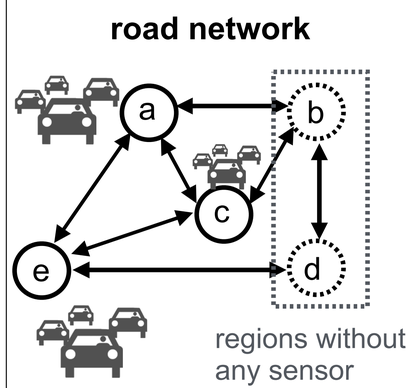
Collective Traffic Prediction with Partially Observed Traffic History using Location-Based Social Media

Xinyue Liu, Xiangnan Kong and Yanhua Li
Worcester Polytechnic Institute

1. Traffic Prediction with Partially Observed History



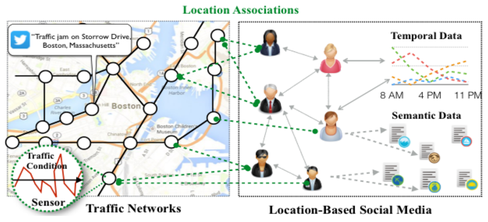
1. Problem Studied: Traffic Prediction with Partially Observed Traffic History.

2. Traffic Prediction: Infer the traffic conditions in the future time span for a geographical area.

3. Partially Observed History: In the target area, some locations are not deployed with any sensors, their historical traffic conditions are not available.

4. The Goal: Make good predictions for every location in the target area based on the partially observed traffic history.

2. Location-Based Social Media (LBSM)



LBSM Used: Twitter

Table 1: Average # of tweets in each region under different spatiotemporal resolutions in our dataset.

Temporal Resolution	Spatial Resolution	Ave. #Tweets
12 hours	1×1	47,113
1 hour	1×1	3,926
1 hour	2×2	1,306
1 hour	3×3	554
1 hour	4×4	389
1 hour	30×30	15

Target Area: Greater Los Angeles

Why LBSM ?

1. Cover much wider range of geographic areas.

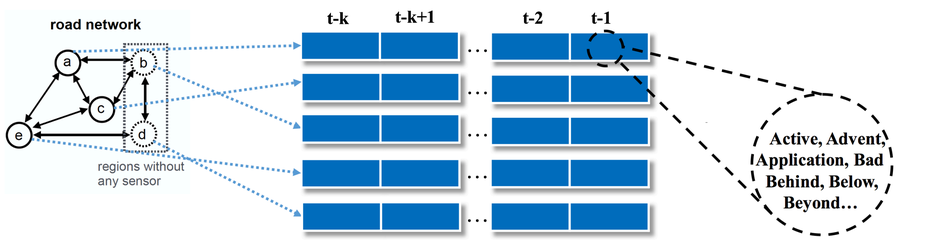
2. Provide abundant information about the road users in real-time.

3. Dictation Systems (e.g. Siri) in smart phones or smart cars allow road user to post contents in LBSM easily.

4. By mining the semantic and spatial information from LBSM, we can effectively infer the future traffic conditions for many areas, including the road segments without sensors.

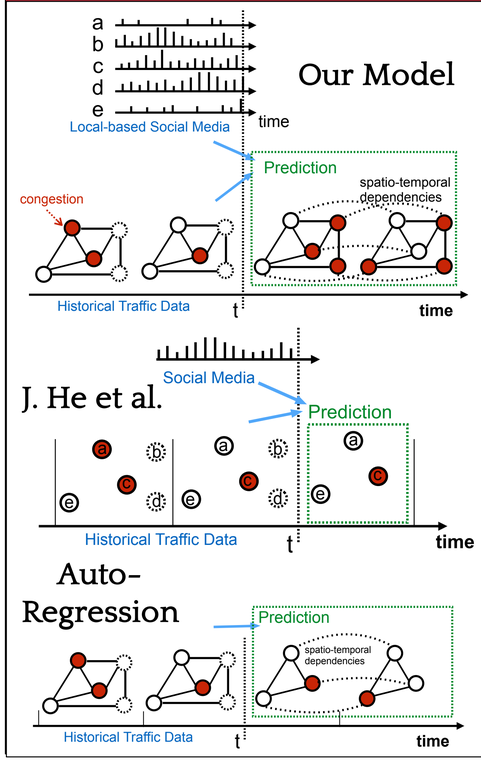
Challenge 0: How to incorporate LBSM semantics into traffic prediction?
Challenge 1: Lack of Traffic History Information for Some Locations.
Challenge 2: Sparsity of LBSM Information at Fine Granularities (Table 1).

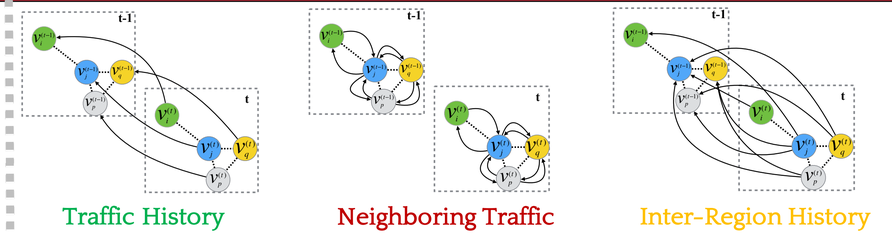
Bag-of-words Representation

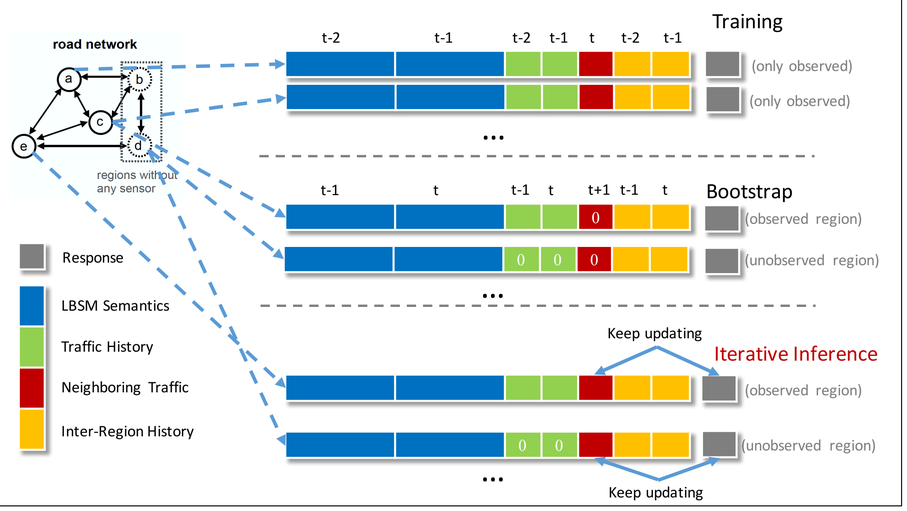


Stemmed Words

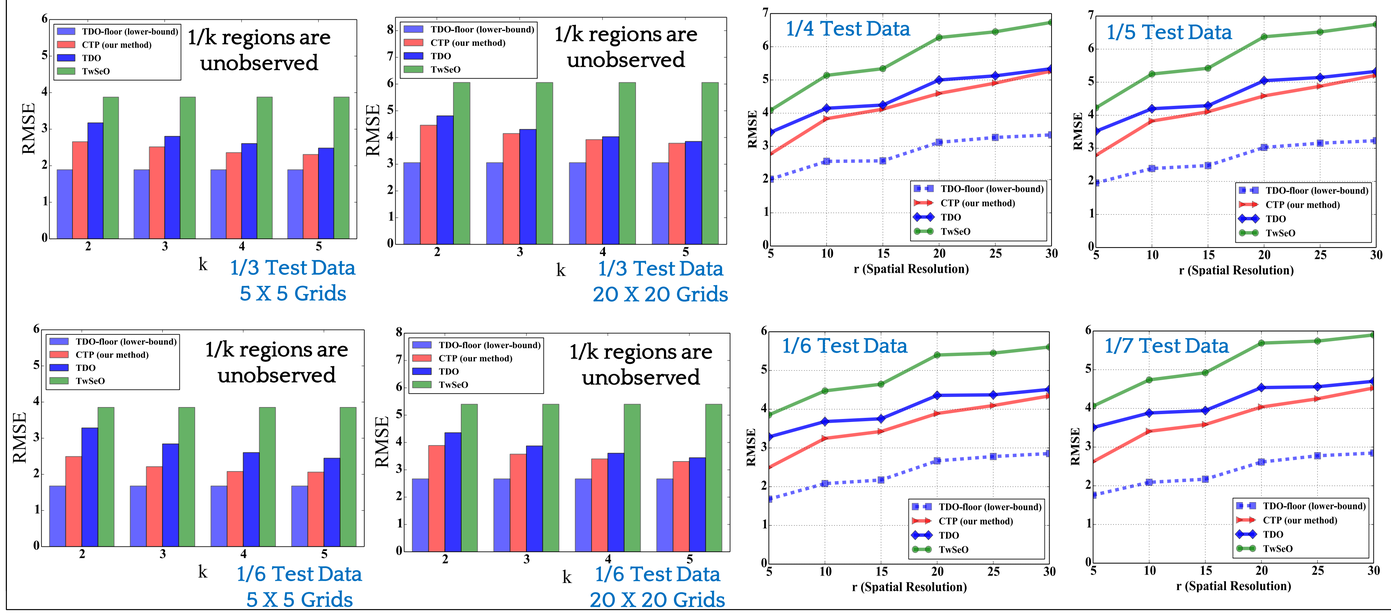
3. Collective Inference Framework







4. Experiment Results



5. Acknowledgement

This work is supported in part by the National Science Foundation through grant CNS-1626236.