D²FAS (Section 5): We prove that D²FAS can scale better than existing state-of-the-art centralized algorithms with a large number of observations and sensors;

- Theoretically guaranteeing the predictive performance of the decentralized data fusion component of D²FAS to be equivalent to that of a sophisticated centralized sparse approximation for the GP model (Section 3): The computation of such a sparse approximate GP model can thus be parallelized and distributed among the mobile sensors (in a Google-like MapReduce paradigm), thereby achieving efficient and scalable prediction;

- Theoretically guaranteeing the performance of the partially decentralized active sensing component of D²FAS, from which various practical environmental conditions can be established to improve its performance;

- Developing a relational GP model whose correlation structure can exploit both the road segment features and road network topology information (Section 2.1);

- Empirically evaluating the predictive performance, time efficiency, and scalability of the D²FAS algorithm on a real-world traffic phenomenon (i.e., speeds of road segments) dataset over an urban road network (Section 6): D²FAS is more time-efficient and scales significantly better with increasing number of observations and sensors while achieving predictive performance close to that of existing state-of-the-art centralized algorithms.

2 Relational Gaussian Process Regression

The Gaussian process (GP) can be used to model a spatiotemporal traffic phenomenon over a road network as follows: The traffic phenomenon is defined to vary as a realization of a GP. Let \( V \) be a set of road segments representing the domain of the road network such that each road segment \( s \in V \) is specified by a \( p \)-dimensional vector of features and is associated with a realized (random) measurement \( z_s \) \((Z_s)\) of the traffic condition such as speed if \( s \) is observed (unobserved). Let \( \{Z_s\}_{s \in V} \) denote a GP, that is, every finite subset of \( \{Z_s\}_{s \in V} \) follows a multivariate Gaussian distribution (Rasmussen and Williams, 2006). Then, the GP is fully specified by its prior mean \( \mu_s \triangleq \mathbb{E}[Z_s] \) and covariance \( \sigma_{ss'} \triangleq \text{cov}[Z_s, Z_{s'}] \) for all \( s, s' \in V \). In particular, we will describe in Section 2.1 how the covariance \( \sigma_{ss'} \) for modeling the correlation of measurements between all pairs of segments \( s, s' \in V \) can be designed to exploit the road segment features and the road network topology.

A chief capability of the GP model is that of performing probabilistic regression: Given a set \( D \subset V \) of observed road segments and a column vector \( z_D \) of corresponding measurements, the joint distribution of the measurements at any set \( Y \subset V \setminus D \) of unobserved road segments remains Gaussian with the following posterior mean vector and covariance matrix

\[
\mu_{Y|D} \triangleq \mu_Y + \Sigma_{YD} \Sigma_{DD}^{-1}(z_D - \mu_D) \tag{1}
\]
\[
\Sigma_{YY|D} \triangleq \Sigma_{YY} - \Sigma_{YD} \Sigma_{DD}^{-1} \Sigma_{DY} \tag{2}
\]

where \( \mu_Y \) \((\mu_D)\) is a column vector with mean components \( \mu_s \) for all \( s \in Y \) \((s \in D)\), \( \Sigma_{YD} \) \((\Sigma_{DD}) \) is a covariance matrix with covariance components \( \sigma_{ss'} \) for all \( s, s' \in D \) \((s, s' \in D)\), and \( \Sigma_{DY} \) is the transpose of \( \Sigma_{YD} \). The posterior mean vector \( \mu_{Y|D} \) is used to predict the measurements at any set \( Y \) of unobserved road segments. The posterior covariance matrix \( \Sigma_{YY|D} \) is defined to exploit both the features and topology information, which will be described next.

### Definition 1 (Road Network)

Let the road network be represented as a weighted directed graph \( G \triangleq (V, E, m) \) that consists of

- a set \( V \) of vertices denoting the domain of all possible road segments,
- a set \( E \subseteq V \times V \) of edges where there is an edge \((s, s')\) from \( s \in V \) to \( s' \in V \) if the end of segment \( s \) connects to the start of segment \( s' \) in the road network, and
- a weight function \( m : E \to \mathbb{R}^+ \) measuring the standardized Manhattan distance \((Borg and Groenen, 2005)\)

\[
m((s, s')) \triangleq \sum_{i=1}^{p} \|[[s]]_i - [[s']_i]\|/r_i \text{ of each edge } (s, s') \text{ where } [[s]]_i \text{ (}[[s']_i]\text{)} \text{ is the } i\text{-th component of the feature vector specifying road segment } s \text{ (}s'\text{)}, \text{ and } r_i \text{ is the range of the } i\text{-th feature. The weight function } m \text{ serves as a dissimilarity measure between adjacent road segments.}

The next step is to compute the shortest path distance \(d(s, s')\) between all pairs of road segments \( s, s' \in V \) (i.e., using Floyd-Warshall or Johnson’s algorithm) with respect to the topology of the weighted directed graph \( G \). Such a distance function is again a measure of dissimilarity, rather than one of similarity, as required by a kernel function. Fur-