Analytics on Sensor Networks

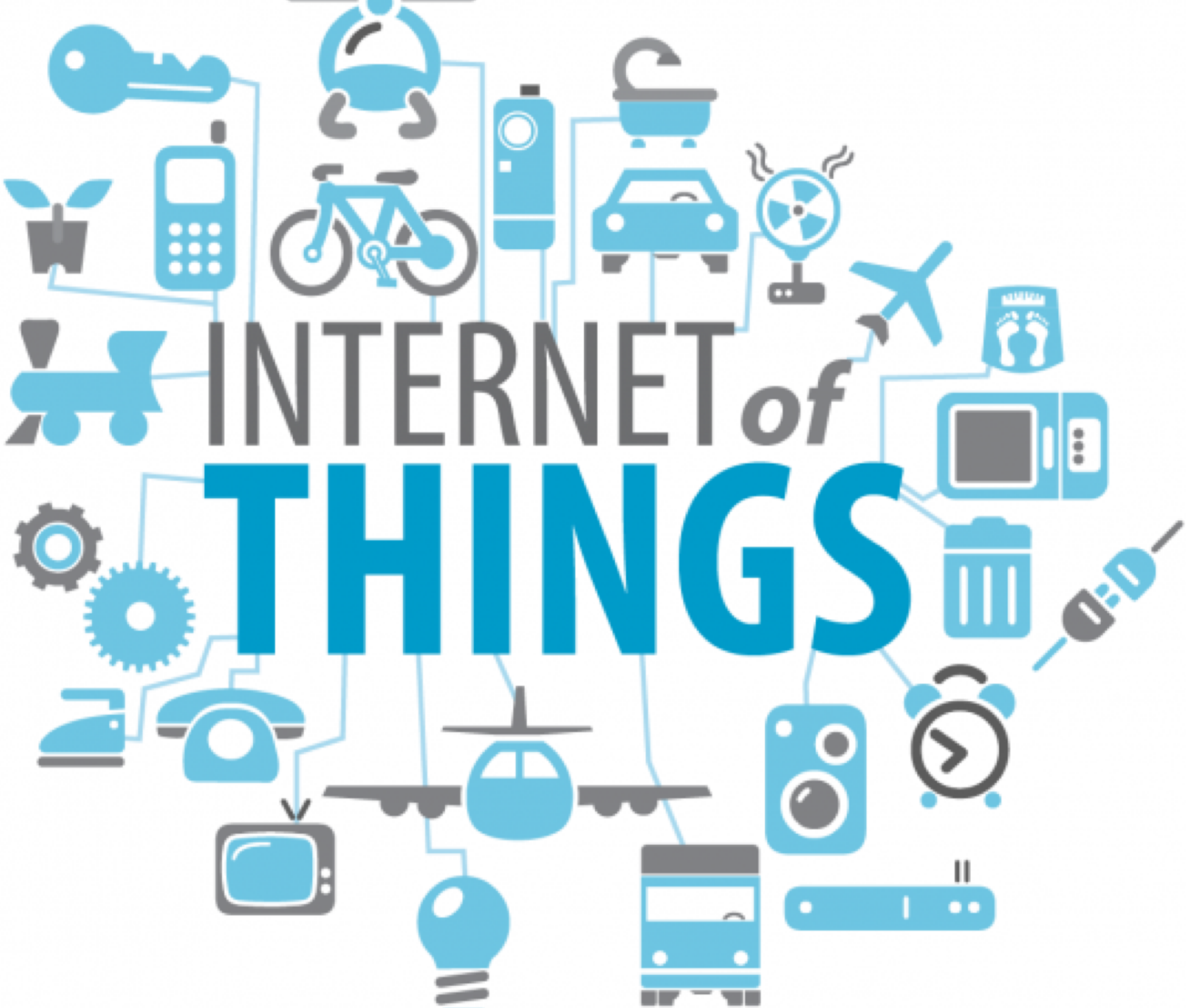
Joint work with **D. Hallac**, S. Vare, S. Bhooshan, R. Sosic, S. Boyd, and VW

Jure Leskovec

STANFORD INFOLAB

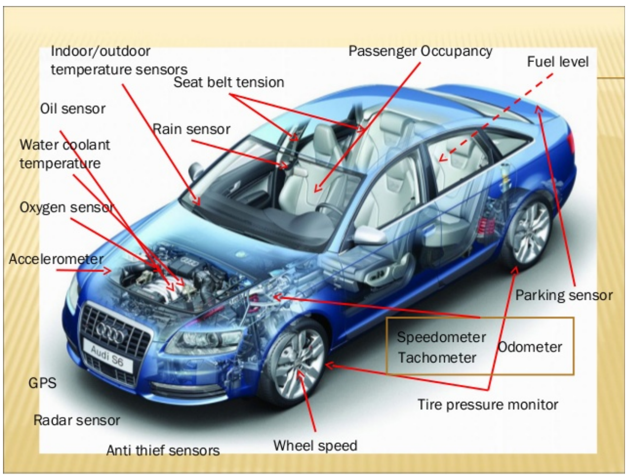
LELAND STANFORD JUNIOR UNIVERSITY ORGANIZED 1891

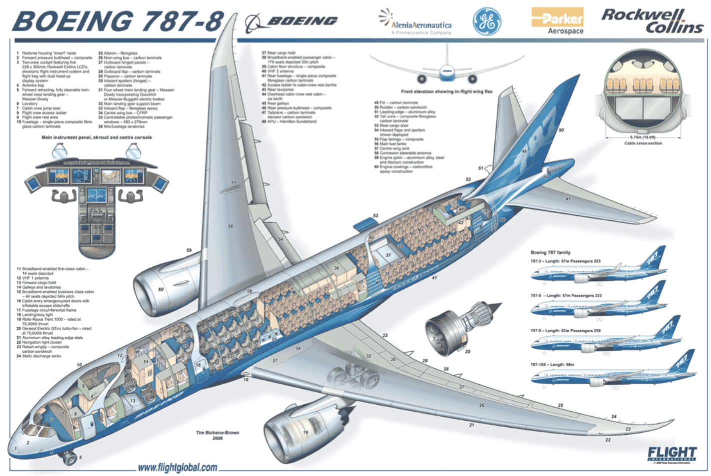
CHAN ZUCKERBERG BIOHUB


INTERNET of
THINGS

Sensors are Everywhere

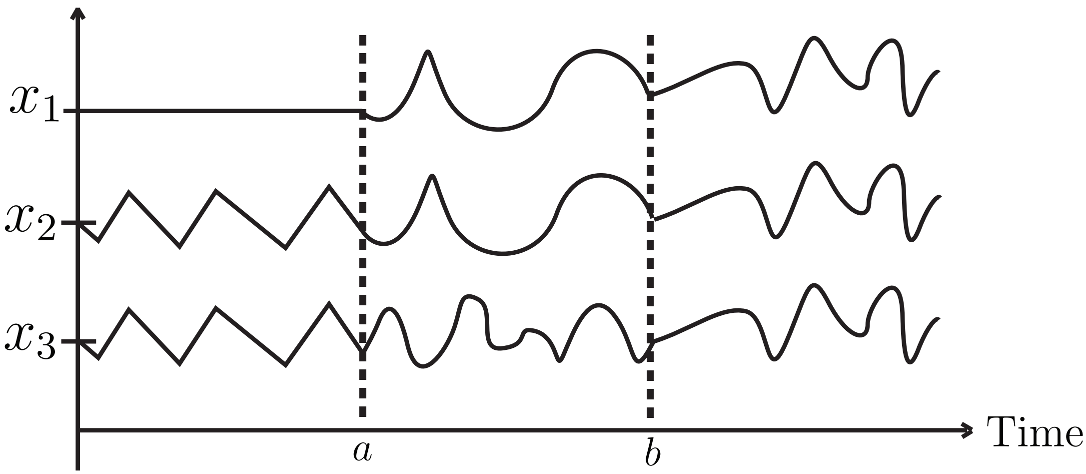
- Sequences of time stamped observations





Sensor Data: Time Series

- Sensors generate lots of time-series data



Challenges

- This data is
  - High-dimensional
  - Unlabeled
  - High-velocity
  - Dynamic
  - Heterogeneous

…But it Can be Very Valuable!

- Caterpillar shipping
- Discovered correlation between fuel usage and refrigerated containers
  - Realized that in certain regimes they needed to re-optimize their engine configuration parameters
- Saved $650,000+/year

Success Stories

- Pella Corporation
  - Large window and door manufacturing
- Owns 10 manufacturing plants
  - Large % of costs comes from energy bill
- Deployed sensor network across their plants
  - To monitor usage and provide real-time feedback to operators
- 16% decrease in energy costs!

Discovering Structure in the Data

- **Without proper methods, it is not possible to capitalize on the promise of “big data”**
- **Unsupervised learning** methods are needed to allow humans to interpret and act on these large datasets

How do we describe the structure of the time series so we can obtain insights and make predictions?

Key Questions

How to break down time series datasets into simple, interpretable components?

- …without pre-defining the structure, which leaves us open to biases!

How can we identify breakpoints, outliers, and labels for this time series data in a scalable way?

- Streaming settings increasingly common

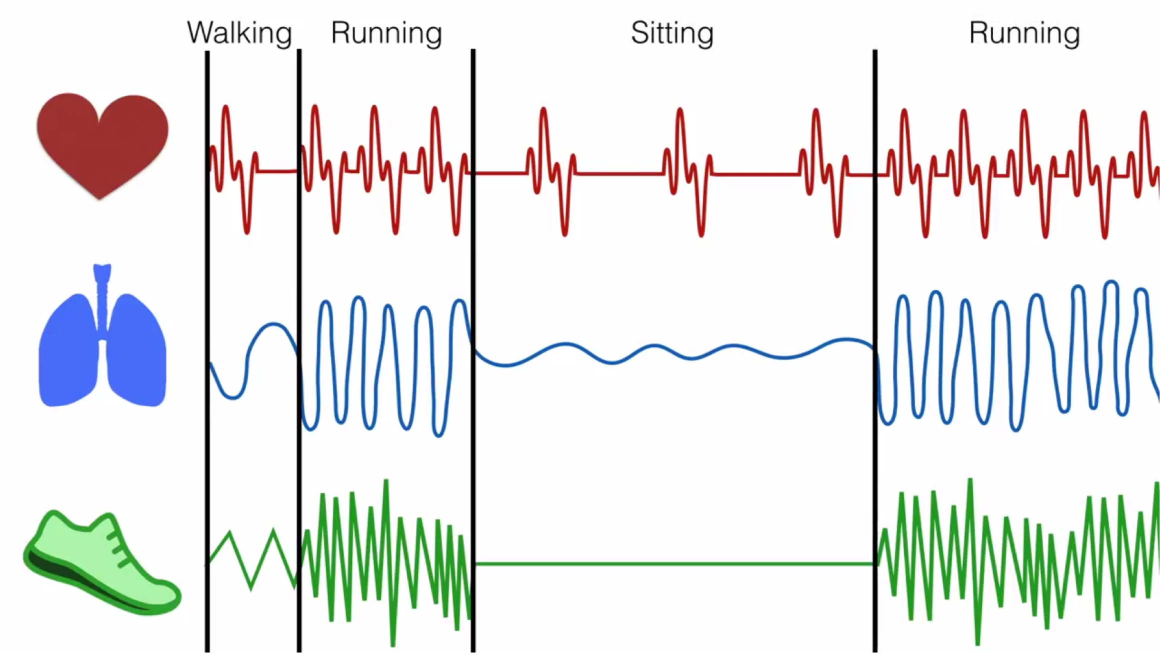
Today's Talk

- Toeplitz inverse covariance-based clustering (TICC)
- Drive2Vec
- Overview of future research directions in time series analysis
  - Deep learning
  - Open-source tools
  - Applications

Toeplitz Inverse Covariance-based Clustering (TICC)

Interpreting a Time Series

Value in "breaking down" the data into a sequence of states



Simultaneous Segmentation and Clustering

- **In general, these “states” are not predefined**
  - We do not know what they are, nor what they refer to…

- Instead, we need to discover these states in an **unsupervised** way!

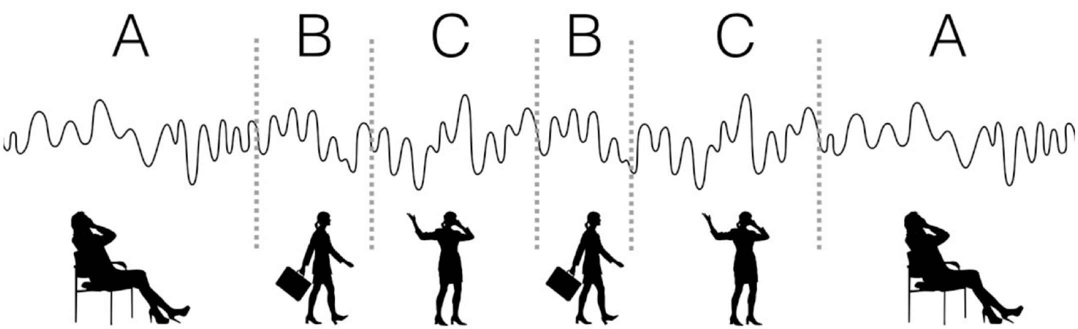
What is a Time Series?

- *T* sequential observations
  - $x_1, x_2, \ldots, x_T$
- Each observation x_i is *n*-dimensional
  - i.e., coming from *n* different sensors

- Observations can be synchronous or asynchronous
- There may be missing data
  - For example, if certain sensors are sampled at a higher rate than others

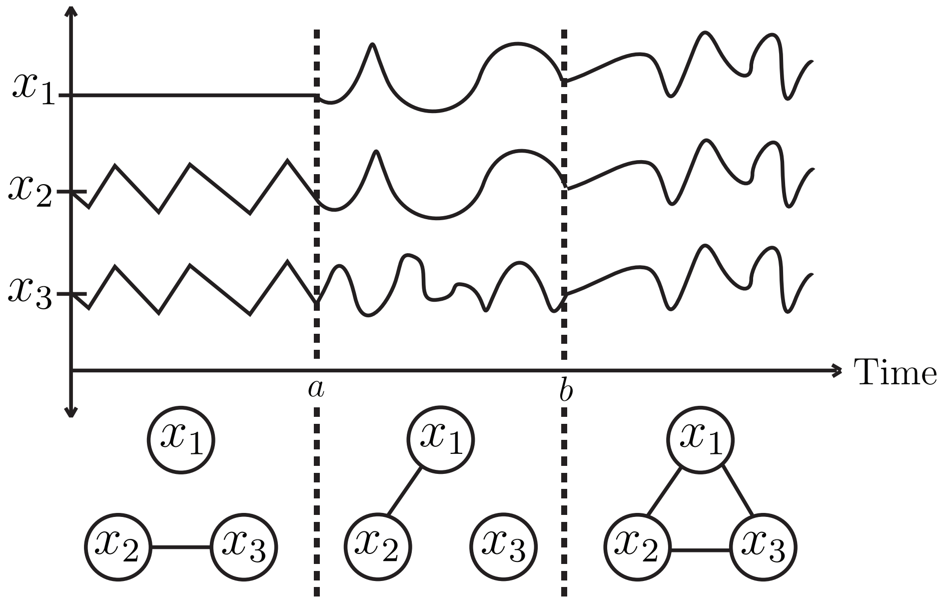
Goal

- <u>Given</u>: Multivariate time series
- <u>Goal:</u> Assign each point into one of *K* different states (or *clusters*), each defined by a simple "pattern"



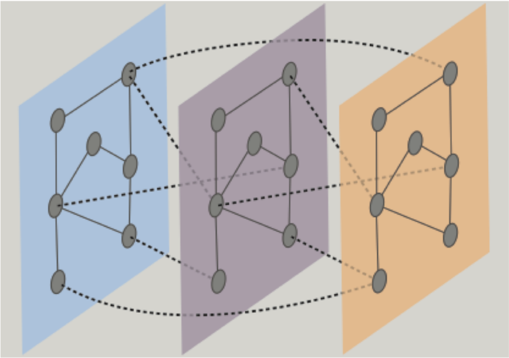
Definition of a Cluster

Convert a sequence of timestamped observations into a time-varying network



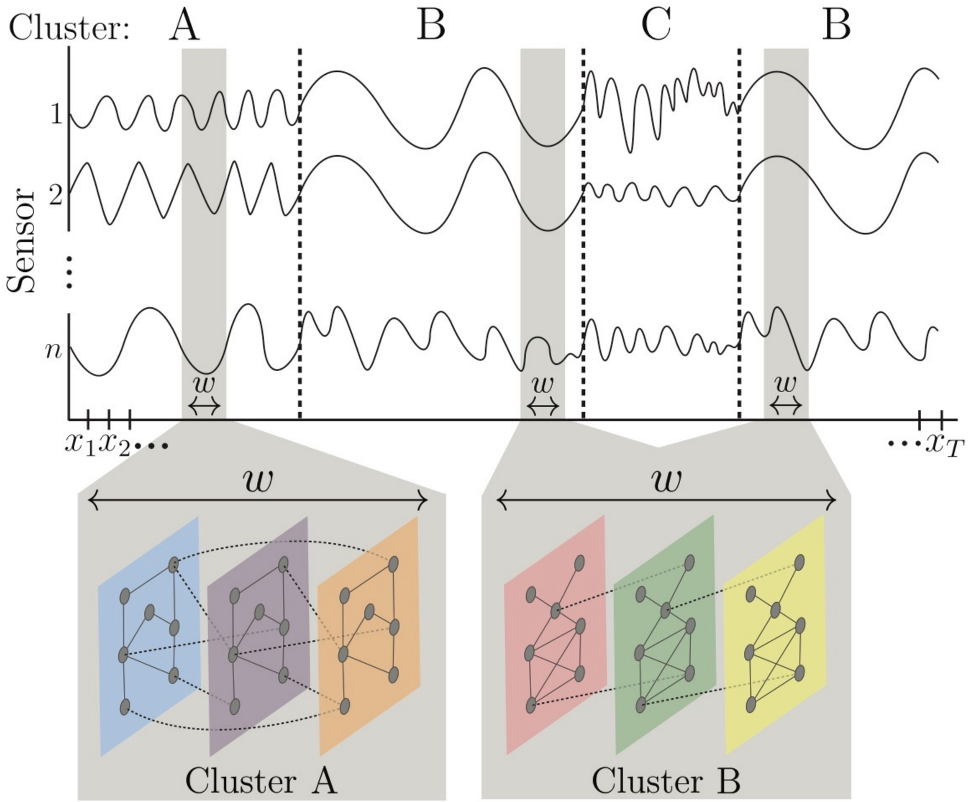
Definition of a Cluster

- Each cluster is defined by a multilayer correlation network, or a *Markov Random Field* (MRF)
  - Contains both intra-layer and inter-layer edges

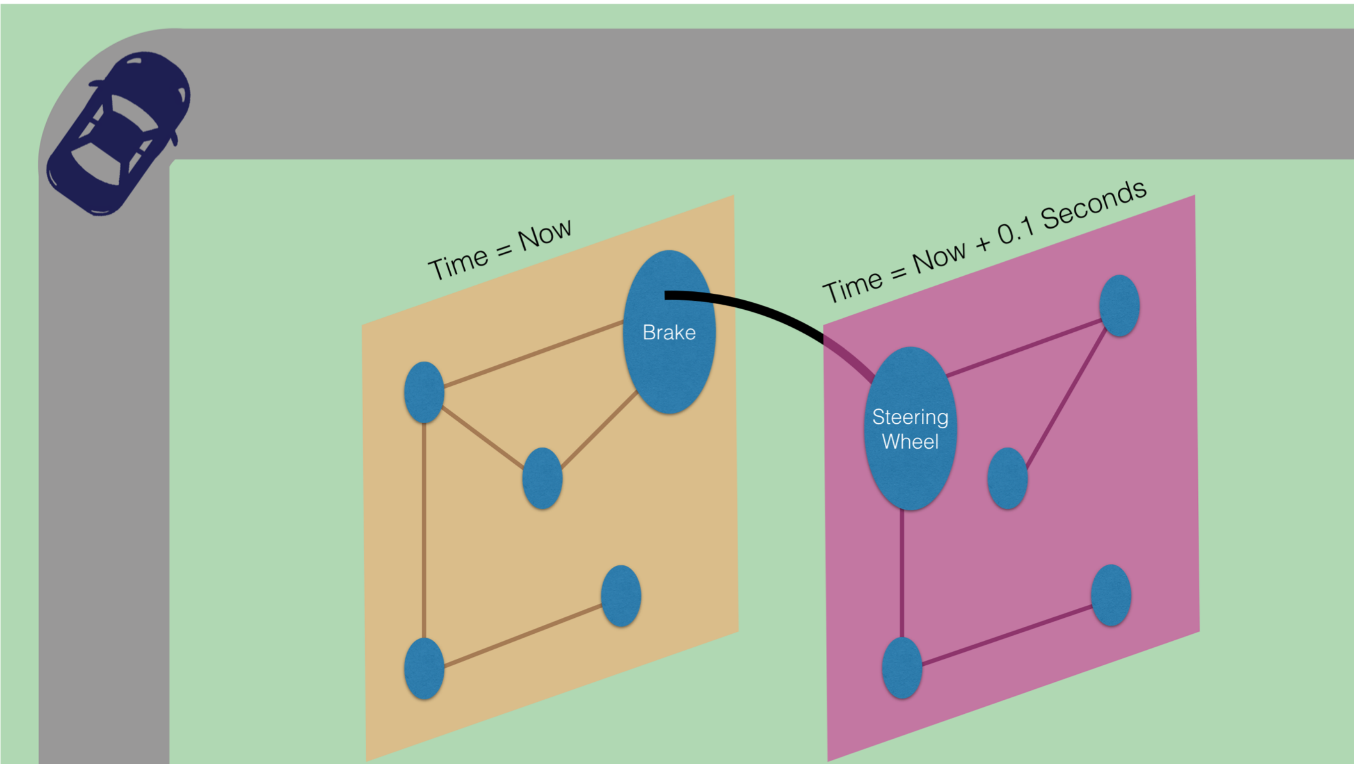


- MRFs encode **structural relationships** between the sensors

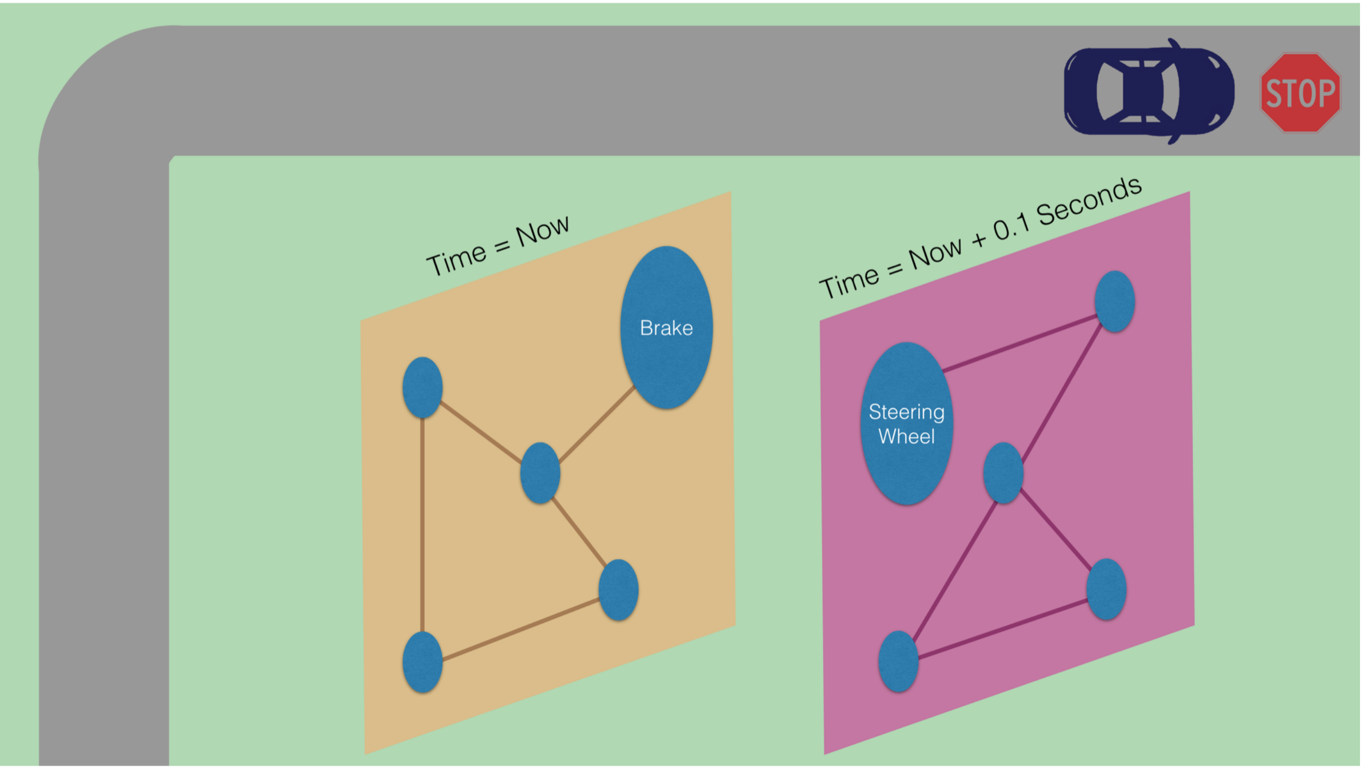
Example



Automobile – “Turning” State



Automobile – “Stopping” State



TICC Problem Setup

- Formal definition:

$$\underset{\Theta \in \mathcal{T}, \mathrm{P}}{\operatorname{argmin}} \sum_{i=1}^{K} \left[\overbrace{\|\lambda \circ \Theta_i\|_1}^{\text{sparsity}} + \sum_{X_t \in P_i} \left(\overbrace{-\ell\ell(X_t, \Theta_i)}^{\text{log likelihood}} + \overbrace{\beta \mathbb{1}\{X_{t-1} \notin P_i\}}^{\text{temporal consistency}} \right) \right]$$

where,

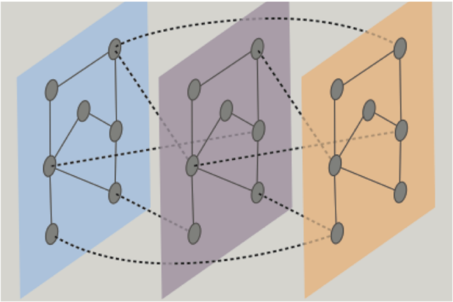
$$\ell\ell(X_t, \Theta_i) = -\frac{1}{2}(X_t - \mu_i)^T \Theta_i (X_t - \mu_i) + \frac{1}{2} \log \det \Theta_i - \frac{n}{2} \log(2\pi)$$

Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data. D. Hallac, S. Vare, S. Boyd, J. Leskovec. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017

Block Toeplitz Matrices

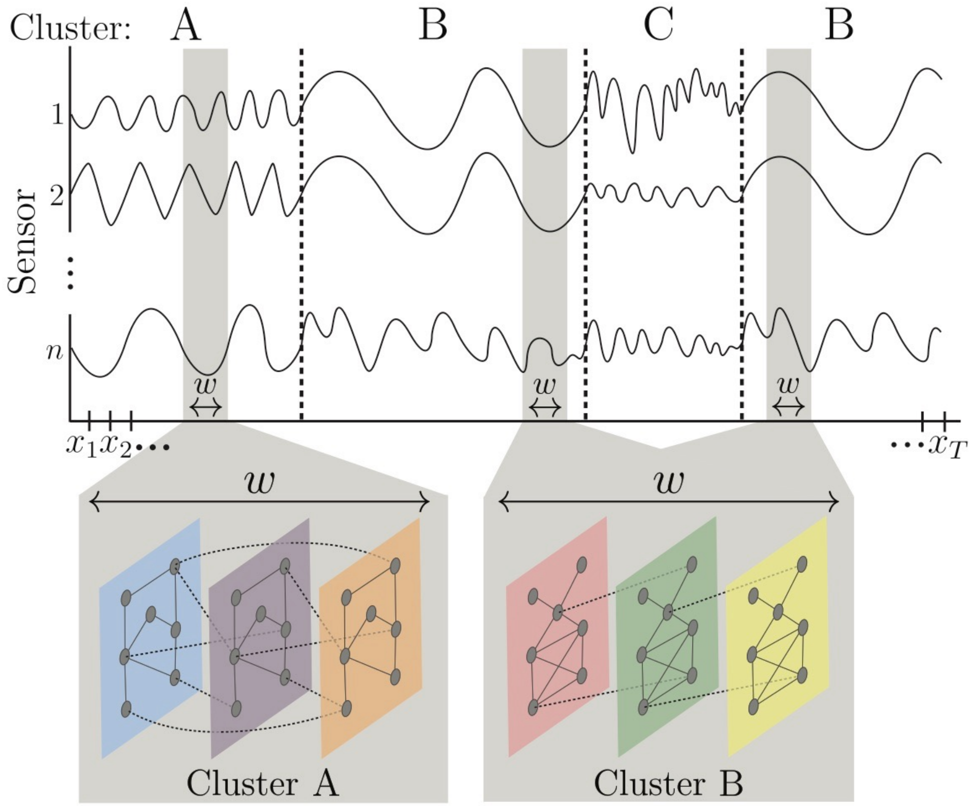
- Sparsity in the Toeplitz matrix defines the MRF edge structure

$$\Theta_i = \begin{bmatrix} A^{(0)} & (A^{(1)})^T & (A^{(2)})^T \\ A^{(1)} & A^{(0)} & (A^{(1)})^T \\ A^{(2)} & A^{(1)} & A^{(0)} \end{bmatrix}$$



- Toeplitz constraint enforces time invariance

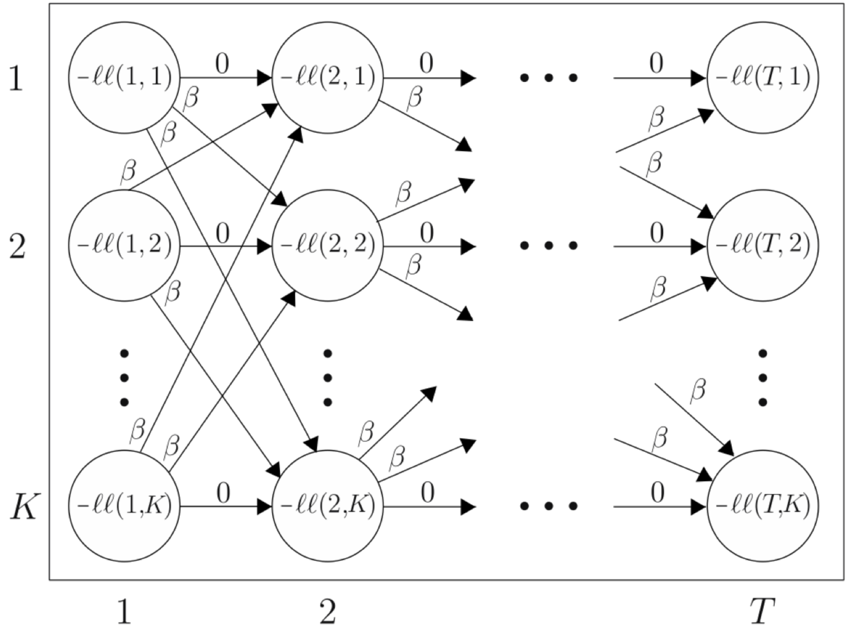
Running Example



Approach: EM

- TICC is highly non-convex
  - But we can use an EM-like approach to solve it!
- Alternate between…
  - Assigning points to clusters in a temporally consistent way
  - Updating the cluster parameters

Assigning Points to Clusters



We can solve this with dynamic programming!

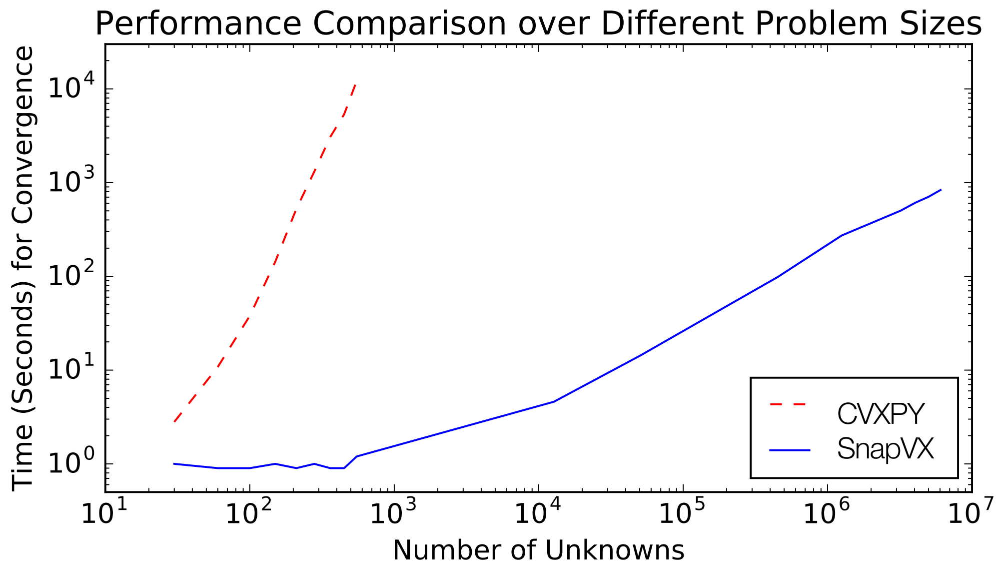
Updating Cluster Parameters

- Toeplitz Graphical Lasso:

$$\begin{array}{ll}\text{minimize} & -\log\det\Theta_i + \mathrm{tr}(S_i\Theta_i) + \frac{1}{|P_i|}\|\lambda \circ \Theta_i\|_1 \\ \text{subject to} & \Theta_i \in \mathcal{T}\end{array}$$

- We derive an ADMM solution (with closed-form proximal operators) to solve this problem efficiently

TICC: Scalability



- Can scale to problems with tens of millions of observations!

[SnapVX: A Network-Based Convex Optimization Solver](). D. Hallac, C. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd, J. Leskovec. Journal of Machine Learning Research (JMLR), 18(4):1−5, 2017.

How to Use TICC

- **Black box solver** that returns
  - Segmentation of the time series
  - Structural network defining each state
- Key parameter: Number of states
  - Statistical methods of choosing the optimal parameter value
- How to *understand* the results?

Case Study: Automobiles

- We analyzed 1 hour of driving data
  - 36,000 samples @ 10Hz
- We observed seven sensors
  - Brake pedal position
  - Forward (X-)acceleration
  - Lateral (Y-)acceleration
  - Steering wheel angle
  - Vehicle velocity
  - Engine RPM
  - Gas Pedal Position

Interpreting the Clusters

- We run TICC with K = 5 clusters and plot the betweenness centrality score of each node in each cluster

	Interpretation	Brake	X-Acc	Y-Acc	SW Angle	Vel	RPM	Gas
#1	Slowing Down	25.64	0	0	0	27.16	0	0
#2		0	4.24	66.01	17.56	0	5.13	135.1
#3		0	0	0	0	16.00	0	4.50
#4		0	0	0	0	32.2	0	26.8
#5		4.52	0	4.81	0	0	0	94.8

Interpreting the Clusters

- We run TICC with K = 5 clusters and plot the betweenness centrality score of each node in each cluster

	Interpretation	Brake	X-Acc	Y-Acc	SW Angle	Vel	RPM	Gas
#1	Slowing Down	25.64	0	0	0	27.16	0	0
#2	Turning	0	4.24	66.01	17.56	0	5.13	135.1
#3		0	0	0	0	16.00	0	4.50
#4		0	0	0	0	32.2	0	26.8
#5		4.52	0	4.81	0	0	0	94.8

Interpreting the Clusters

- We run TICC with K = 5 clusters and plot the betweenness centrality score of each node in each cluster

	Interpretation	Brake	X-Acc	Y-Acc	SW Angle	Vel	RPM	Gas
#1	Slowing Down	25.64	0	0	0	27.16	0	0
#2	Turning	0	4.24	66.01	17.56	0	5.13	135.1
#3	Speeding Up	0	0	0	0	16.00	0	4.50
#4		0	0	0	0	32.2	0	26.8
#5		4.52	0	4.81	0	0	0	94.8

Interpreting the Clusters

- We run TICC with K = 5 clusters and plot the betweenness centrality score of each node in each cluster

	Interpretation	Brake	X-Acc	Y-Acc	SW Angle	Vel	RPM	Gas
#1	Slowing Down	25.64	0	0	0	27.16	0	0
#2	Turning	0	4.24	66.01	17.56	0	5.13	135.1
#3	Speeding Up	0	0	0	0	16.00	0	4.50
#4	Driving Straight	0	0	0	0	32.2	0	26.8
#5		4.52	0	4.81	0	0	0	94.8

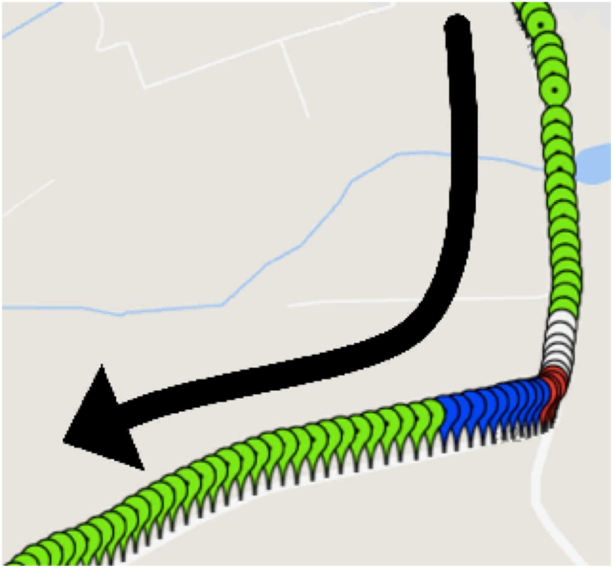
Interpreting the Clusters

- We run TICC with K = 5 clusters and plot the betweenness centrality score of each node in each cluster

	Interpretation	Brake	X-Acc	Y-Acc	SW Angle	Vel	RPM	Gas
#1	Slowing Down	25.64	0	0	0	27.16	0	0
#2	Turning	0	4.24	66.01	17.56	0	5.13	135.1
#3	Speeding Up	0	0	0	0	16.00	0	4.50
#4	Driving Straight	0	0	0	0	32.2	0	26.8
#5	Curvy Road	4.52	0	4.81	0	0	0	94.8

Plotting the Resulting Clusters

- Green = straight, white = slowing down, red = turning, blue = speeding up
- Results are very consistent across the data!



Implications

- Auto-labeling of data in an unsupervised way
  - Big cost for autonomous vehicles
- **Search engine** for discovering motifs in the time series
- Discover unique characteristics of individual drivers
- Can be used to identify more granular behaviors
  - Lane changes, near-accidents, etc.

Predicting the Future

(but without feature engineering)

Key Question

Can you aggregate all of car’s sensors and embed them into a single, low-dimensional **state**?

Our Approach

This state should be **predictive of both the short and long-term** future

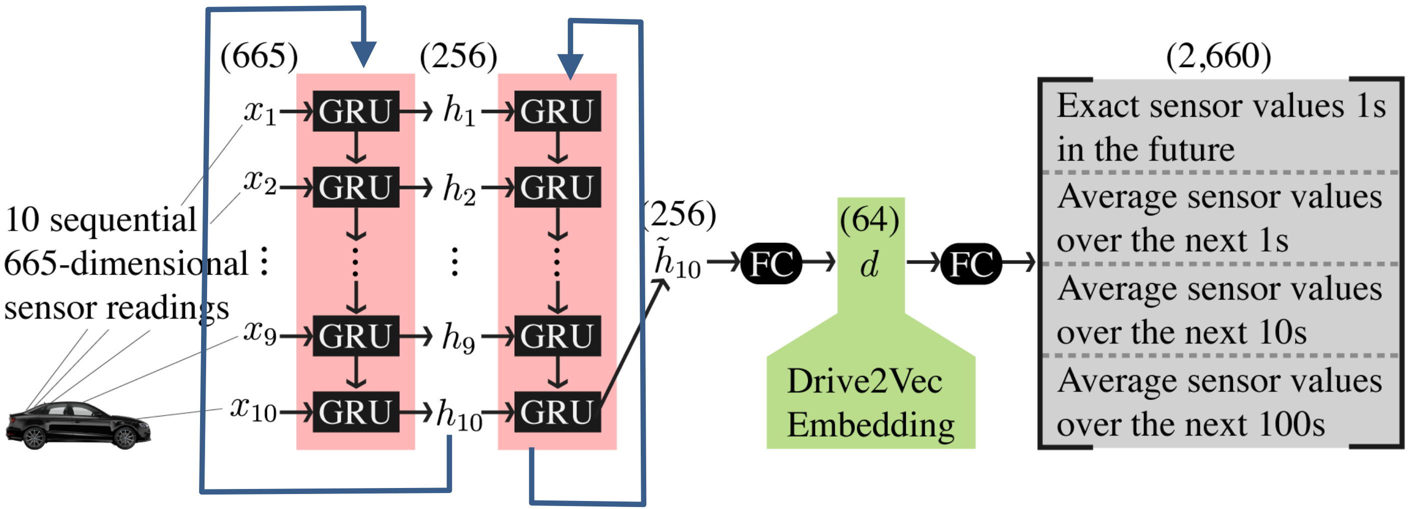
- *First order effects* – what the car is about to do
- *Second order effects* – the environment that the car is currently in (location, driver style, etc…)

Key Insight

Key insight: Attempt to predict the future at multiple granularities simultaneously:

- Combine multiple RNNs so they can learn at different levels of abstraction
- Learn to encode future at various time-scales

Drive2Vec Architecture



- Recurrent Neural network based on stacked Gated Recurrent Units (GRUs)

Problem Setup

- Dataset: Automobile data containing 1,400 sensors recording at 10 Hz.
- Goal: Predict driver actions 1 sec before they occur
  - Left/Right blinker
  - Accelerate (gas pedal > threshold)
  - Hard braking (brake pedal < threshold)

Driver Identification Using Automobile Sensor Data from a Single Turn. D. Hallac, A. Sharang, R. Stahlmann, A. Lamprecht, M. Huber, M. Roehder, R. Sosic, J. Leskovec IEEE International Conference on Intelligent Transportation Systems (ITSC), 2016.

Drive2Vec Goal

- **Given:** a 1 second window (10 samples) of 665-dimensional data
- **Goal:** Embed this data into a single 64-dimensional state that can be used to predict the short and long-term future of the car

Drive2Vec Experiments

- This single 64-dimensional embedding can:
  - A) Predict **exact** sensor values in short-term
  - B) Predict long-term **average** sensor values
  - C) Correctly identify driver (out of 29 potential drivers)
  - D) Be used as a **knowledge base** to identify potentially risky scenarios

Experimental Setup

- Train embeddings on 80% of the data to get mapping from raw data to the embedding
- Evaluate performance on a separate hold-out test set
  - All numbers are reported using the **same** 64-dimensional embedding

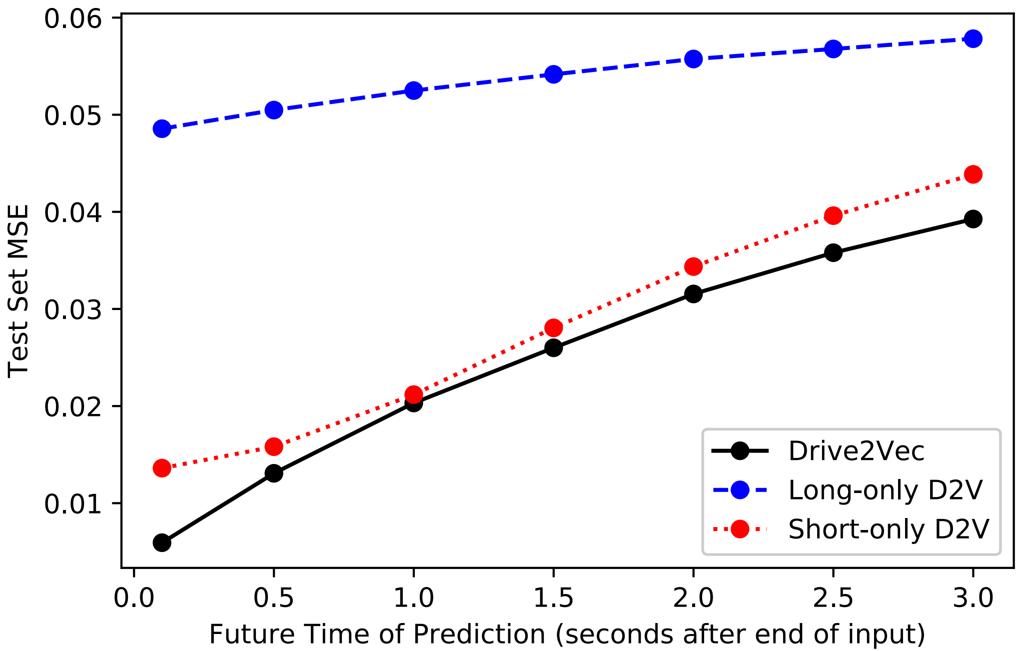
Experiment #1

- Short prediction: 64-dimensional embedding → **exact** 665 sensor values 1 second in the future
- Long prediction: 64-dimensional embedding → average 665 sensor values over the next 100 seconds

	Short prediction MSE	Long prediction MSE
Drive2Vec	0.020	0.021
Short-only D2V	0.021	0.027
Long-only D2V	0.052	0.021
PCA 64	0.174	0.036
Last timestep	0.204	0.069

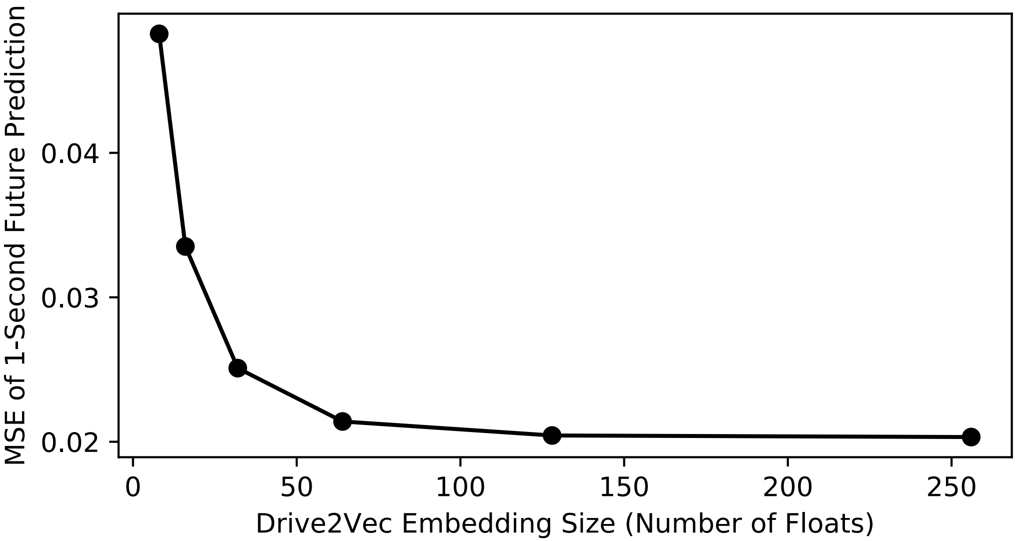
Experiment #2

- MSE vs. “time in future” of short-term prediction



Experiment #3

- MSE vs. Embedding size



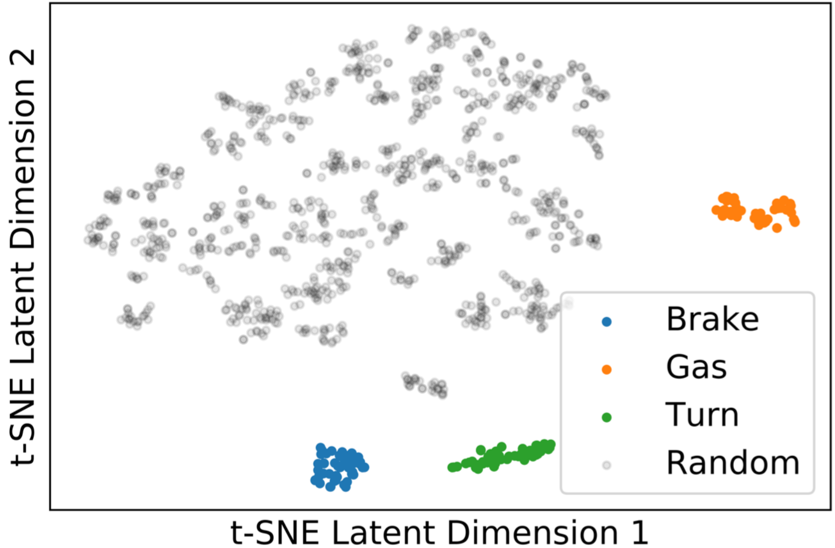
Experiment #4

- F1-score of 29-way driver identification task

Method	Micro F_1-score
Drive2Vec	0.642
Short-only D2V	0.593
Long-only D2V	0.741
PCA 64	0.577
Random	0.046

Case Study #1

- Different scenarios have extremely similar Drive2Vec embeddings!

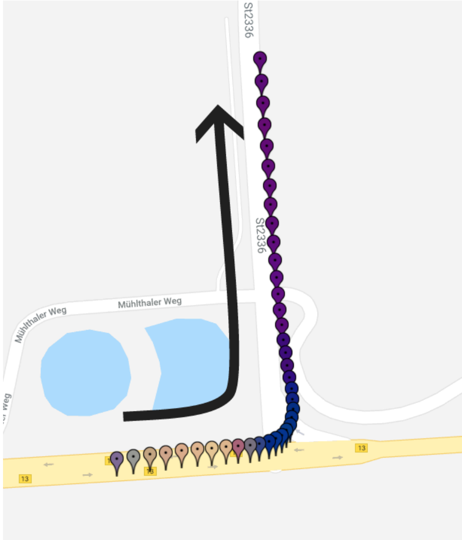


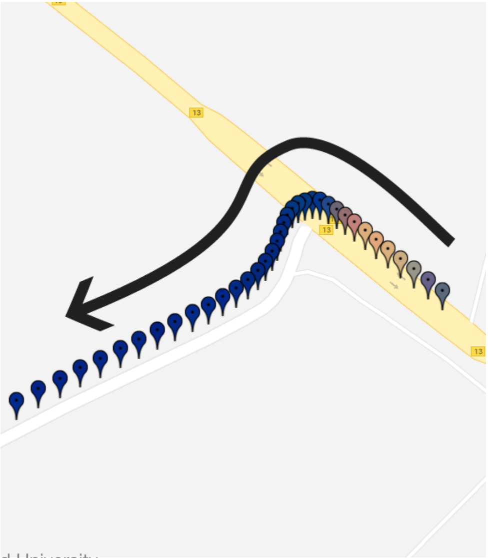
Case Studies #2

- We can identify risky scenario's **before** they occur
  - Predict 0.1s before a "brake slam"
- Similarity search returns AUC of 0.999983 compared to set of 8.5 million non-hard-brake scenarios

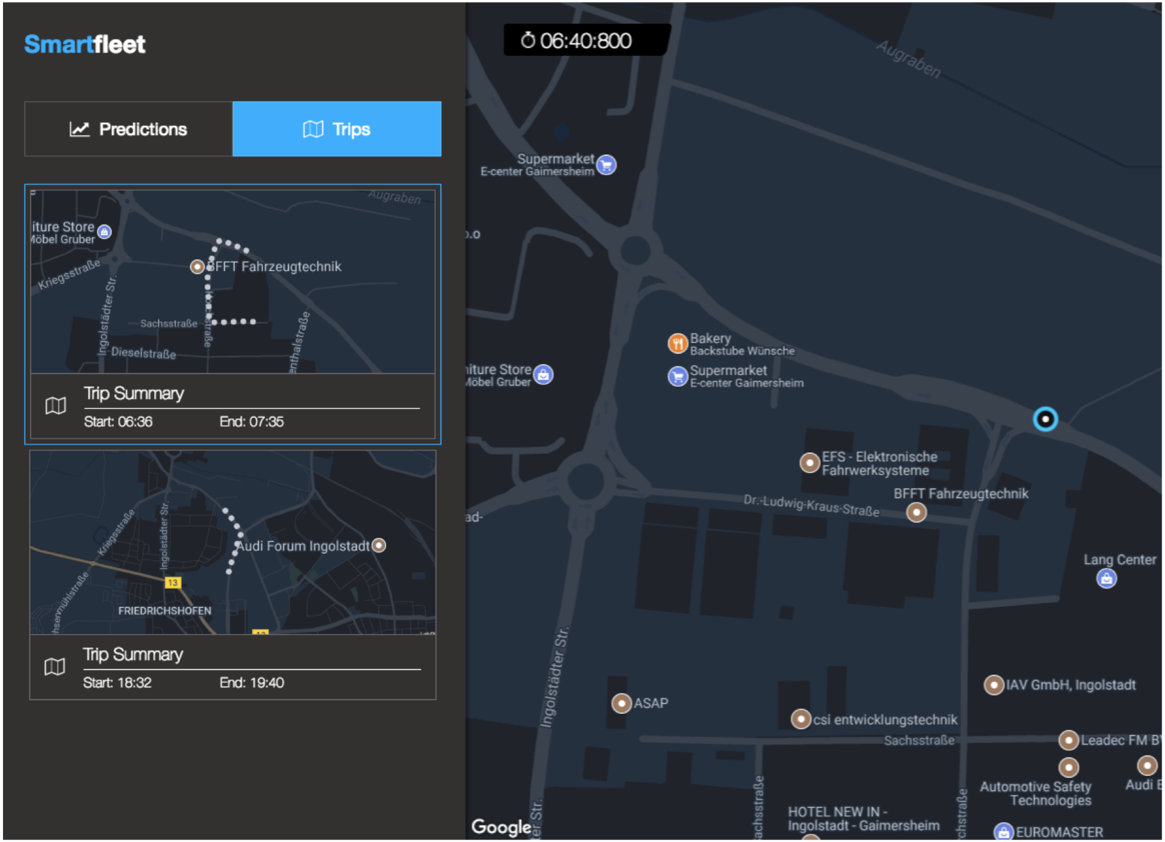
Case Study #3

- Temporal evolutions of embeddings
  - Large shocks occur from highway to rural (both short + long expected values change)

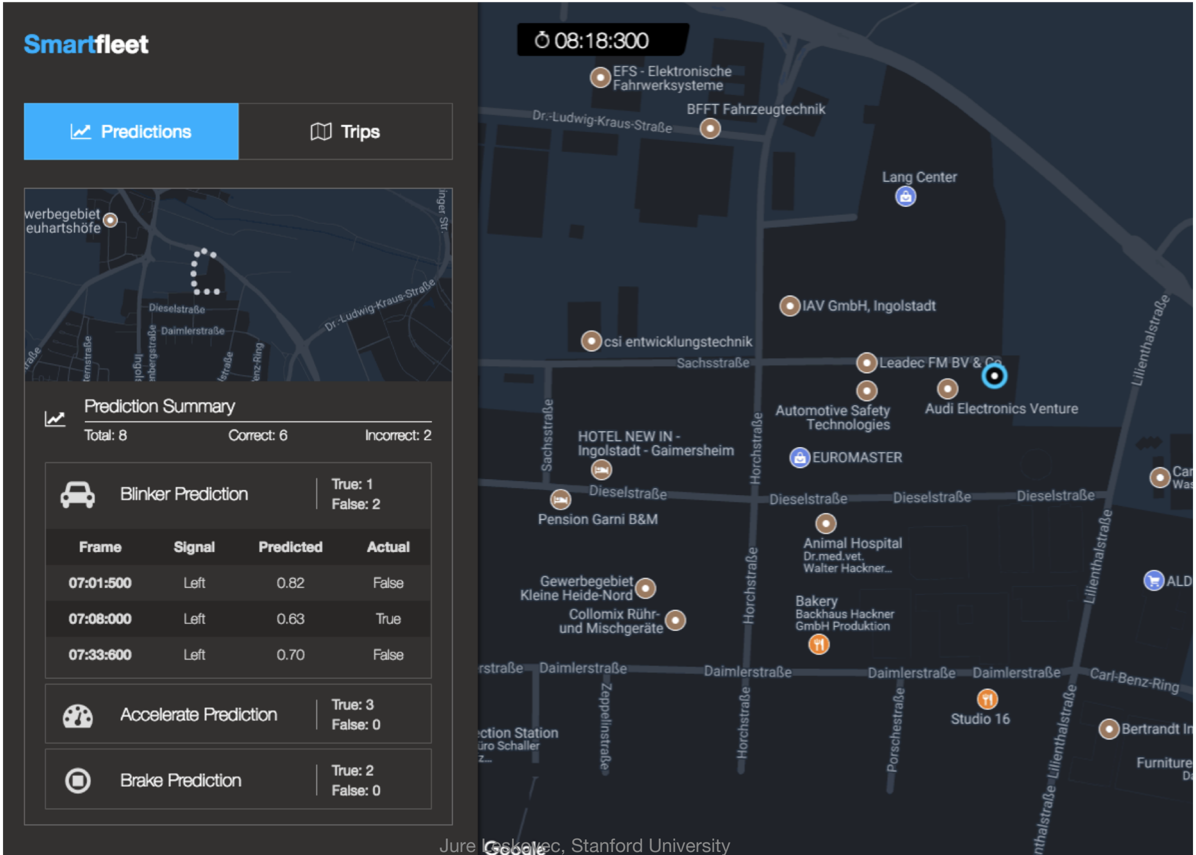




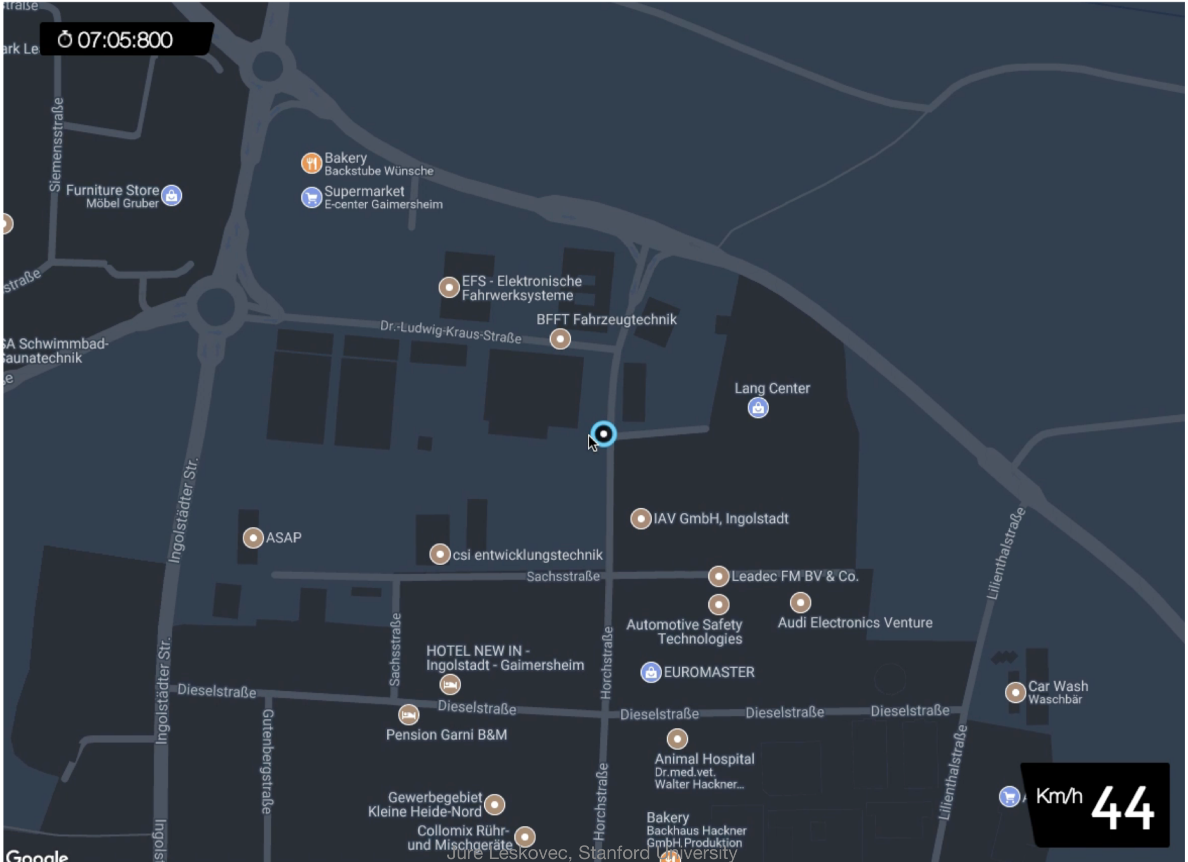
Predicting Driver Actions



Predicting Driver Actions



Predicting Driver Actions

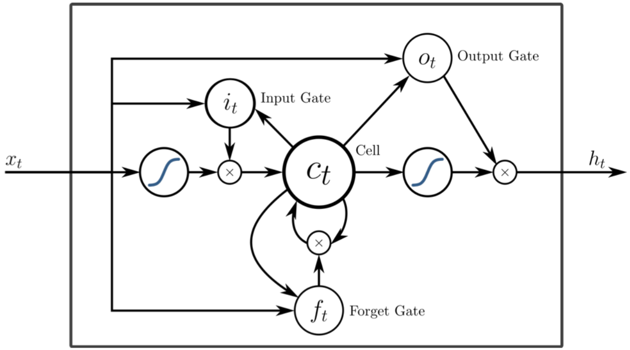


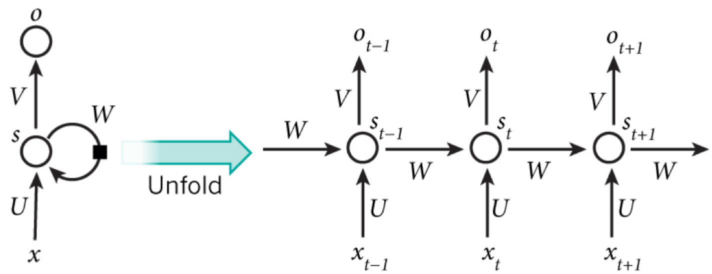
Jure Leskovec, Stanford University

The Future of Time Series Research

Deep Learning

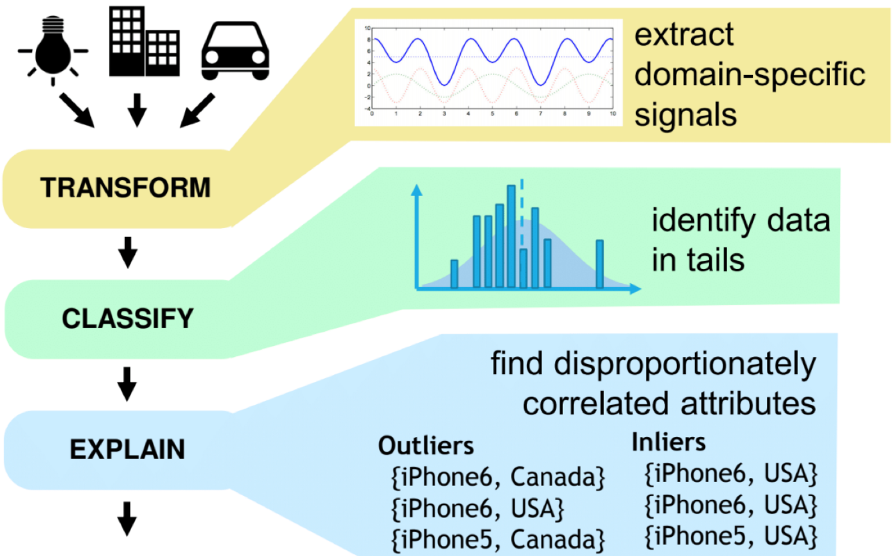
- Long short-term memory (LSTMs)
  - Type of recurrent neural network (RNN)
- Becoming a increasingly powerful method of forecasting/classification on time series
  - However, results are less interpretable





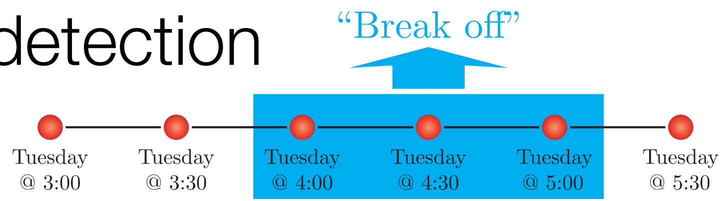
Stanford Project: MacroBase

- Analytics engine that **prioritizes user attention** by combining outlier detection and high-dimensional feature selection routines at scale

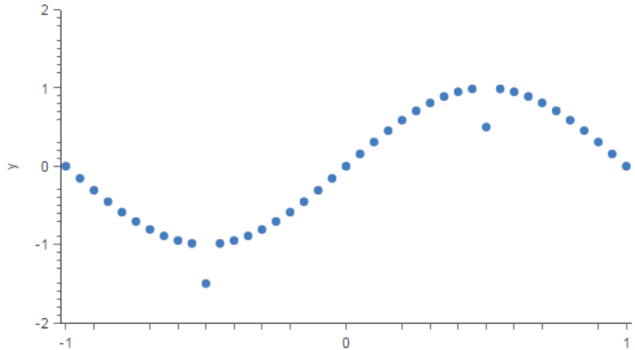


Applications

- Event/anomaly detection



- Important to have principled math/statistics background
  - Not everything is this clean…



Applications

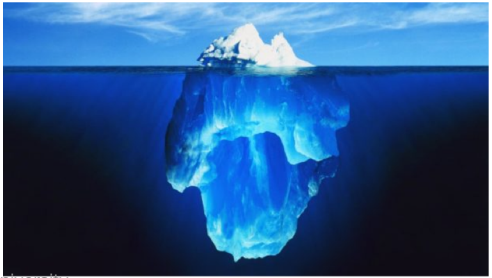
- Predictive maintenance
  - What if you can predict failures before they occur?
    - Potentially huge cost/safety benefits

Applications

- User modeling (personalization)
  - Bridging the gap between online and offline

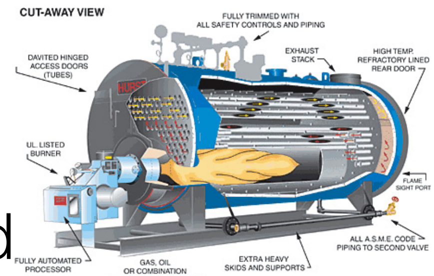
Analyzing Sensor Data

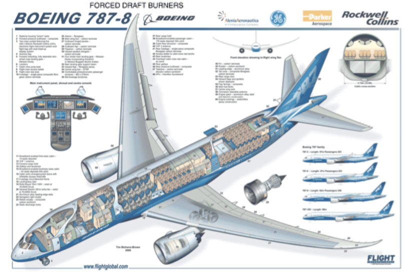
- Lots of exciting research directions
- More and more applications by the day
  - Bringing innovations from the online world to the real world
- However, new and improved methods are required to keep innovating
  - Interpreting and acting on sensor data in an unsupervised way
- We're only at the tip of the iceberg!

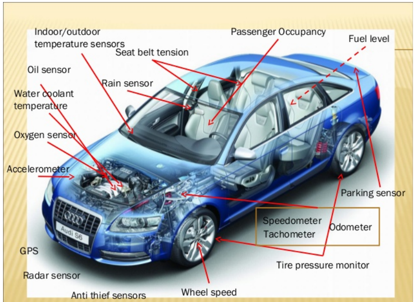


Conclusion

- Complex engineered systems
  - High-dimensional unlabeled time series data collected in real-time
- We need tools to understand these data as well as to make accurate predictions







Industry Partnerships

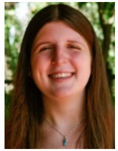
Ford, azumio, Siemens, Huawei, twitter, Pinterest, Under Armour, facebook, Hyundai, Boeing, VW, Wikipedia, Bosch Invented for life, intel, NTT docomo, spinn3r, JD.COM, HITACHI, Observe, Viaduct

PhD Students

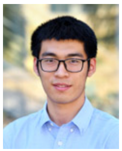

Alexandra Porter


Camilo Ruiz


Claire Donnat

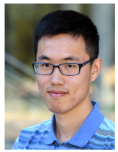

Emma Pierson


Weihua Hu


Jiaxuan You


Bowen Liu


Mohit Tiwari


Rex Ying

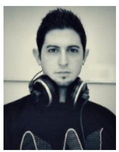
Funding

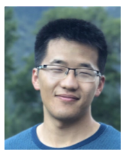
NIH, IARPA, MURI ARO, DARPA, NSF, Office of Naval Research ONR Science & Technology, Chan Zuckerberg Initiative

Post-Doctoral Fellows


Baharan Mirzasoleiman


Marinka Zitnik


Michele Catasta

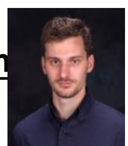

Pan Li


Shantao Li


Srijan Kumar


Hingwei Wang

Research Staff


Adrijan Bradaschia


Rok Sosic

Collaborators

Dan Jurafsky, Linguistics, Stanford University
David Grusky, Sociology, Stanford University
Stephen Boyd, Electrical Engineering, Stanford University
David Gleich, Computer Science, Purdue University
VS Subrahmanian, Computer Science, University of Maryland
Sarah Kunz, Medicine, Harvard University
Russ Altman, Medicine, Stanford University
Jochen Profit, Medicine, Stanford University
Eric Horvitz, Microsoft Research
Jon Kleinberg, Computer Science, Cornell University
Sendhill Mullainathan, Economics, Harvard University
Scott Delp, Bioengineering, Stanford University
James Zou, Medicine, Stanford University

STANFORD INFOLAB

References

- Drive2Vec: Multiscale State-Space Embedding of Vehicular Sensor Data. D. Hallac, S. Bhooshan, M. Chen, K. Abida, R. Sosic, J. Leskovec. *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2018.
- Data-Driven Model Predictive Control of Autonomous Mobility-on-Demand Systems. R. Iglesias, F. Rossi, K. Wang, D. Hallac, J. Leskovec, M. Pavone. *International Conference on Robotics and Automation (ICRA)*, 2018.
- Network Inference via the Time-Varying Graphical Lasso. D. Hallac, Y. Park, S. Boyd, J. Leskovec.ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017.
- Toeplitz Inverse Covariance-Based Clustering of Multivariate Time Series Data. D. Hallac, S. Vare, S. Boyd, J. Leskovec. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017.
- Learning the Network Structure of Heterogeneous Data via Pairwise Exponential Markov Random Fields. Y. Park, D. Hallac, S. Boyd, J. Leskovec. Artificial Intelligence and Statistics Conference (AISTATS), 2017.
- SnapVX: A Network-Based Convex Optimization Solver. D. Hallac, C. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd, J. Leskovec. Journal of Machine Learning Research (JMLR), 18(4):1–5, 2017.
- Driver Identification Using Automobile Sensor Data from a Single Turn. D. Hallac, A. Sharang, R. Stahlmann, A. Lamprecht, M. Huber, M. Roehder, R. Sosic, J. Leskovec IEEE International Conference on Intelligent Transportation Systems (ITSC), 2016.
- Network Lasso: Clustering and Optimization in Large Graphs. D. Hallac, J. Leskovec, S. Boyd. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2015.