Problem statement: Formally define the problem addressed in the paper. Briefly explain the significance of the problem in context of our course.

The paper proposes a method called Multi-task Representation learning model for Arrival Time estimation (MURAT) that creates meaningful representation that preserves various trip properties in the real-world and at the same time leverages the underlying road network and the spatiotemporal prior knowledge.

Theoretically, there are two methods to estimate arrival times - path-based estimation where we first estimate the travel time on individual links and then sum up all the link travel times in the given path. However, this method is computationally expensive and not a faithful estimate as dynamic factors like traffic change the link estimates on the go.

Second method is Origin Destination Estimation where the above issue is mitigated. This approach estimates the trip duration by averaging the scaled travel times of all historical trips with a similar origin, destination and time of the day. In addition to this, the spatio-temporal aspects of a trip are represented and learned along with the links so that estimate is more holistic in itself.

To learn better, the representation of these spatio temporal features are input through spatio temporal smoothness and shared factors among tasks are considered for the same.

Thus the entire problem statement reduces to learning a network that has three embeddings for the links, space feature and time features. Given an origin, destination and departure time, the paper tries to estimate the duration using the set of historical trip dataset X as well as the underlying road network G. The path P is a series of latitude longitude points captured by GPS. And a trip is a function of origin, destination, duration, departure time and path. Hence, a historic trip dataset used for training is a set of trips defined as above which make them different dimensions in the network.

List the major contributions of the paper. Which do you think is most significant and why?

The spatio-temporal feature is embedded through links in the first place which prevents the source and destination to and from following a Euclidean path to a adhere to the network created by both. For the spatial representation, the embedding of a node, is represented as the concatenation of the embeddings of its latitude and longitude, while for the temporal representation is represented as the concatenation of the embedding of “time in day” and “day in week”. By sharing embeddings, it imposes the priors that locations in the same latitude/longitude, and intervals at the same time of the day or day of the week should have similar representations. The latitude and longitude is incorporated in this network along with time estimates at every link.
What are the key concepts behind the approach in this paper. Provide simple explanations of those. Also provide a couple of simple exercises for the audience to check their understanding of the key concepts.

There are three components that make up the system. Each of these components are explained below:

**Representation Learning:-**
The objective of this step is to learn a vector for each link. A link can be understood as a road with a direction. (i.e. opposite sides of the same road are two different links) We can then have a loss function that encourages adjacent links to have similar representations and train our links using that loss function. Additional techniques can also be imposed on this learning process like first learning a representation based on unsupervised graph embedding techniques and then fine tuning the link embedding based on supervised signals. Overall, the goal of this step is to learn a representation of each link to characterize the underlying road network.

**Spatiotemporal Representation Learning:-**

The objective of this step is to integrate the prior knowledge of spatial and temporal signals into the embedding space. You can think of this step as enriching the embedding space from the previous step with additional features.

In this step, there are two spaces which we are trying to learn simultaneously: the spatial and the temporal space. In the spatial space, each node is a grid/region and nearby regions are connected using edges. In the temporal space, each node is a 5 minute interval and adjacent intervals are connected using edges. Now, we learn a fixed length vector representation for each of these nodes by enforcing spatiotemporal smoothness prior and periodicity by applying the Laplacian regularization over these vectors. Basically, adjacent nodes in both graphs will have similar properties.

**Multi-task Representation Learning:-**

Currently, the task of this model is to predict the travel time. But we could have a better, more generalized model if our model is also trained to predict other auxiliary related tasks along with predicting travel time. Another advantage is that this prevents our model from overfitting to the training dataset. So, this model is also trained to predict multiple related tasks like travel distance, number of links in the trip, number of tasks etc along with travel time.
In the final step, the link information, spatial and temporal information are embedded into the learned spaces. Secondly, the embedded representation together with other numerical features are fed as input into a deep residual network which generates the prediction for all K tasks. Thirdly, the objective is defined as the aggregated loss from multiple tasks as well as the graph Laplacian regularizers for the spatial and the temporal graphs. Finally, the embeddings, as well as the weights of the residual network, are jointly optimized using the Adam Optimizer.

What is the validation methodology (e.g. case studies, statistical hypothesis testing, proofs, simulation) used in this paper? Describe the strengths and weaknesses of the methodology. Why did authors choose this methodology?

The validation methodology used in this paper is testing the algorithm on three different datasets and comparing its performance versus other state-of-the-art algorithms. The paper also tests the significance of each of the different components that make up the model by
removing the component that needs to be tested and checking the performance and plugging in the component and checking whether there is a significant improvement in performance.

The strengths of this methodology is that it thoroughly checks that the algorithm does provide an improvement in performance. Three different datasets were considered and three different error metrics were computed. Trends wrt how error metrics are affected when we vary the travel distance and travel time were also justified clearly. Usually in machine learning models, it is easy to obfuscate the model with extraneous features that are noisy and we liked how the paper identified whether each component does contribute to the performance of the model by plugging it in and out of the model and checking the performance. All these factors contributed to an increase in the trustworthiness of the paper.

The main weakness of the testing methodology was the absence of testing the model in a dynamic environment. One of the main motivations for writing this paper is to correct errors in prediction when the actual path deviates from the predicted path due to dynamic changes in the traffic environment. So, it makes sense to test this model in a dynamic environment where, for example, a driver uses a standard state-of-the-art route prediction algorithm and how its travel time keeps changing as the traffic and route changes and verifying how close this model was able to predict the actual arrival time beforehand. As this type of testing was absent, it begs the question as to whether the paper actually went on to solve the problem that it intended to solve.

The authors chose this methodology because this is a prediction problem which has a ground-truth value. Therefore, it makes sense to consider different error metrics based on the ground truth and predicted value.

List the assumptions made by the authors. Critique an assumption that you believe is unreasonable. What is the impact of removing this unreasonable assumption on the solution proposed by the authors?

One of the assumptions that the author makes is that path-based approaches are less practical for fast online travel time because the route prediction step introduces expensive computations. However, the route prediction step is expensive precisely because it takes into consideration the dynamic changes in the environment. If it didn't, then we can precompute/cache a lot of the routes beforehand. But in this paper the author had not introduced dynamic changes in the environment to the model. He trains the model on a dataset and he tests it on a dataset after the model is built. There is no mention of how the model might go about introducing dynamic changes to the model. If we want to achieve that in this model, will the prediction step be more or less computationally expensive than the current state-of-the-art models?

The impact of removing this assumption would have forced the authors to have dynamic tests and also have experiments where they test the computational performance of the model instead of presuming that this model definitely performs better computationally. It is not clear as to whether this model fully captures all such dynamic changes such that online updates to the
model are not necessary and also whether the algorithm can be deployed as-is in a real life application.

If you were to rewrite this paper today, what would you preserve and what would you revise? Briefly justify.

I would definitely preserve the rigorous and meticulous testing of the different components in the system. This ensures that the model is lean and efficient.

I would revise how the testing was done like a traditional ML model rather than requiring the dynamic testing that seeks to answer whether the paper actually solves the problem it intends to solve.