Signature 24, for example, as seen in Fig. 9c.

### Analysis of New York City Data: Error Analysis

We second carry out the error analysis of Sect. 3. We will use



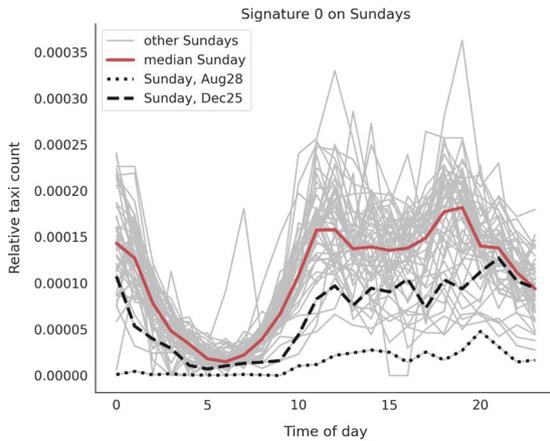
(A) Signature 21 during the Wisconsin Labor Rally.



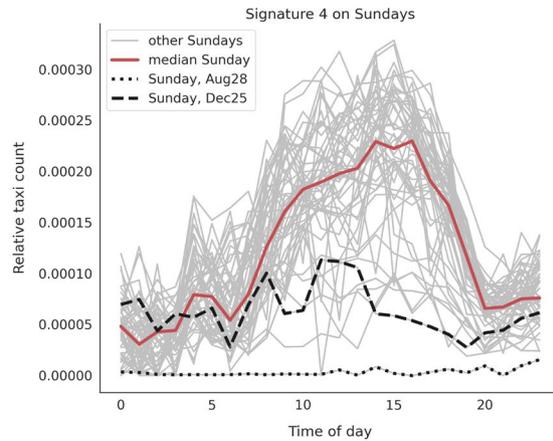
(B) A map of signature 21 with location of Wisconsin Labor Rally.

**Fig. 7** Signature 21 during February 26 2011. The other days with anomalous traffic compared to median signature 21 traffic were April 2 (International Pillow fight Day which involved public gatherings at the Union Square as well as several police blockades) and April 30 (cause for anomalous traffic unknown). This underscores the fact that

the matrix factorization can pick out patterns and anomalous behavior but it is up to us to make sense of it and assign a sensible cause to each anomaly. **a** Signature 21 during the Wisconsin Labor Rally. **b** A map of signature 21 with location of Wisconsin Labor Rally



(A) Signature 0 during Christmas day. The other marked day represents Hurricane Irene day 4 (August 28).



(B) Signature 4 during Christmas day. The other marked day represents Hurricane Irene day 4 (August 28).

**Fig. 8** Anomalous traffic during Christmas day. **a** Signature 0 during Christmas day. The other marked day represents Hurricane Irene day 4 (August 28). **b** Signature 4 during Christmas day. The other marked day represents Hurricane Irene day 4 (August 28)

$$\mathcal{N}^{\text{def}} = \{40, 50, 60, 70, 80, 90\}$$

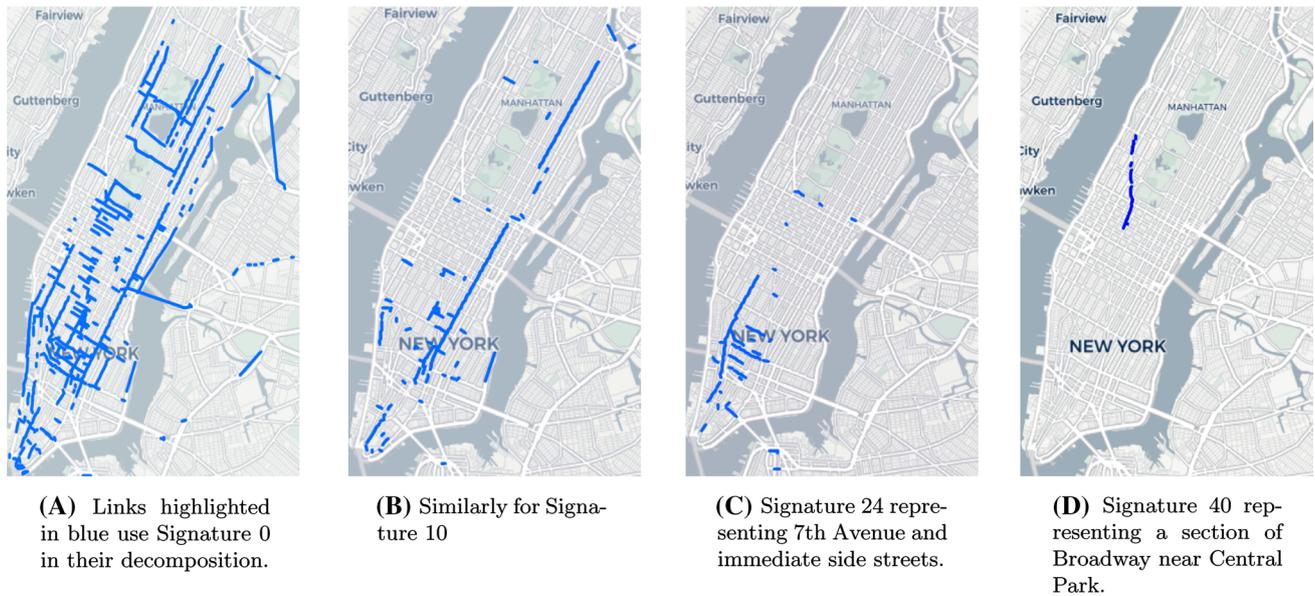
$$\mathcal{B}^{\text{def}} = \{0, 1000, 2000, \dots, 9000\}.$$

as the sets of (11).

Figure 10 gives us Plot 1. We see that Sparsity increases with an increase in  $\beta$  for every value of  $N \in \mathcal{N}$ . This

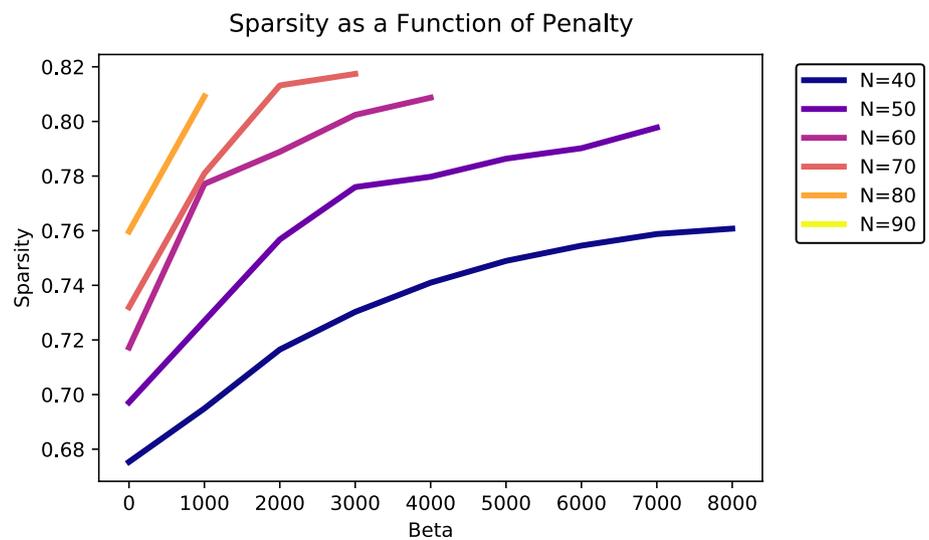
confirms that we can reparametrize  $\beta$  as Sparsity\* from Sect. 3.2 for fixed values of  $N$ .

Figures 11, 12, 13 14 gives Plot 2. Namely, for each season  $\mathcal{T}$ , it gives (Sparsity\*( $N, \beta$ ),  $\mathcal{E}^*(N, \beta, \mathcal{T})$ ). We compare each season with the yearly average and note that in spring and winter, the seasonal (relative) error of our approximation



**Fig. 9** Spatial Patterns. **a** Links highlighted in blue use Signature 0 in their decomposition. **b** Similarly for Signature 10 **(c)** Signature 24 representing 7th Avenue and immediate side streets. **d** Signature 40 representing a section of Broadway near Central Park

**Fig. 10** Sparsity as a function of  $\beta$  for different values from  $\mathcal{N}$  (Plot type 1)



is lesser than the yearly (relative) error. This indicates that spring and winter traffic are comparatively low-rank and thus more orderly. The fall and summer traffic are more disordered and have higher approximation error than the yearly average. We also note that the gap between the season and annual approximation errors is more pronounced in fall and winter compared to spring and summer. This indicates that spring and summer are more representative of the annual traffic trends while fall and winter capture more of the anomalous behavior in our traffic dataset.

Figures 15, 16, 17, 18 gives Plot 3. As with Figures 11, 12, 13, 14, we see a bit more order in the spring and winter, and a bit more disorder in the summer and fall. (Since we

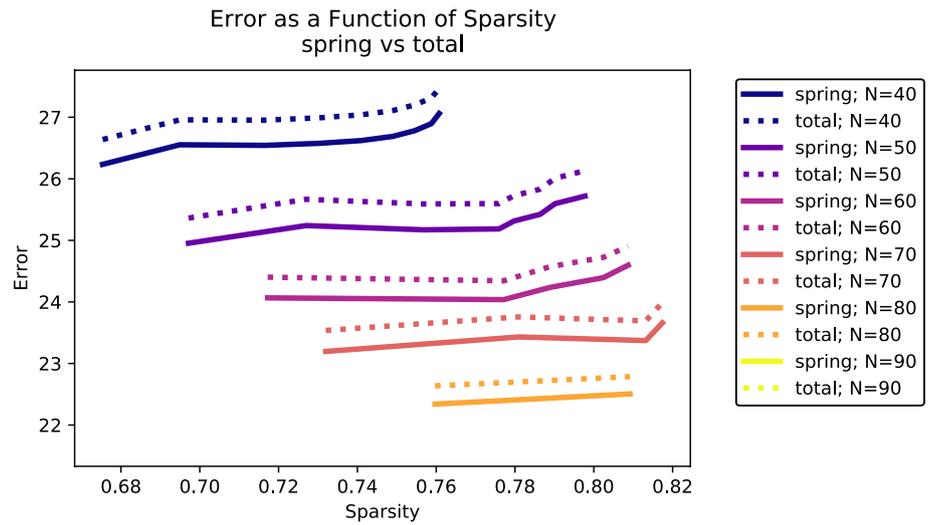
restrict the analysis to links missing at most 30 days (1/4 of a season), these conclusions we are likely to be statistically meaningful).

### Variables Other Than Taxi-Counts

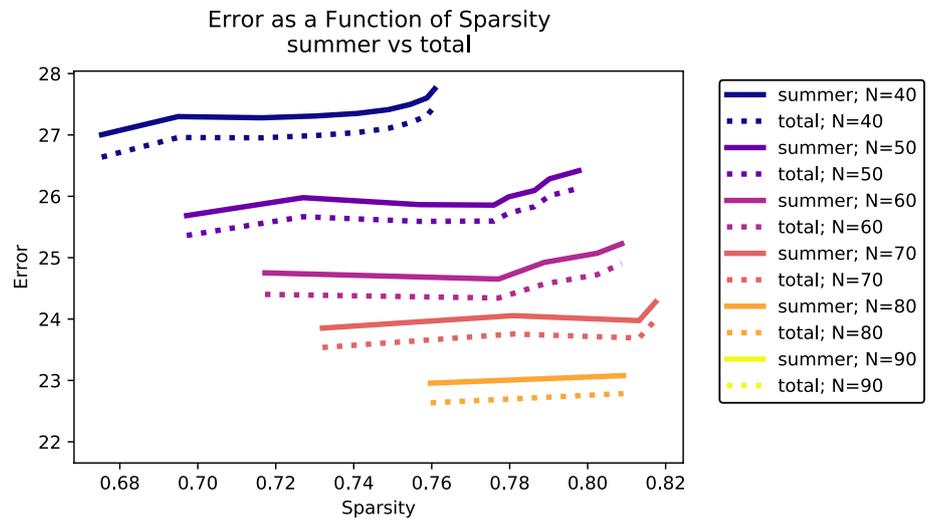
We looked at variables other than taxi counts and observed similar low-rank (periodic) behavior in the data. Figure 19 shows 7-day periodicity in taxi travel times, speeds, pace (reciprocal of speed) and taxi density (which was computed as a ratio of the taxi counts to link lengths).

A matrix factorization similar to the one shown here for taxi counts might be mathematically carried out for each of

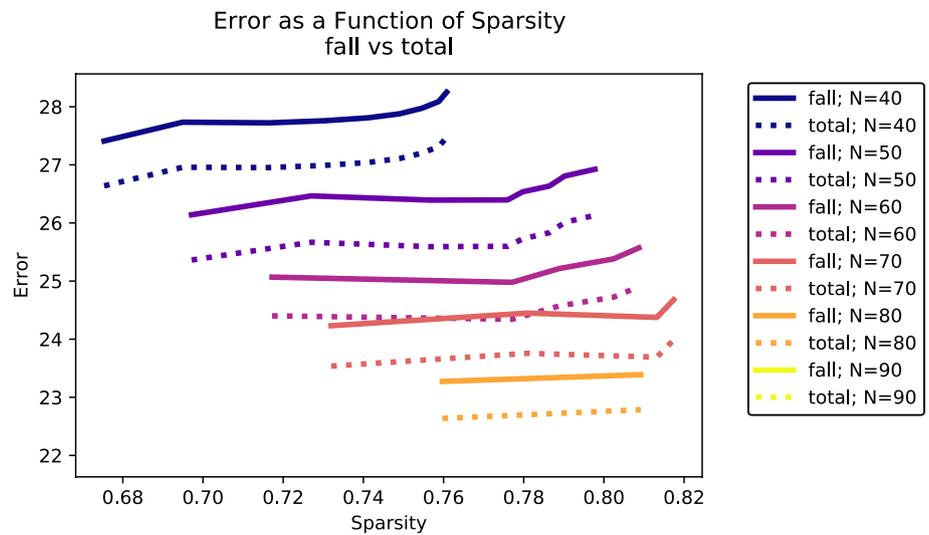
**Fig. 11** Seasonal Error as a function of Sparsity (Plot type 2): Spring



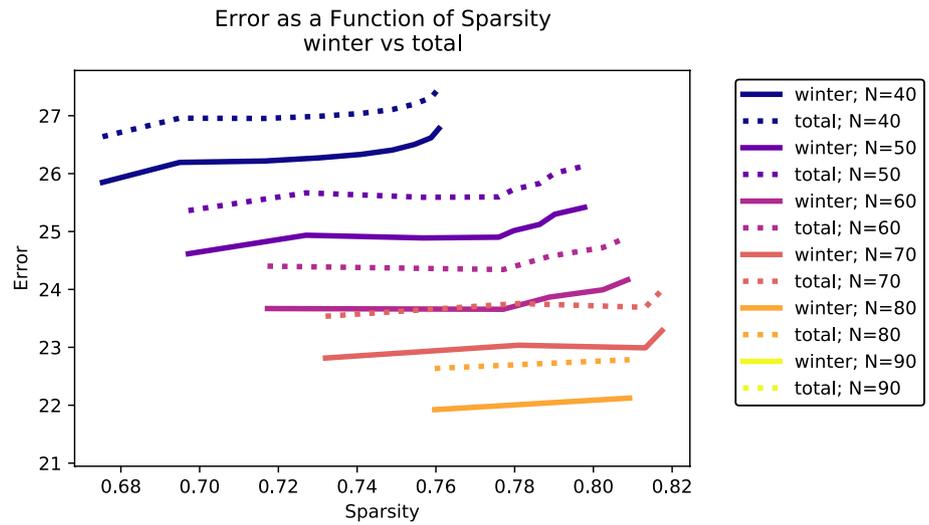
**Fig. 12** Seasonal Error as a function of Sparsity (Plot type 2): Summer



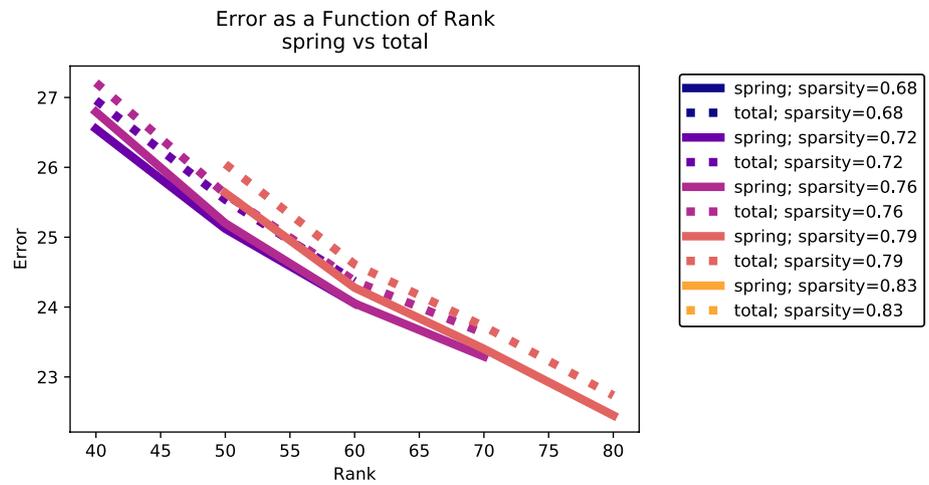
**Fig. 13** Seasonal Error as a function of Sparsity (Plot type 2): Fall



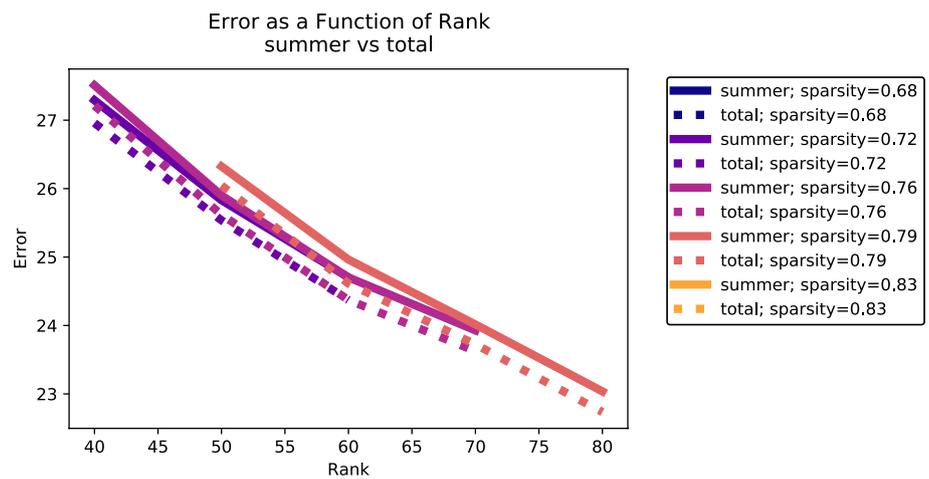
**Fig. 14** Seasonal Error as a function of Sparsity (Plot type 2): Winter



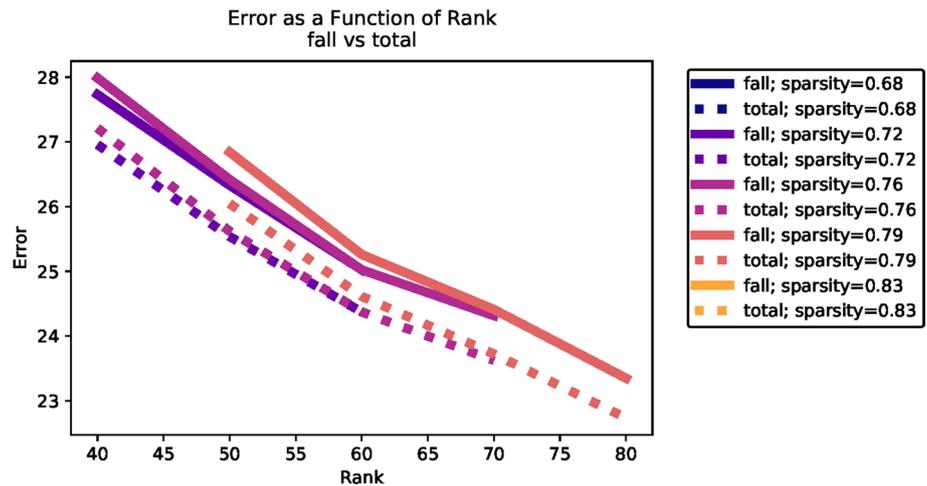
**Fig. 15** Seasonal Error as a function of rank  $N$  (Plot type 3): Spring



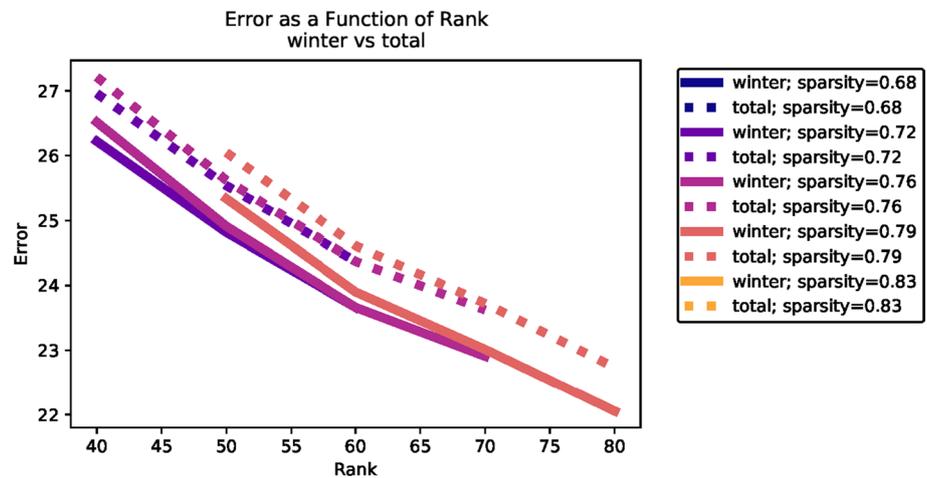
**Fig. 16** Seasonal Error as a function of rank  $N$  (Plot type 3): Summer



**Fig. 17** Seasonal Error as a function of rank  $N$  (Plot type 3): Fall



**Fig. 18** Seasonal Error as a function of rank  $N$  (Plot type 3): Winter



these variables. However, since pace and speed are not additive, the interpretation of these matrix factorizations would have to be carefully motivated.

## Conclusions

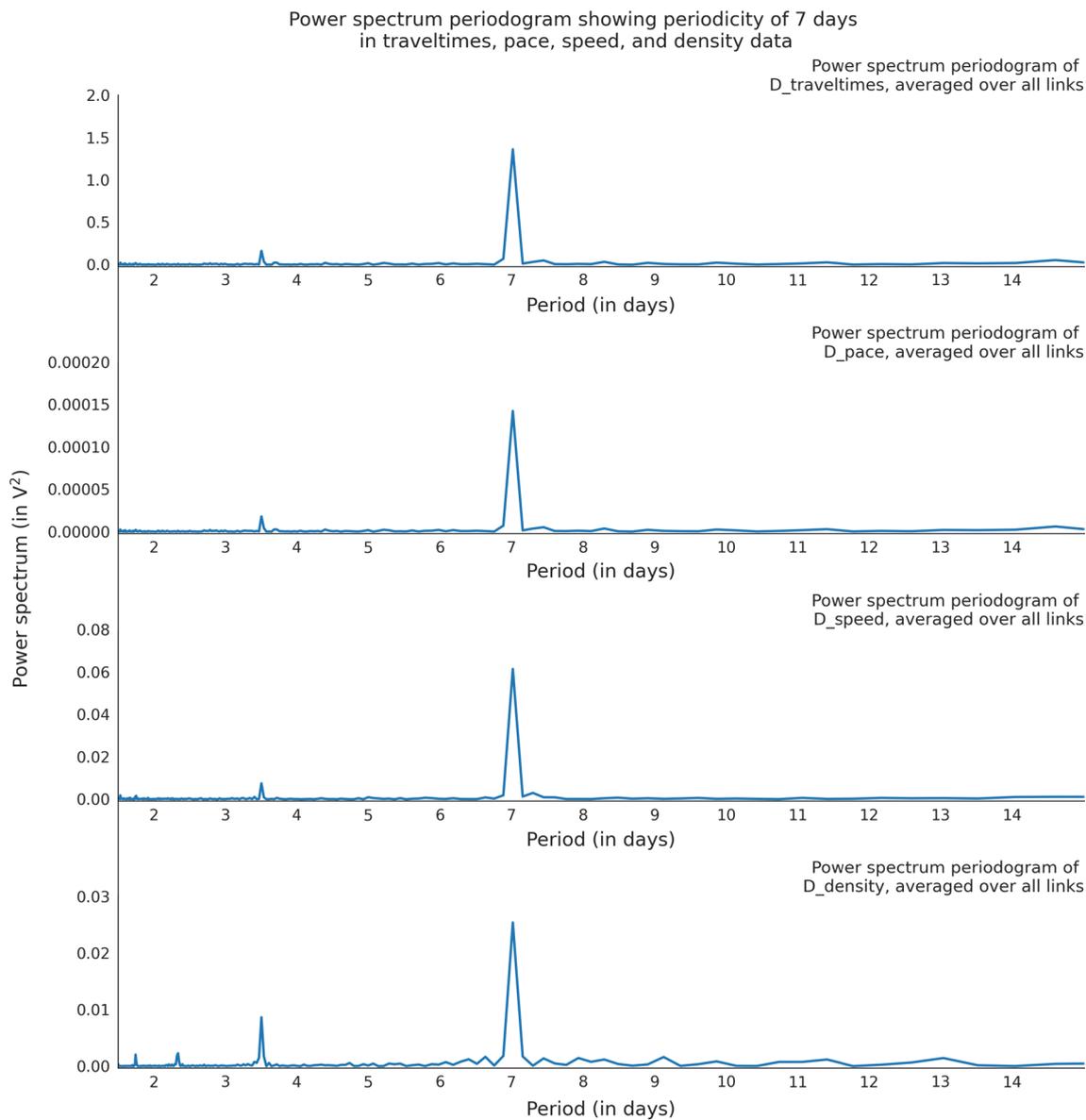
In this paper, we studied a New York City taxi traffic dataset using SNMF techniques. This gave us some insight into underlying behavior with a small number of signatures. It also enabled identification of anomalous traffic patterns which we visualized in Sect. 4.4. These visualizations captured several widespread anomalous events.

We also developed an analysis which might suggest comparisons between traffic patterns in different circumstances (viz. different time intervals). Various city attributes can be connected to different traffic conditions and signatures. This might be relevant for adaptive urban planning and scheduling.

One result of any coarse-graining analysis, such as ours, is to help focus exploration of large datasets around patterns and events of interest. If one already understands the data and is designing, e.g., an incident detection algorithm, then other (perhaps model-based) approaches might give more refined and/or complementary insights.

This work did not directly address generalizability of this approach to other cities. Comparable datasets from other cities would enable us to search for city-specific idiosyncratic behavior as opposed to behavior which might be somewhat invariant to the urban environment.

Starting with our analysis of seasonal error in Sect. 5, another interesting future direction would be a closer look at how different signatures (which last over the interval (13) of a year) might preferentially represent behavior of different seasons. Informally, the rank of signatures representing a particular season's behavior might be another indication of the complexity of traffic in that season.



**Fig. 19** A power-spectrum periodogram of taxi-traffic data for travel times, pace (i.e. inverse of speed), speed and density (i.e. taxi count divided by link lengths). The data were averaged over all links. These periodograms are very similar to the one shown in Fig. 5 for taxi counts

### Appendix

For completeness, let’s write down the calculations leading to the algorithm of Sect. 2.5.

Writing out  $\mathcal{E}_{\beta,\eta}$  of (5), we get

$$\mathcal{E}_{\beta,\eta}(W, H) = \sum_{(t,\ell) \in \mathcal{I}} |D_{t,\ell} - (WH)_{t,\ell}|^2 + \beta \sum_{n=1}^N \left( \sum_{\ell=1}^L H_{n,\ell} \right)^2 + \eta \sum_{n=1}^N \sum_{t=1}^T W_{\ell,t}^2.$$

We seek to minimize this by alternating between minimization problems in  $W$  and  $H$ . Namely, if we start with a fixed  $(W, H) \in \mathbb{R}_+^{T \times N} \times \mathbb{R}_+^{N \times L}$ , we can construct a descent step for the function  $\mathcal{E}_{\beta,\eta}(W, \cdot)$  and then, letting  $H'$  be the result, we can construct a descent step for  $\mathcal{E}_{\beta,\eta}(\cdot, H')$ . This should decrease the value of  $\mathcal{E}_{\beta,\eta}$ , and we can then proceed iteratively.

The gradients of  $\mathcal{E}_{\beta,\eta}$  in the directions of  $W$  and  $H$  are given by

$$\begin{aligned} \frac{\partial \mathcal{E}_{\beta, \eta}}{\partial W_{i, \hat{n}}}(W, H) &= -2 \sum_{\ell: (i, \ell) \in \mathcal{I}} \left( D_{i, \ell} - \sum_{n=1}^N W_{i, n} H_{n, \ell} \right) H_{i, \ell} \\ &\quad + \eta W_{i, \hat{n}} \\ &= -2([D - WH]_{\mathcal{I}} H^T + \eta W)_{i, \hat{n}} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \mathcal{E}_{\beta, \eta}}{\partial H_{\hat{n}, \ell}}(W, H) &= -2 \sum_{t: (t, \ell) \in \mathcal{I}} \left( D_{t, \ell} - \sum_{n=1}^N W_{t, n} H_{n, \ell} \right) W_{t, \hat{n}} \\ &\quad + 2\beta \left( \sum_{n=1}^N H_{n, \ell} \right) \\ &= -2(W^T [D - WH]_{\mathcal{I}})_{\hat{n}, \ell} \\ &\quad + 2\beta(\mathbf{1}_{N \times N} H)_{\hat{n}, \ell}. \end{aligned}$$

As in Kim and Park (2008), we want to iteratively find the critical points of  $\mathcal{E}_{\beta, \eta}$ , i.e. the solutions of

$$\begin{aligned} [WH]_{\mathcal{I}} H^T - [D]_{\mathcal{I}} H^T + \eta W &= 0 \\ W^T [WH]_{\mathcal{I}} - W^T [D]_{\mathcal{I}} + \beta \mathbf{1}_{N \times N} H &= 0 \end{aligned}$$

The above formulae suggest a multiplicative descent rule (which need not be *gradient descent*; see Lee and Seung (2001)). Fix  $(W, H) \in \mathbb{R}_+^{T \times N} \times \mathbb{R}_+^{N \times L}$ . Assume that

$$\frac{\partial \mathcal{E}_{\beta, \eta}}{\partial H_{n, \ell}} > 0; \quad (14)$$

we can then decrease the value of  $\mathcal{E}_{\beta, \eta}$  by decreasing  $H_{n, \ell}$ . Rewriting (14) as

$$-2(W^T [D - WH]_{\mathcal{I}})_{n, \ell} + 2\beta(\mathbf{1}_{N \times N} H)_{n, \ell} > 0$$

or rather

$$(W^T [WH]_{\mathcal{I}})_{n, \ell} + \beta(\mathbf{1}_{N \times N} H)_{n, \ell} > (W^T [D]_{\mathcal{I}})_{n, \ell},$$

since  $W$ ,  $H$ , and  $D$  all have nonnegative entries, both sides of this equation are nonnegative. This in turn can be written as  $\chi_{n, \ell}^h(W, H) < 1$  where

$$\chi_{n, \ell}^h(W, H) \stackrel{\text{def}}{=} \frac{(W^T [D]_{\mathcal{I}})_{n, \ell}}{(W^T [WH]_{\mathcal{I}})_{n, \ell} + \beta(\mathbf{1}_{N \times N} H)_{n, \ell}}.$$

Thus, another way to decrease  $H_{n, \ell}$  while still retaining non-negativity is to multiply it by  $\chi_{n, \ell}^h(W, H)$ . Reviewing these steps, we also see that if  $\frac{\partial \mathcal{E}_{\beta, \eta}}{\partial H_{n, \ell}} < 0$ , we want to increase  $H_{n, \ell}$ , and can again multiply by  $\chi_{n, \ell}^h(W, H)$ . Finally, if  $\frac{\partial \mathcal{E}_{\beta, \eta}}{\partial H_{n, \ell}} = 0$  (i.e., we have found a critical point)  $\chi_{n, \ell}^h(W, H) = 1$ , so multiplying  $H_{n, \ell}$  by  $\chi_{n, \ell}^h(W, H)$  leaves  $H_{n, \ell}$  unchanged.

The update rule for  $W_{t, n}$  is similar. To start, assume that

$$\frac{\partial \mathcal{E}_{\beta, \eta}}{\partial W_{t, n}} > 0; \quad (15)$$

then we can decrease  $\mathcal{E}_{\beta, \eta}$  by decreasing  $W_{t, n}$ . We can rewrite (15) as

$$-2([D - WH]_{\mathcal{I}} H^T + \eta W)_{t, n} > 0.$$

We can again rewrite this as the comparison of two non-negative quantities;

$$([WH]_{\mathcal{I}} H^T + \eta W)_{t, n} > ([D]_{\mathcal{I}} H^T)_{t, n};$$

This in turn is equivalent to  $\chi_{t, n}^w(W, H) < 1$  where

$$\chi_{t, n}^w(W, H) \stackrel{\text{def}}{=} \frac{([D]_{\mathcal{I}} H^T)_{t, n}}{([WH]_{\mathcal{I}} H^T + \eta W)_{t, n}}$$

In other words, we can decrease  $W_{t, n}$  by multiplying by  $\chi_{t, n}^w(W, H)$ . One can similarly see that if  $\frac{\partial \mathcal{E}_{\beta, \eta}}{\partial W_{t, n}} < 0$ , gradient descent again increases or decreases  $W$  with the same sign as multiplying by  $\chi_{t, n}^w(W, H)$ .

Our proposed update rule for  $W$  and  $H$  is now

$$W'_{t, n} = W_{t, n} \chi_{t, n}^w(W, H)$$

$$H'_{n, \ell} = H_{n, \ell} \chi_{n, \ell}^h(W, H).$$

which is equivalent to (6).

## Compliance with Ethical Standards

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Asif MT, Kannan S, Dauwels J, Jaillet P (2013) Data compression techniques for urban traffic data. In: 2013 IEEE symposium on computational intelligence in vehicles and transportation systems (CIVTS), pages 44–49. IEEE
- Ahmadi P, Kaviani R, Gholampour I, Tabandeh Mahmoud (2015) Modeling traffic motion patterns via non-negative matrix factorization. In 2015 IEEE international conference on signal and image processing applications (ICSIPA), pages 214–219. IEEE
- Alonso-Mora J, Samaranayake S, Wallar A, Frazzoli E, Rus D (2017) On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. In: Proceedings of the National Academy of Sciences, page 201611675
- Ban XJ, Hao P, Sun Z (2011) Real time queue length estimation for signalized intersections using travel times from mobile sensors. Trans Res Part C 19(6):1133–1156
- Boquet G, Morell A, Serrano J, Vicario JL (2020) A variational autoencoder solution for road traffic forecasting systems: missing data imputation, dimension reduction, model selection and anomaly detection. Trans Res Part C 115:102622

- Brunet J-P, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci* 101(12):4164–4169
- Chagoyen M, Carmona-Saez P, Shatkay H, Carazo JM, Pascual-Montano A (2006) Discovering semantic features in the literature: a foundation for building functional associations. *BMC Bioinformatics* 7(1):41
- Chen X, He Z, Sun L (2019) A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Trans Res Part C* 98:73–84
- Cazabet R, Jensen P, Borgnat P (2018) Tracking the evolution of temporal patterns of usage in bicycle-sharing systems using nonnegative matrix factorization on multiple sliding windows. *Int J Urban Sci* 22(2):147–161
- Carmona-Saez P, Pascual-Marqui Roberto D, Tirado Francisco, Carazo Jose M, Pascual-Montano Alberto (2006) Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics* 7(1):78
- Carrasco DR, Tonon G, Huang Y, Zhang Y, Sinha R, Feng B, Stewart JP, Zhan F, Khatry D, Protopopova M et al (2006) High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer cell* 9(4):313–325
- Deri JA, Moura JMF (2015) Taxi data in new york city: a network perspective. In: 2015 49th asilomar conference on signals, systems and computers, pages 1829–1833, Nov 2015
- Donovan B, Mori A, Agrawal N, Meng Y, Lee J, Work D (2016) New York City hourly traffic estimates (2010–2013). [https://doi.org/10.13012/B2IDB-4900670\\_V1](https://doi.org/10.13012/B2IDB-4900670_V1)
- Dueck D, Morris Quaid D, Frey BJ (2005) Multi-way clustering of microarray data using probabilistic sparse matrix factorization. *Bioinformatics* 21(suppl-1):i144–i151
- Donovan B, Work Daniel B (2015) Using coarse GPS data to quantify city-scale transportation system resilience to extreme events. arXiv preprint [arXiv:1507.06011](https://arxiv.org/abs/1507.06011)
- Djenouri Y, Zimek A, Chiarandini M (2018) Outlier detection in urban traffic flow distributions. In 2018 IEEE international conference on data mining (ICDM), pages 935–940
- Ermagun A, Levinson D (2018) Spatiotemporal traffic forecasting: review and proposed directions. *Trans Res* 38(6):786–814
- Ferreira N, Poco J, Vo HT, Freire J, Silva CT (2013) Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. *IEEE Trans Vis Comput Graph* 19(12):2149–2158
- Gao Y, Church G (2005) Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics* 21(21):3970–3975
- Guan X, Chen C, Work D (2016) Tracking the evolution of infrastructure systems and mass responses using publicly available data. *PloS one* 11(12):e0167267
- Geroliminis N, Daganzo CF (2008) Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Trans Res Part B* 42(9):759–770
- Guo J, Huang W, Williams BM (2015) Real time traffic flow outlier detection using short-term traffic conditional variance prediction. *Trans Res Part C* 50:160–172
- Gong Y, Li Z, Zhang Jian, Liu W, Zheng Y, Kirsch C (2018) Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization. In: Proceedings of the 27th ACM international conference on information and knowledge management, pages 1243–1252. ACM
- Hoffleitner A, Herring R, Bayen A, Han Y, Moutarde F, De La Fortelle A (2012) Large scale estimation of arterial traffic and structural analysis of traffic patterns using probe vehicles. In Transportation Research Board 91st Annual Meeting (TRB'2012)
- Han Y, Moutarde F (2011) Analysis of network-level traffic states using locality preservative non-negative matrix factorization. pages 501–506, 10
- Han Y, Moutarde F (2013) Statistical traffic state analysis in large-scale transportation networks using locality-preserving non-negative matrix factorisation. *IET Intell Trans Syst* 7(3):283–295
- Han Yufei, Moutarde Fabien (2016) Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. *Int J Intell Trans Syst Res* 14(1):36–49
- Hoyer PO (2002) Non-negative sparse coding. In: Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on, pages 557–565. IEEE
- Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. *J Mach Learn Res* 5(Nov):1457–1469
- Herman R, Prigogine I (1979) A two-fluid approach to town traffic. *Science* 204(4389):148–151
- Ronald H, Steffen R, Bernd S (2000)  $C^*$ -algebras and numerical analysis. CRC Press, Boca Raton
- Ito K, Ito M, Miyazaki K, Tanimoto K, Sezaki K (2017) Data analysis on train transportation data with nonnegative matrix factorization. In: 2017 IEEE international conference on big data (Big Data), pages 4080–4085. IEEE
- Krichene W, Castillo MS, Bayen A (2016) On social optimal routing under selfish learning. In: IEEE transactions on control of network systems
- Kim H, Park H (2007) Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23(12):1495–1502
- Kim H, Park H (2008) Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J Matrix Anal Appl* 30(2):713–730
- Kim PM, Tidor B (2003) Subsystem identification through dimensionality reduction of large-scale gene expression data. *Genome Res* 13(7):1706–1718
- Karve V, Yager D, Abolhelm M, Work D, Sowers R NYC Traffic Patterns cSNMF Source Code. <https://gitlab.engr.illinois.edu/TrafficPatterns/CSNMF.git>
- Liu Z, Cao J, Yang J, Wang Q (2017) Discovering dynamic patterns of urban space via semi-nonnegative matrix factorization. In: 2017 IEEE international conference on big data (Big Data), pages 3447–3453. IEEE
- Lv Y, Duan Y, Kang W, Li Z, Wang FY (2015) Traffic flow prediction with big data: a deep learning approach. *IEEE Trans Intell Trans Syst* 16(2):865–873
- Li Stan Z, Hou XW, Zhang HJ, Cheng QS (2001) Learning spatially localized, parts-based representation. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–I. IEEE
- Li Q, Jianming H, Yi Z (2007) A flow volumes data compression approach for traffic network based on principal component analysis. In: 2007 IEEE intelligent transportation systems conference, pages 125–130
- Lee T, Matsushima S, Yamanishi K (2016) Traffic risk mining using partially ordered non-negative matrix factorization. In: 2016 IEEE international conference on data science and advanced analytics (DSAA), pages 622–631. IEEE
- Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. In: Advances in neural information processing systems, pages 556–562
- Li L, Xiaonan S, Zhang Y, Lin Y, Li Z (2015) Trend modeling for traffic time series analysis: an integrated study. *IEEE Trans Intell Trans Syst* 16(6):3430–3439
- Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatry D, Feng B, Sinha R, Louis DN et al (2006) Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 66(23):11502–11513
- Ma X, Li Y, Chen P (2018) Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative

- matrix factorization algorithm. *Journal of Traffic and Transportation Engineering (English Edition)*
- Mahmassani HS, Williams JC, Herman R (1984) Investigation of network-level traffic flow relationships: some simulation results. *Trans Res Record* 971:121–130
- Nagy AM, Simon V (2018) Survey on traffic prediction in smart cities. *Pervasive Mobile Comput* 50:148–163
- Pavlyuk D (2019) Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review. *Euro Trans Res Rev* 11(1):6
- Caltrans Performance Measurement System. <http://pems.dot.ca.gov/>
- Alberto P-M, Maria CJ, Kieko K, Dietrich L, Pascual-Marqui RD (2006) Nonsmooth nonnegative matrix factorization (nsnmf). *IEEE Trans Pattern Anal Mach Intell* 28(3):403–415
- Paul Pauca V, Piper J, Plemmons RJ (2006) Nonnegative matrix factorization for spectral data analysis. *Linear algebra and its applications* 416(1):29–47
- Pauca VP, Shahnaz F, Berry MW, Plemmons RJ (2004) Text mining using non-negative matrix factorizations. In: *Proceedings of the 2004 SIAM international conference on data mining*, pages 452–456. SIAM
- Pehkonen P, Wong G, Törönen P (2005) Theme discovery from gene lists for identification and viewing of multiple functional groups. *BMC bioinformatics* 6(1):162
- Lijun S, Kay WA. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, 91:511–524
- Hongzhi W, Mohamed JB, Mohamed H. Progress in outlier detection techniques: A survey. *IEEE Access*, 7:107964–108000
- Wisconsin worries: Labor rallies in NY. <https://nypost.com/2011/02/26/wisconsin-worries-labor-rallies-in-ny/>
- NY subway system shuts down due to Hurricane Irene (updated). <http://www.wnyc.org>
- Xu L, Wang Y, Yu H, Li H (2015) Feature extraction of urban traffic network data based on locally sensitive discriminant analysis algorithm
- Yangyang X, Yin W, Wen Z, Zhang Y (2012) An alternating direction algorithm for matrix completion with nonnegative factors. *Front Math China* 7(2):365–384
- Yang S, Qian S (2019) Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents. *IEEE Intell Trans Syst Mag* 11(3):12–28
- Zhang Z, He Q, Tong H, Gou J, Li X (2016) Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network. *Trans Res Part C* 71:284–302
- Zheng J Liu HX (2017) Estimating traffic volumes for signalized intersections using connected vehicle data. *Trans Res Part C* 79:347–362
- Yuan Z, Kaan O, Kun X, Hong Y (2016) Using big data to study resilience of taxi and subway trips for hurricanes sandy and irene. *Trans Res Record* 2599:70–80
- Zhan X, Ukkusuri SV, Zhu F (2014) Inferring urban land use using large-scale social media check-in data. *Netw Spatial Econ* 14(3):647–667. <https://doi.org/10.1007/s11067-014-9264-4>
- Zhang S, Wang W, Ford J, Makedon Fillia Learning from Incomplete Ratings Using Non-negative Matrix Factorization, pages 549–553

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Vaibhav Karve<sup>1</sup> · Derrek Yager<sup>2</sup> · Marzieh Abolhelm<sup>3</sup> · Daniel B. Work<sup>4</sup> · Richard B. Sowers<sup>5</sup> 

✉ Richard B. Sowers  
r-sowers@illinois.edu

Vaibhav Karve  
vkarve2@illinois.edu

Derrek Yager  
deryage@sandia.gov

Marzieh Abolhelm  
abolhel2@illinois.edu

Daniel B. Work  
dan.work@vanderbilt.edu

<sup>2</sup> Sandia National Laboratories, Albuquerque, NM 87123, USA

<sup>3</sup> Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA

<sup>4</sup> Civil and Environmental Engineering, Vanderbilt University, Nashville, TN 37235, USA

<sup>5</sup> Department of Industrial and Enterprise Systems Engineering, Department of Mathematics University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA

<sup>1</sup> Department of Mathematics, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA