Understanding day-to-day traffic patterns during disruption: an incremental nonnegative matrix factorization approach

Jingyan Li, Sebastian Brulin, Nithin Santhanam, Sean Qian

2026

Preprint:

This is an accepted article published in Transportation Research Board Annual Meeting. The final authenticated version is available online at:
https://doi.org/[DOI not available]

Understanding Day-to-Day Traffic Patterns During Disruption: An Incremental Nonnegative Matrix Factorization Approach

Carnegie Mellon University

Jingyan Li[1], Sebastian Brulin[2], Nithin Santhanam[3], Sean Qian[1]

[1]*Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA*

[2]*Honda Research Institute Europe, Germany*

[3]*Honda Research Institute USA, US*

Honda Research Institute

BACKGROUND

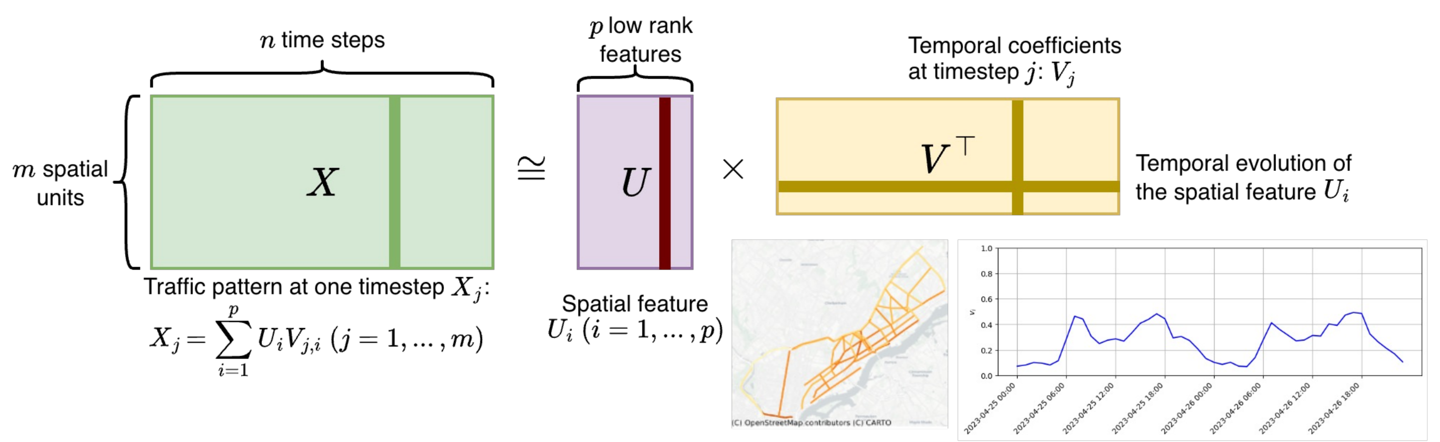
Traffic data is in low rank

- Traffic data, as high-dimensional spatiotemporal measurements, exhibit strong **temporal regularities** such as daily and weekly periodicity (morning/evening peaks, weekday/weekend differences), and **spatial correlations** among nearby or functionally similar links. That is, many links share a small number of typical "profiles" over time and these profiles are repeated across days.
- The traffic data matrix/tensor (e.g., sensor × time, or sensor × time-of-day × day) can be approximated by a low-dimensional subspace or **a low-rank tensor**, capturing most variance with relatively few components.
- Low-rank approximations drastically reduce computational and storage complexity for large-scale networks. Low-rank factors (e.g., basis patterns over time and space) provide interpretable building blocks—such as typical peak-period profiles or spatial congestion patterns—that support pattern discovery and domain insight.

Non-negative matrix factorization (NMF)

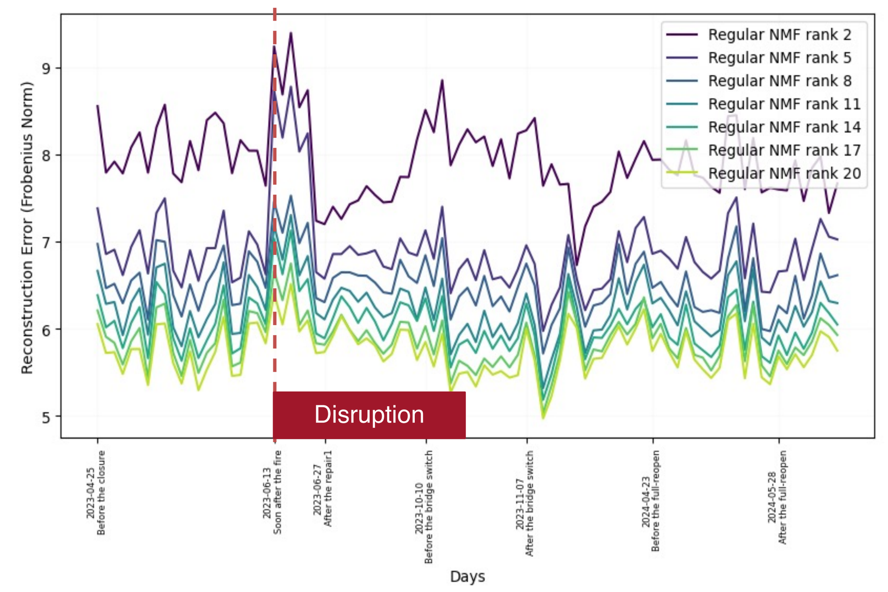
Let $X \in \mathbb{R}^{m \times n}$ be traffic observations for n timesteps in a m-link transportation network. NMF decomposes traffic records into two non-negative matrices that denote spatial features and their temporal evolution correspondingly:

$$X \cong UV^\top, \quad U \succcurlyeq 0, V \succcurlyeq 0$$



Disruptions introduce new patterns

- Disruptive events such as crashes, lane closures, and weather-related blockages generate non-recurrent congestion, with abrupt speed drops and capacity losses near the incident location and shockwaves that propagate upstream. Spatially, disruptions can shift demand to parallel routes and alternative modes, inducing atypical load patterns on normally less-congested links and intersections.



- Studies that factorize urban traffic data into low-rank components (normal patterns) plus sparse anomalies show that when abnormal days or events become more prevalent, more components are needed to approximate the data well.
- Thus, robust PCA or tensor decomposition methods used for anomaly detection cannot be used for system disruption cases: disruptions become region-wide and long-lasting, which violates the assumptions -- the "sparse" component is small. As atypical patterns become frequent, they start to look like additional regimes or bases rather than isolated outliers.

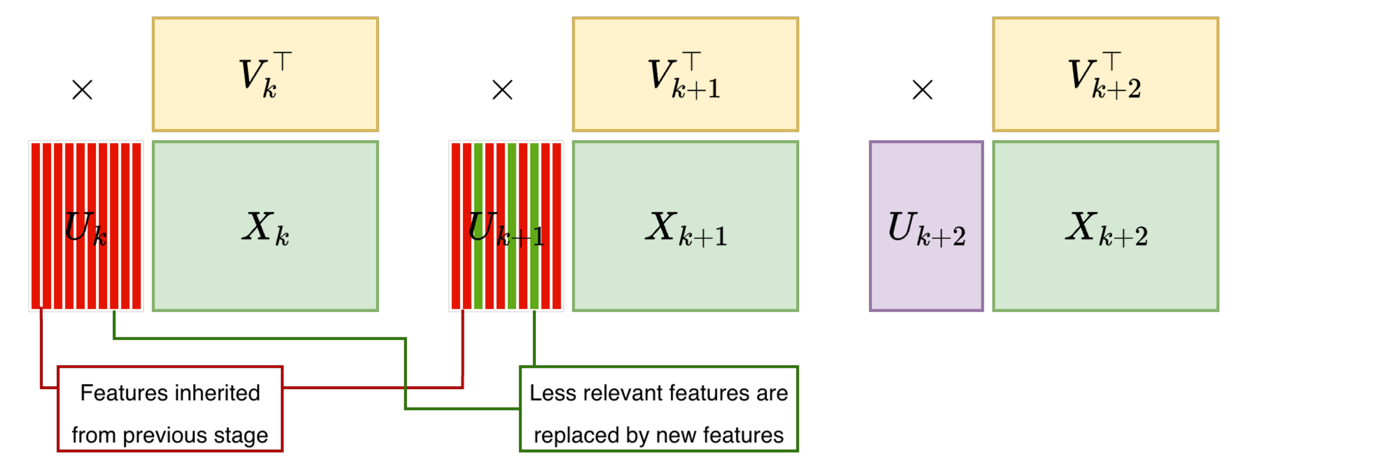
RESEARCH OBJECTIVES

- To understand *how* travel patterns adapt over time from real-world data
- To derive interpretable *latent* structure of day-to-day network dynamics that support pattern updating at disruption period
- To facilitate traffic status tracking, predicting and traffic operation after disruptions

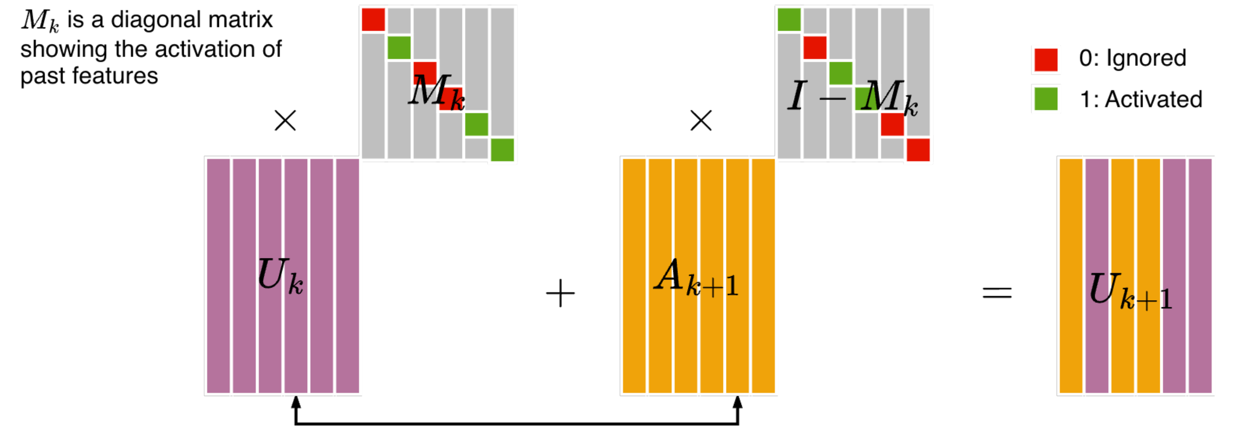
METHODOLOGY

Incremental NMF supporting day-to-day pattern discovery

Let $X_k \in \mathbb{R}^{m \times n_k}$ be the observations of stage k with n_k observation samples. Then, the next stage has observations $X_{k+1} \in \mathbb{R}^{m \times n_{k+1}}$.



Control the memory between stages: $U_k M_k + A_{k+1}(I - M_k) = U_{k+1}$



Spatial autocorrelation: two spatial units that are adjacent have higher similarity in their embeddings. Given the spatial feature matrix $U \in \mathbb{R}_{\geq 0}^{m \times p}$, the graph regularization is

$$\mathcal{R}_{\mathrm{g}}(U) = \frac{1}{2} \sum_{j,l=1}^{m} \|\mathbf{u}_j - \mathbf{u}_l\|^2 W_{jl} = \sum_{j=1}^{m} \mathbf{u}_j \mathbf{u}_j^\top D_{jj} - \sum_{j,l=1}^{m} \mathbf{u}_j \mathbf{u}_l^\top W_{jl}$$
$$= \mathrm{Tr}\left(U^\top L U\right)$$

Subregion activation: enforce the spatial feature to learn only a subregion of the network. Set a binary spatial mask vector $\mathbf{s}_i \in \{0,1\}^m$ for each spatial feature, where 1 denotes the corresponding road segment is allowed in the spatial feature. The matrix formulation is

$$U = S \odot \tilde{U}$$

Optimization problem

The problem can be solved by gradient descent or AO-ADMM algorithm.

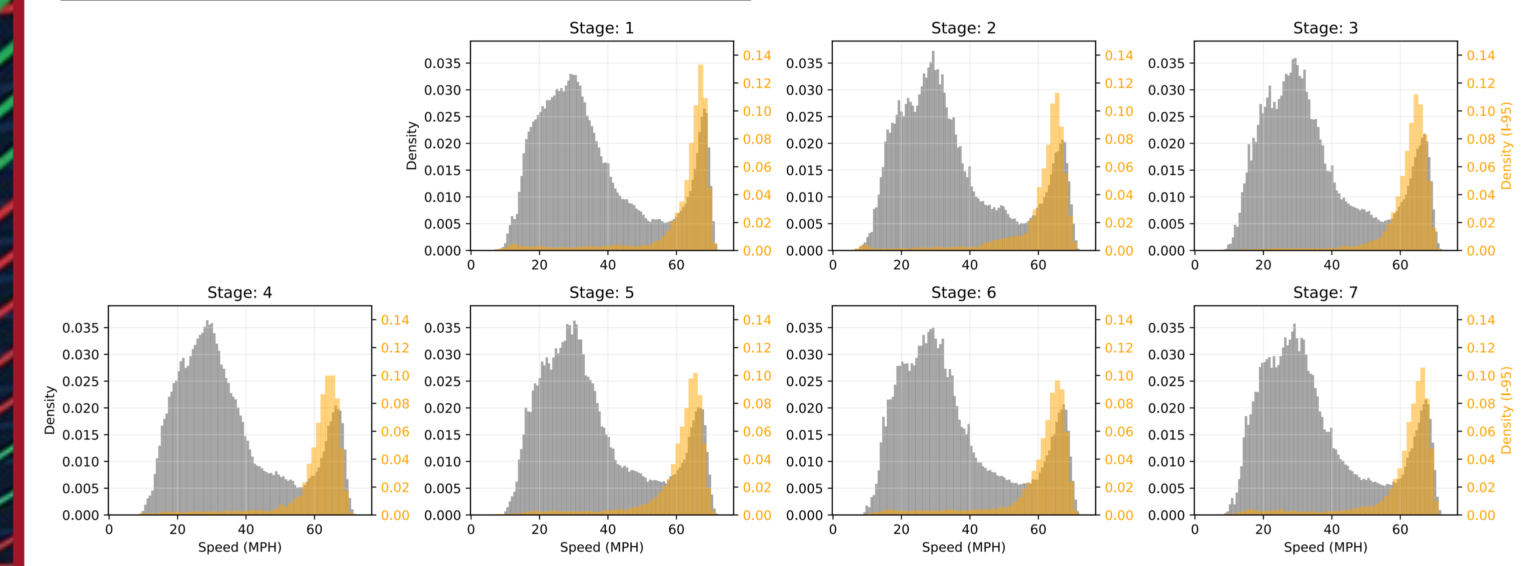
$$\begin{aligned} \min_{\tilde{U}_{k+1}, V_{k+1}, M_{k+1}, A_{k+1}} \quad & \left\| X_{k+1} - U_{k+1} V_{k+1}^\top \right\|_F^2 + \lambda \mathcal{R}_g \\ \text{subject to} \quad & \text{(C1) } U_{k+1} = S \odot \tilde{U}_{k+1}, \\ & \text{(C2) } \tilde{U}_{k+1} = U_k M_{k+1} + A_{k+1}(I - M_{k+1}), \\ & \text{(C3) } \operatorname{diag}\left(A_{k+1}^\top U_k\right) = 0, \\ & \text{(C4) } M_{k+1} \text{ is diagonal with } m_i \in \{0,1\} \text{ for all } i, \\ & \text{(C5) } \mathcal{R}_g = \mathrm{Tr}(\tilde{U}_{k+1}^\top L \tilde{U}_{k+1}), \\ & \text{(C6) } \tilde{U}_{k+1} \succeq 0, \quad V_{k+1} \succeq 0. \end{aligned} \tag{1}$$

CASE STUDIES

I-95 bridge collapse, June 11, 2023

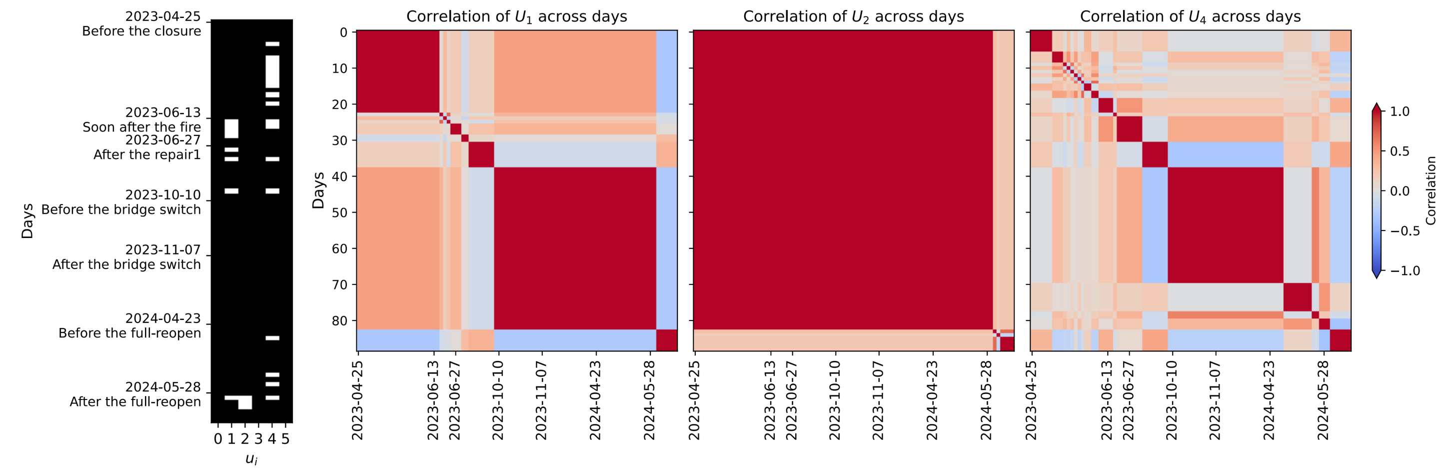
The fire damaged bridge structure and caused road full closure for two weeks. Major alternative corridors (i.e., U.S. 1, NJT) and local reroutes are recommended by local administration.

Stage	Start date	End date	Description
1	2023-04-25	2023-06-10	Before the closure
2	2023-06-11	2023-06-22	Soon after the fire
3	2023-06-24	2023-07-22	After the repair
4	2023-10-10	2023-11-02	Before the bridge switch
5	2023-11-07	2023-12-07	After the bridge switch
6	2024-04-23	2024-05-23	Before the full-reopen
7	2024-05-24	2024-06-22	After the full-reopen

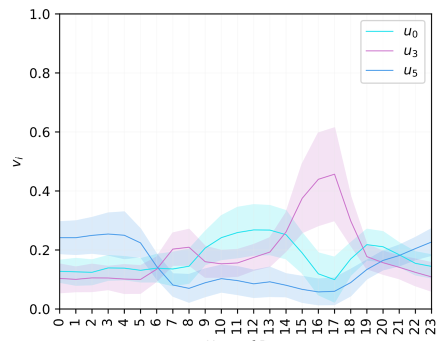


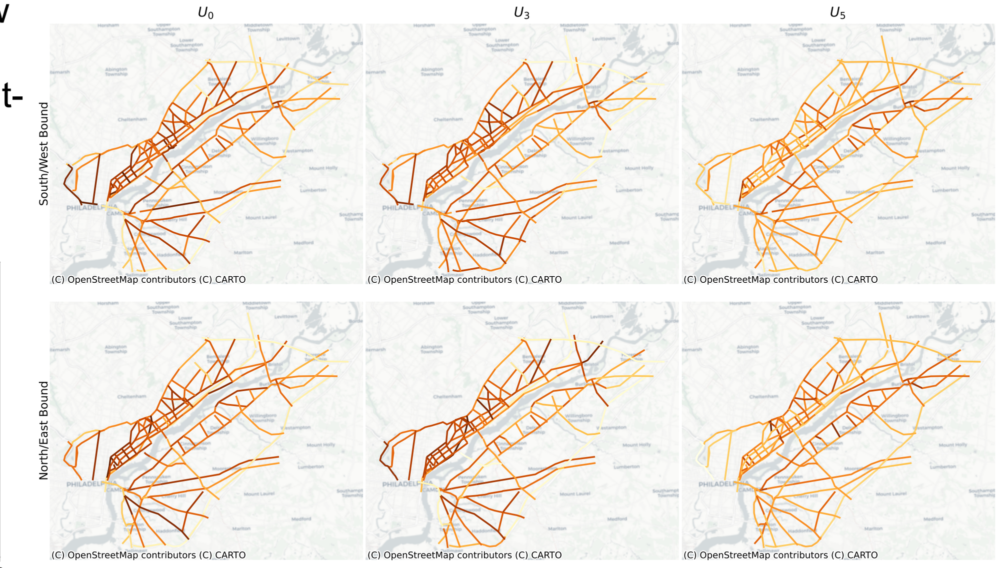
Q1: Are there any new patterns emerging and old patterns replaced during disruptions?

- Incremental NMF learns whether to update the spatial feature day to day.
- M matrix shows how the spatial features (u_i) are updated (white cell) or preserved (black cell)
- Pearson's correlation coefficient between the original spatial feature and its updated version suggests how "big" changes between spatial features.



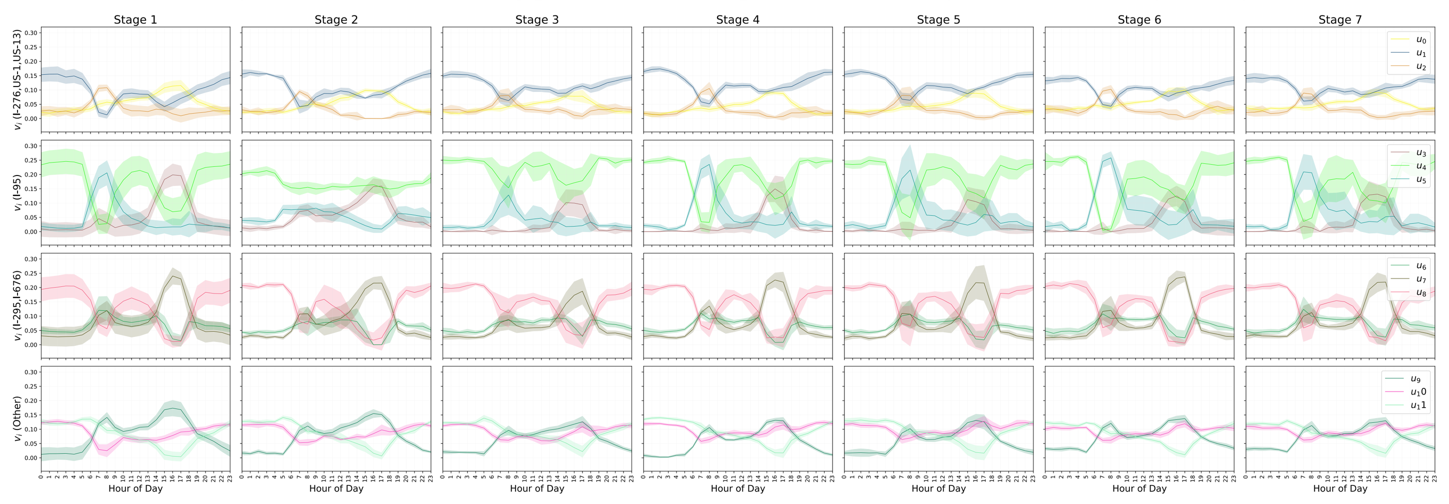
- New patterns show up soon after the fire, though with some similarities as before; the other spatial patterns (U_0, U_2, U_3, U_5) consistently exist.
- Completely new patterns show up after the full-reopen one year after the accident.
- The updating spatial features quickly converges to the new spatial pattern within 1 or 2 days after the disturbance.
- Unvaried features show morning / afternoon commuting peaks, night-time off-peak patterns, and day-time off-peak patterns





Q2: How are alternative routes impacted?

- Spatial hard mask on major rerouting corridors
- Different spatial features suggests morning / afternoon peak and off-peak patterns



Q3: Does the new equilibrium look alike the previous one?

- Afternoon peak on I-95 does not come back after full reopening
- Local roads do not show peak as sharp as before

CONCLUSION

- Provides **system-level understanding** of disruption propagation; derived from real-world data, offer guiding scenarios to calibration of transportation network modeling
- Scalable to multi-event, multi-city, multi-modal networks
- Support resilience assessment and disruption planning