Electric Vehicle Charge Station Placement via Node Classification, Structural Role Extraction, and Community Detection

Justin Luke
Civil and Environmental Engineering
jthluke@stanford.edu

Addie Petersen
Computer Science
adelinep@stanford.edu

Tom Mikkelson
Mathematical and Computational Science
tmikk11@stanford.edu

Major (Yue) Zeng
Civil and Environmental Engineering
majorzeng@stanford.edu

I. Introduction

Transportation and electricity/heat production sectors contribute 14% and 25% of global greenhouse gas (GHG) emissions, respectively, with both sectors expected to continue growing in emissions volume [1]. The electrification of road vehicles, which currently account for nearly 75% of transportation sector GHG emissions [2], is a significant driver of transportation decarbonization, particularly when progressing in tandem with the growth of renewable electricity generation. Due to the stochastic, intermittent, and distributed nature of renewable generation and uncontrolled EV charging, their integration to the power grid at a significant penetration level is limited by several technical challenges. These include unfavorable load shapes (e.g. steeper generation ramp-ups and larger evening peak demand), greater uncertainty in safe grid operations, voltage violations, and accelerated aging of grid assets. However, optimal control of autonomous EV fleet routing and charging can serve as a synergistic solution to support both low-carbon technologies' adoption [3] [4]. When coordinated, EVs can act as grid storage to buffer variability in renewable generation and avoid curtailment of overgenerated rooftop PV solar, while distributed PV offers lower electricity rates to the EV fleet operator and minimizes the need to upgrade transformers or start-up fossil fuel peaker plants to meet charging demand [5]. For such coordination to be effective, public charging infrastructure need to be sized and placed strategically to best serve the autonomous EV fleet's travel demands while also taking advantage of co-located distributed solar.

For our project, we will model San Francisco's transportation system as a network and analyze city travel using node classification, structural role extraction, and community detection graph techniques. With structural role extraction, we will determine the different roles of nodes within the graph and use these to inform features of the nodes in our graph. We also used two different community detection methods, the Louvain algorithm and spectral clustering, to determine EV charger placement. The motivation behind using community detection algorithms was to find "commute" communities, where there is a high degree of traffic flowing between nodes, and strategically place chargers within each community. Finally, we wanted to use iterative classification for node classification to assign each TAZ a number of EV chargers.

Our methods expand upon related work presented in Section II and are described in detail in Section III. Section IV discusses the model results and determines the quantity of EV chargers to place at each location informed by the analysis. We conclude in Section V where we describe extensions for future work to consider co-location with distributed rooftop solar as a feature and alternative methods of determining EV charger placement.

II. Related Work

This section presents three relevant previous works and how our methods extend their results.

A. On modeling transportation networks

Estimation of traffic flow changes using networks in networks approach (Hackl et al.) [6] seeks to model the relationship between dynamic traffic flow and the underlying transportation network structure. This paper details a novel "networks in networks" representation, a solution to the trade off between showing topological and spatial-temporal features. The authors conclude that the networks in networks approach performed better in assigning best paths to passengers when there was interruption in traffic flow compared to a single layer network.

However, this approach used travelers' perceived costs for a given path for edge weights, and thus the model must assume that travelers have prior information about interruptions in the transportation network. We would suggest leveraging additional data sets to better inform the travel time associated with each path, the edge weights for the second layer of the network, in order to improve the accuracy of the model.

B. Node classification for identifying charger locations

Many recent papers focus on the placement of new charging stations in urban areas to meet the future demand of EVs and incentivize their adoption through the mitigation of range anxiety. While EV charger placement is typically an optimization problem, some papers first formulate the problem as a network.

Optimizing the deployment of electric vehicle charging stations using pervasive mobility data (Vazifeh et al. 2019) [7] proposes a way to meet the demand of an urban region while minimizing cost to drivers. Cost is mainly determined by the total excess driving distance required to reach new chargers, with the potential to also consider the number of chargers and their energy overhead.

Excess driving distance is computed by comparing the shortest path on the road network from node C_j to C_i to the shortest path from C_j to C_i which also passing through a charging station. The estimated proposed solution using a Genetic Algorithm considering the whole network more accurately finds the optimal path than both true locally optimized solutions and the current deployment.

However, current infrastructure deployment is not taken into account when formulating the new optimal arrangement, so any potential gaps between starting from the current state and a blank canvas were not explored. Our research can take into account current charging locations when deciding where to best place future chargers. Additionally, while the algorithm's framework allows for considering cost variance in different charging locations, Vazifeh et al. considered cost equal for each installation.

Determination of charging infrastructure location for electric vehicles (Csonk and Csiszár 2017) [8] considered both the existing stations and the differing costs associated with each location in proposing a weighted multi-criteria method to meet local demands. In taking a closer look at the details surrounding each location, each node contains environmental, economic, and transportation related attributes.

The ability to meet intra-city demands are influenced by available services for daytime parking and the population of the surrounding area for nighttime parking. Csonk and Csiszár determine the installation potential of a hexagonal region mainly by the number of charging stations in the surrounding area and the average parking time at a service point.

The paper concedes that many details, including the optimal number of chargers in an area and the type of charger, are

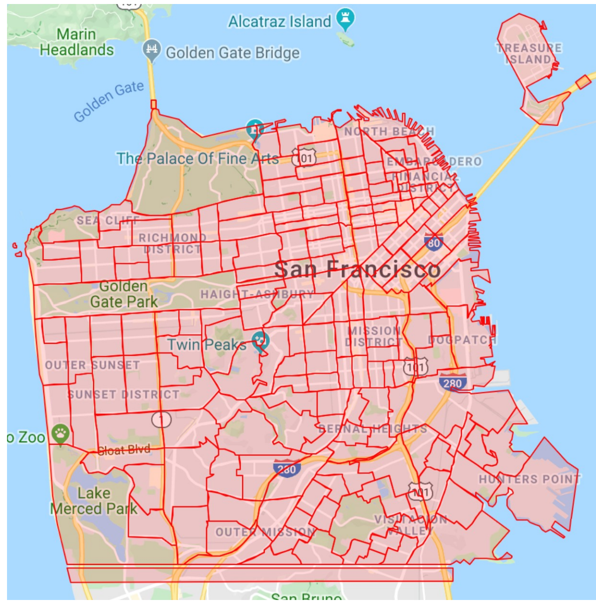

Fig. 1. 190 Transportation Analysis Zones in San Francisco. The additional three pass-through zones are indicated by the box above Treasure Island, the bar below Southern San Francisco, and the box above the Presidio along the Golden Gate Bridge.

to be derived later using local knowledge. Further research directions include seasonal and real time changes in traffic volume and charging demand. As a whole, the weighting system of the model appears quite arbitrary, and the model only considers the population of each point in the city. This leaves the addition of actual traffic flow and local data to the model a very promising line of further research, features that we are able to incorporate into each one of our models to inform EV charger placement.

III. Methodology

A. Data Collection and Analysis

Our main data set is origin-destination traffic data of San Francisco from Streetlight Inc. [9], which uses mobile GPS traces to determine traffic patterns. We divided San Francisco into its 190 transportation analysis zones (TAZ), which are standard partitions used for traffic study, shown in Figure 1. Additionally, we also added three pass-through zones to cover inbound and outbound traffic from the Bay Bridge, the Golden Gate Bridge, and all roads south of San Francisco. The TAZs were obtained as a shapefile from the Metropolitan Transportation Commission (MTC). The corresponding graph has 190 TAZ nodes and 3 additional pass-through nodes and 37155 edges between unique node pairs. For each origin-destination node pair, Streetlight provides the traffic volume, average trip distance, average trip duration, and demographic data of passengers. The data is taken from Sept-Oct 2018 and Mar-April 2019 and aggregated across all days of the week and all hours of the day.

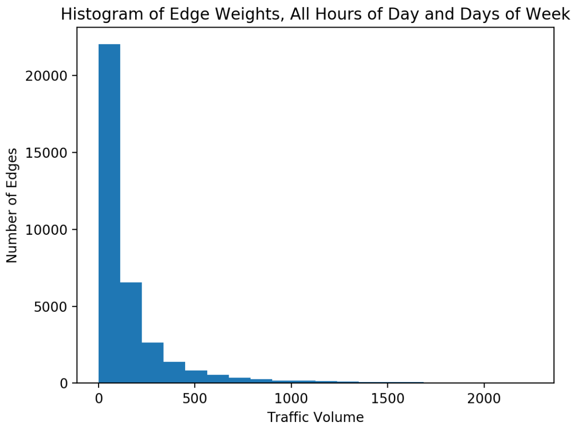

Fig. 2. Histogram of total edge weights (summed across all hours of day, days of week) from 1 to 2250 cars. Max volume is 10771, which is not shown in figure range.

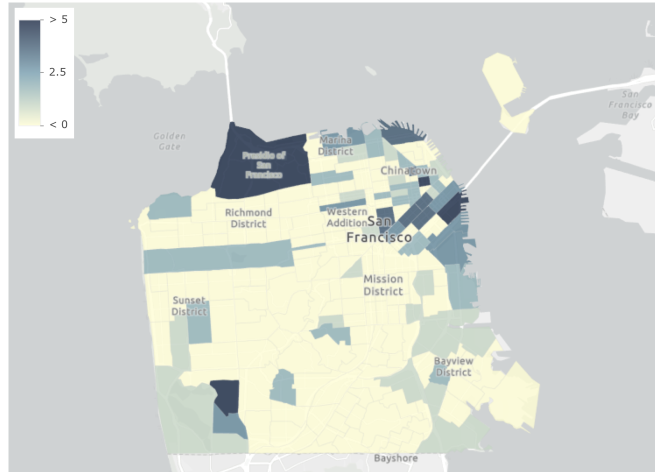

Fig. 3. Shows the present distribution of Level 2 and fast EV chargers across San Francisco. Units are number of chargers.

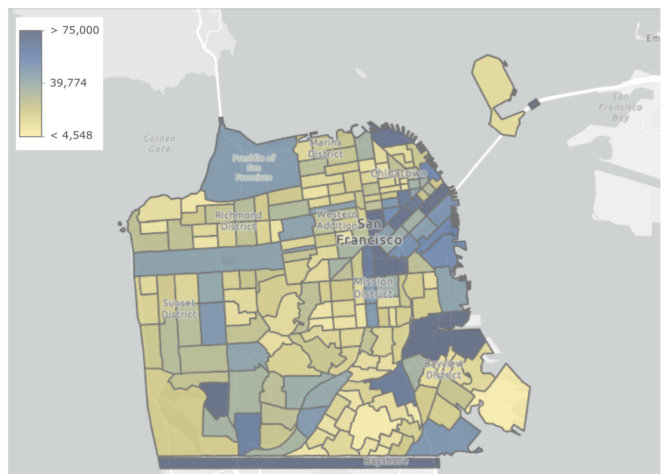

Fig. 4. Total (in and out flow) daily traffic volume

As we are working within one city and because even a small number of vehicles travelling between two nodes creates an edge, the road network of San Francisco is nearly fully connected. Thus, its graph visualization (not shown) and degree distribution are trivial; the distribution of total edge weight provides a better characterization of the network, shown in Figure 2. From the total edge weight histogram, we observe that most traffic volume between two zones are between 1 and 100 cars, but the max is 10771.

B. Baseline

Our problem of discovering transportation network features and placing public charging infrastructure for a future autonomous EV fleet is an unsupervised classification problem. Nevertheless, we propose two methods against which to evaluate our results.

We compare our produced placement of charging stations to where current charging stations have already been placed. To do so, we looked at current placements of Tesla and EVgo DC fast chargers as well as the Open Charger Map. From the Tesla, EVGo, and Open Charger Map websites, we extracted the latitude and longitude coordinates of the charging stations. We then manually mapped these to the zones within our San Francisco TAZ data set, and found that there is one Tesla supercharger, 20 EVgo chargers, and 124 chargers (of all charging speeds) within our region of study. A plot of these existing stations are shown Figure 3. We plan to use these data sets as our baseline to gain an intuitive sense of where chargers should be placed. However, one limitation to this approach is that EV penetration is small in present day and the future placement of charging infrastructure serving *all* San Francisco traffic, as we are studying, may be dramatically different from the few placements today.

Another method to validate results is to qualitatively interpret the placements in the context of San Francisco districts (e.g. commercial, industrial, residential) and total travel volume at each location, as shown in Figure 4. We expect a larger share of public charging infrastructure to be placed at commercial regions, tourist attractions, and high volume traffic zones.

C. Structural Role Extraction

Our approach to structural role extraction adapts the RolX discovery method as introduced in [10]. Given the road network is nearly fully connected, the conventional definition of egonet is useless as are the egonet-derived features, as all nodes will have identical egonet and features. We propose a method of choosing neighbors for each node u based on their relative contribution to the total in-flow and out-flow traffic of node u. Specifically, we greedily add neighbors to form the egonet until a threshold of total node traffic is covered. Thus, the base features (depth $K = 0$) of each node are:

$$\begin{bmatrix} \textit{node degree (total traffic = in + out)}, \\ \textit{total traffic within ``egonet''}, \\ \textit{total traffic from ``egonet'' to rest of graph}, \end{bmatrix}$$

When recursing, the recursive node feature at depth K aggregates features by:

$$\begin{bmatrix} \textit{feature at depth K-1}, \\ \textit{mean of egonet's depth K-1 features}, \\ \textit{sum of egonet's depth K-1 features}, \end{bmatrix}$$

From these recursive features, we determine structural roles placements based on cluster centroids obtained from k-means clustering. Examination of cluster centroids and the geo-spatial map of cluster labels can provide insight on structural roles in the transportation network. The number of k-means clusters is determined by the "elbow-joint" method which chooses k based on the inflection point of the distortion (mean sum-squares distance to centroid) curve.

D. Community Detection

For this section, we implemented the Louvain algorithm for community detection as well as spectral clustering [11]. Once we had detected the communities with each algorithm, we implemented two different methods of converting community membership to EV charger placement, where both assigned a total of 500 EV chargers. The decision of allocating 500 chargers was made because we wanted the average number of chargers per zone to be greater than 1 so that we would be able to see a dynamic range of chargers across zones. The first method was to allocate the same number of chargers to each community, and then to allocate those chargers within the community based on each node‘s relative out degree within the community. The second method was to allocate chargers to each community based on the each community‘s relative traffic flow, and then to distribute chargers within a community by each node‘s relative contribution to the community‘s traffic flow. We also considered a third method of converting communities to charger placement, where we would allocate a weighted number of chargers to each community based on the traffic volume, and then allocate a uniform number of chargers between the nodes of each community. However, we chose to not show the plots generated from this method, as it would be very similar to the plots showing the community segmentation.

Louvain algorithm for community detection: The Louvain algorithm greedily maximizes the total modularity of the graph and partition by iterating through 2 phases. The first phase assigns nodes membership to different communities to maximize the modularity, and the second phase aggregates each community into a super node to create a graph for the next phases. These phases are repeated until the modularity has been maximized. For our algorithm, we both implemented our own Louvain algorithm that could be used with a SNAP graph, as well as leveraged a built-in version of the Louvain algorithm compatible with igraphs. For both the built-in and our implemented Louvain algorithms, we used the traffic flow between two nodes as the directed edge weight between TAZs. Thus, we converted our data to both an igraph and a SNAP graph. We included the pass through zones in the community detection for Louvain so that we could understand which pass-through zones had high traffic between which communities of the San Francisco TAZs.

Spectral Clustering: k-way spectral clustering uses the following a 4-step process to find the minimum normalized k-way cut of a graph:

1) Building a Laplacian matrix of the graph
2) Find the eigenvectors and eigenvalues of the Laplacian
3) Build a k-dimensional representation of each node from their corresponding values in the eigenvectors of the 2^{nd} to k+1^{th} smallest eigenvalues.
4) Then, use k-means to cluster the nodes into communities

As our network was directed, we could not use the standard $L = D - A$ to build the Laplacian, so we instead used the Laplacian matrix for directed networks approximation described in pages 40-42 of the paper by Malliaros and Vazirgiannis [12].

$$L_d = I - \frac{\Pi^{\frac{1}{2}} P \Pi^{\frac{-1}{2}} + \Pi^{\frac{-1}{2}} P^T \Pi^{\frac{1}{2}}}{2} \quad (1)$$

Where P is the transition matrix built from our weighted and directed adjacency matrix A, i.e. $P_{ij} = \frac{A_{ij}}{k_i^{out}}$, and Π is a diagonal matrix with the probability of finding each node on a random walk (the stationary distribution of the random walk), i.e. $\Pi = diag(\pi_1, ..., \pi_n)$. This definition of the Laplacian allows for the eigenvector corresponding to the second smallest non-zero eigenvalue of L_d to be used to create the best cut in the network. This can be extended from a bi-partition to our case of a k-partition by looking at the k+1 smallest eigenvalues of L_d and their corresponding eigenvectors.

E. Iterative Classification

This method is a semi-supervised relational classification method, which predicts charging station placement based on existing patterns extracted from Open Charge Map [13]. It assumes that current station placement is ideal and that future stations will be placed at similar locations as current ones. It classifies nodes based on both node features and labels of neighboring nodes. In this particular implementation, a linear regression model is trained on a concatenated vector for each node including node features (population in the zone, total traffic volume, and average trip duration, etc) and relational feature (the sum of its neighbor's labels weighted by the traffic in between). Then the method iteratively updates relational features, classification model and labels until the labels stabilize. The algorithm is shown below

Algorithm 1: Iterative Classification Algorithm

Result: Number of charging station prediction
Initialization
$n_i \leftarrow$ node features vector for $i = 1, 2, 3 \ldots 193$
$N_j \leftarrow$ training zones where l_j number of public charging stations is known, $|N_j| = 45$
Bootstrap Phase
Local linear classifier $f(n_i)$ to predict label Y_i
$r_i \leftarrow$ relational feature $\sum_{j \leftarrow N_i} Y_j \cdot w_{ij}$
Global linear classifier $g([n_i, r_i])$ to predict label Y_i'
Iterative Phase
while $||Y - Y'|| < 0.001$ **do**
  $Y_i \leftarrow Y_i'$
  $r_i \leftarrow$ relational feature $\sum_{j \leftarrow N_i} Y_j \cdot w_{ij}$
  Global linear classifier $g([n_i, r_i])$ to predict label Y_i'
end

IV. Results and Findings

A. Structural Role Extraction

Three experimental parameters were considered: whether to include or exclude the three pass-through zones, the depth of recursion for producing recursive features, and the threshold of total traffic volume covered when defining the "egonet" of a node. A grid search was performed to determine the recursion depth and egonet traffic threshold, which found a depth of 3 and neighbors covering 33% of a node's total traffic maximized graph modularity.

When pass-through zones are included, the model was able to identify that the two bridge pass-through zones have a special role and distinguish them as a separate cluster. The corresponding centroid has the greatest total node flow, within-egonet travel, and egonet-to-graph travel which agrees with the high volume on the bridges. It also has egonet nodes with lowest total node flow possibly because a diverse combination of zones, including less significant ones, all have traffic leading to the bridges.

Because the pass-through zones have much higher volume and skew the edge weight distribution, excluding them allows for analysis of travel within the 190 San Francisco TAZs. Figure 5 shows the labeling of roles in San Francisco. Examining the recursive feature vector of the centroids of the resulting $k = 4$ clusters, cluster 0 represents nodes with the greatest total node flow, greatest within-egonet traffic, and having egonet nodes that also have high total flow. Notably, high-tourist locations, such as Golden Gate Park, Golden Gate Bridge, Twin Peaks have this label. Cluster 1 also has a high total flow but the lowest within-egonet and egonet-to-graph flows, suggesting these nodes have greater within-node (self-loop) travel. This may agree with the large spatial size of the TAZs with this label, such as those in Sunset and Bayview

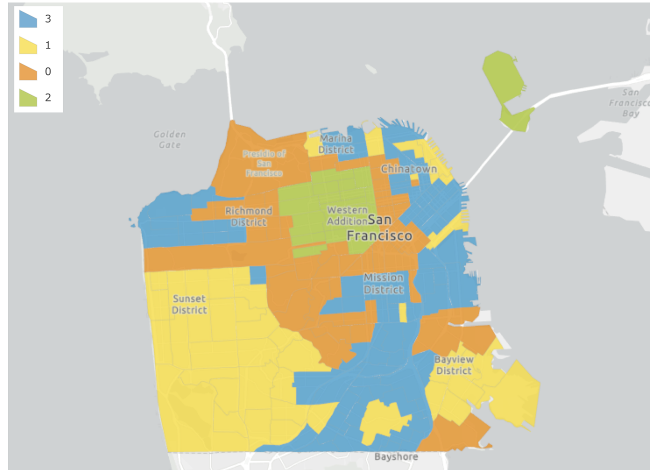

Fig. 5. Structural roles identified by RolX. Each color denotes a different role detected by RolX.

districts, as they are more likely to have passengers traveling by vehicle within a TAZ. On the contrary, Cluster 2 has low node total flow but high within-egonet and egonet-to-graph weight, suggesting these nodes are not themselves significant travel destinations but connect to more cohesive locations, which agrees with the assignment of islands in Figure 5 and the Western Addition neighborhood, which is more interior. Lastly, cluster 3 has egonet nodes that have high total flow. The positioning of nodes of this label along coastal and boundary regions indeed suggests their traffic are shared with inner significant zones. These discovered roles can inform public charging station placement: locations with roles like cluster 0 should have a larger share as high-travel destinations whereas role cluster 1 and 2 should have a lower share having shorter distance travel and less frequent visits.

B. Community Detection

Louvain Algorithm: We ran both the built-in implementation with an igraph representation of our data, as well as our own implementation of the Louvain algorithm on a SNAP graph. Due to runtime intractability with our implementation of the Louvain algorithm, we chose to only move forward with the built-in implementation of the Louvain algorithm with an igraph representation of the data.

Figure 6 shows the results of community detection and exact charger placement. From the top geospatial graph, we can see that the graph was segmented into six total communities that look to be influenced by geographic proximity, as well as separating the Golden Gate pass through zone into its own community and Treasure Island and the Bay Bridge pass through zone into its own community. These communities intuitively make sense, because there is a greater degree of traffic between TAZs that are closer to one another. We next examine the EV charger placement based on the community segmentation. For the first charger placement policy, EV chargers are uniformly distributed across communities and

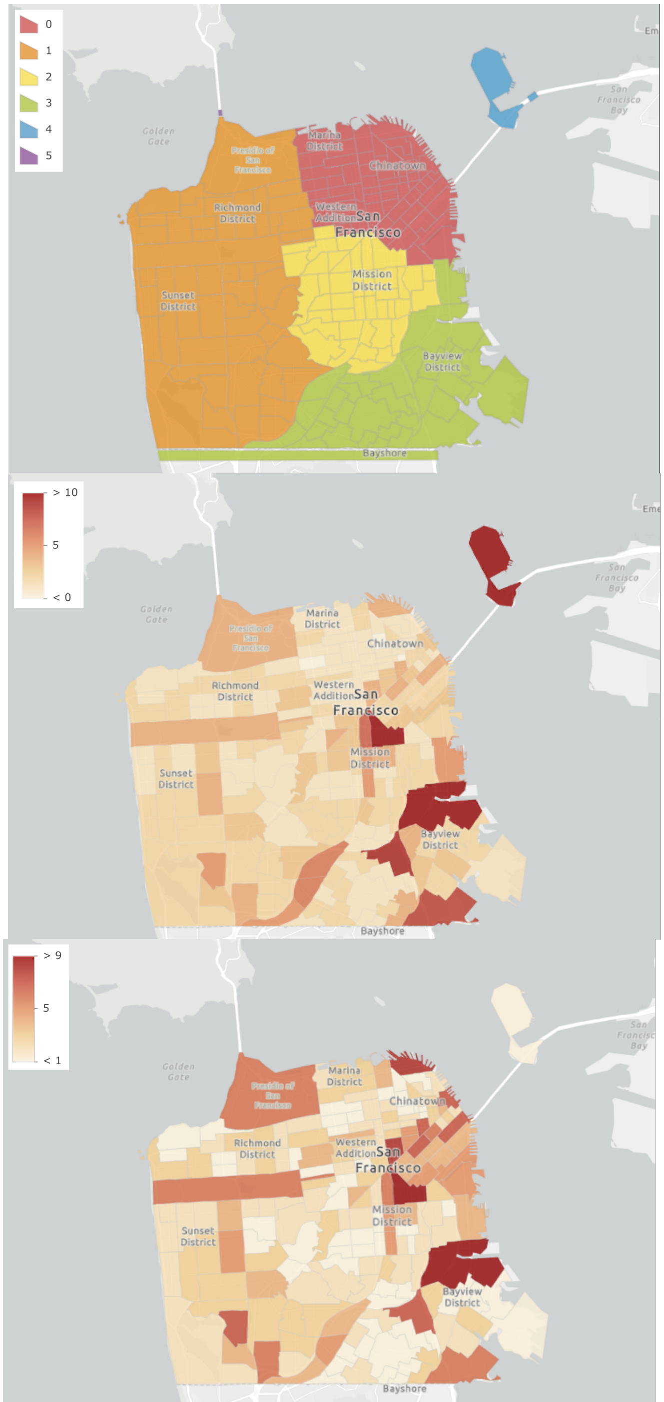

Fig. 6. Community results from Louvain algorithm and EV charger placement. (Top) Shows the community segmentation of TAZs from the Louvain algorithm, where each color denotes different community membership. (Middle) Shows the EV charger placement according to method 1, where each community gets the same number of chargers. Units in legend are number of chargers. (Bottom) Shows the EV charger placement according to method 2, where each community gets the number of chargers proportional to its traffic flow. Units in legend are number of chargers.

then placed within communities by each node‘s out-degree, we can see that the placement is relatively similar to the traffic flow plot (Figure 1), but there are more chargers placed near the pass-through zones (see Treasure Island, Golden Gate park, and zones along the southern border). For the second charger placement policy, EV chargers are distributed across communities by the community‘s traffic flow and then placed within communities by each node‘s out-degree; we can see that this plot now well-approximates traffic flow plot (Figure 1), which makes sense given that the weight attribute of the community detection was done with traffic flow.

Spectral Clustering: After carrying out the first two steps of spectral clustering to build the Laplacian and calculate the eigenvalues and eigenvectors, in which we ignore the three pass-through nodes to achieve more balanced eigenvectors for partitioning, we were left with some choices on how to carry out steps 3 and 4.

For step 3, building the k-dimensional representation of each node, we could either build this feature vector with 1s and 0s in the i$^{\text{th}}$ position depending on whether each node's partition in the i+1$^{\text{th}}$ eigenvector, partitioning on whether this value is greater than zero or less than zero, respectively. Alternatively, the i$^{\text{th}}$ position could include the actual value in the i+1$^{\text{th}}$ eigenvector. We decided on the latter, as the clustering on these continuous attributes provided contiguous community clusters, where partitioning based on the binary resulted in some split communities when plotted on the real map. This method also resulted in better community scores, as determined in step 4.

For step 4, we were faced with how to determine the optimal k, the number of communities. We carried out spectral clustering for a range of ks and for each k we calculated the modularity of the graph as a whole using the method described by Dugué and Perez [11].

$$Q_d = \frac{1}{m}\sum_{i,j}\Big[A_{ij} - \frac{d_i^{in} d_j^{out}}{m}\Big]\delta c_i, c_j \qquad (2)$$

As can be seen in (Figure 7), the modularity peaked at 5 communities, so we chose to use this cut when assigning charging stations.

The two methods for charger placement showed very different results, seen in (Figure 8). The uniform approach, dividing the 500 chargers uniformly to each of the 5 communities and divided those 100 station within each zone based on degree, appear to under serve the high traffic downtown area, while zones outside of the downtown community with higher traffic relative to their community receive many charges despite relatively lower traffic overall, as seen in two communities in the Lake Merced and Bayview area.

The second method still found the Bayview and Lake Merced area to be important, along with the North Waterfront

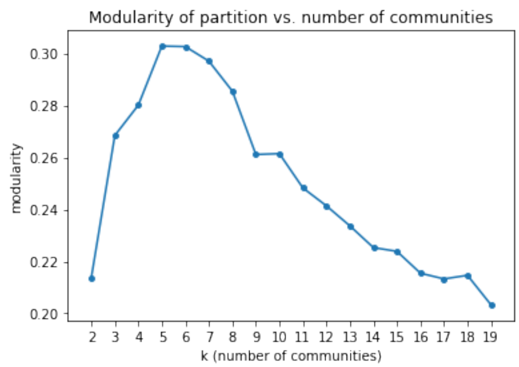

Fig. 7. Finding optimal number k communities for spectral clustering.

and one zone in the Mission district. Scattered zone throughout the downtown and financial district also received elevated importance. A non-uniform approach makes the most sense for spectral clustering as the relative importance, demand, and number of nodes of each zone is not kept equal, and instead the focus is on finding the most distinct communities.

C. Iterative Classification

Based on the weight of each features (normalized) in the final regression model, the most relevant node feature in this prediction are average trip length and duration for trips origins or destinies in each zone. Demographic features like median income or level of education are less important.

The predicted number of charging stations in each transportation analysis zone are shown in Figure 9 below. The total number of chargers is 421. In general, it predicts more charger stations on the outskirt of the city and less in the interior of the city. It places more chargers around Richmond District, SOMA, South Park, Mission Bay, Potrero Hill and Oceanview.

D. Comparison of Methods

The communities produced by the Louvain algorithm and spectral clustering are fairly similar, given both are relating to graph modularity maximization. They succeed in distinguishing regional districts in San Francisco. Should chargers be placed based on community size, both would place more chargers in the dense traffic region in the city's northwest (e.g. downtown, financial district), which agrees with both traffic flow data and current charging placements.

The predicted number of chargers from iterative classification, the Louvain algorithm and spectral clustering for each transportation analysis zone are averaged and shown in Figure 10. It places high importance around Mission Bay and Dogpatch area, Richmond area and Lake Merced area. We also plotted the TAZs for which all three methods placed a number of chargers that was greater than the median (colored in red), and the TAZs where all three methods placed a number of chargers less than the median (colored in blue) 10. Because a majority of the TAZs are colored in this plot, this is further evidence of the similarity in the charger

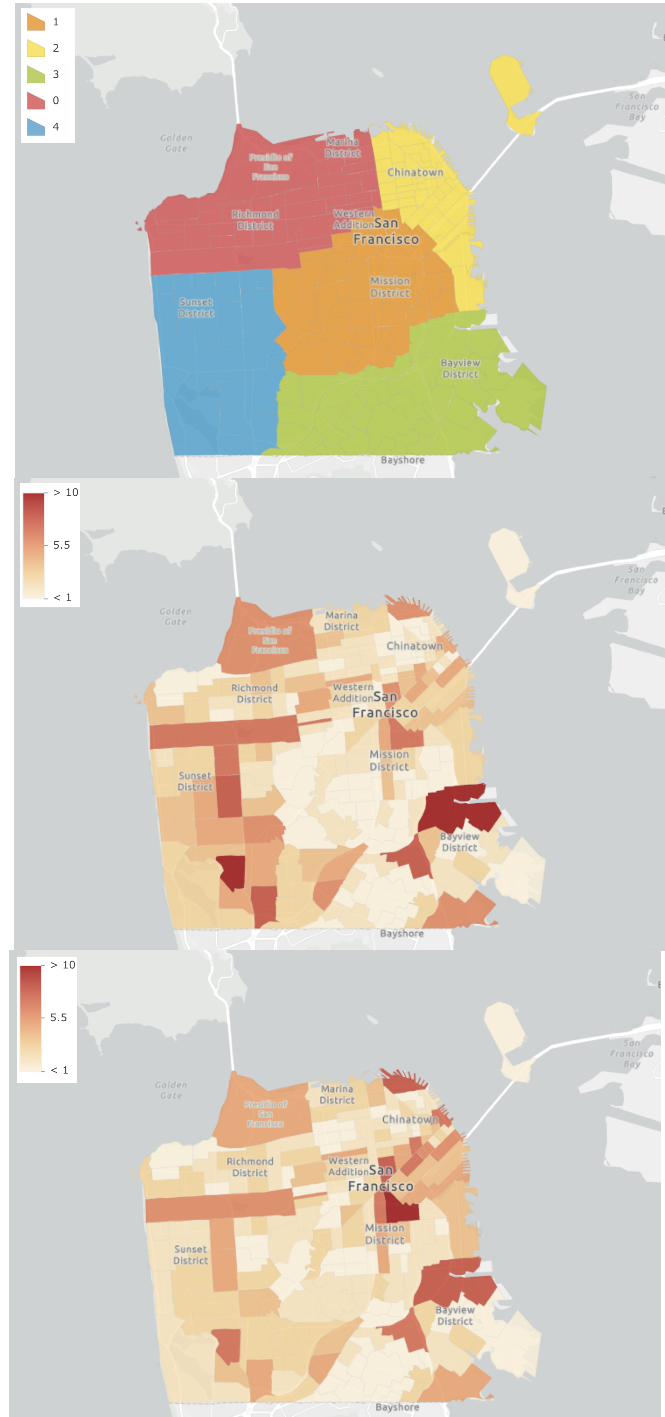

Fig. 8. Community results from spectral clustering and EV charger placement. (Top) Shows the community segmentation of TAZs from spectral clustering following by k-means with $k = 5$, where each color denotes different community membership. (Middle) Shows the EV charger placement according to method 1, where each community gets the same number of chargers. Units in legend are number of chargers. (Bottom) Shows the EV charger placement according to method 2, where each community gets the number of chargers proportional to its traffic flow. Units in legend are number of chargers.

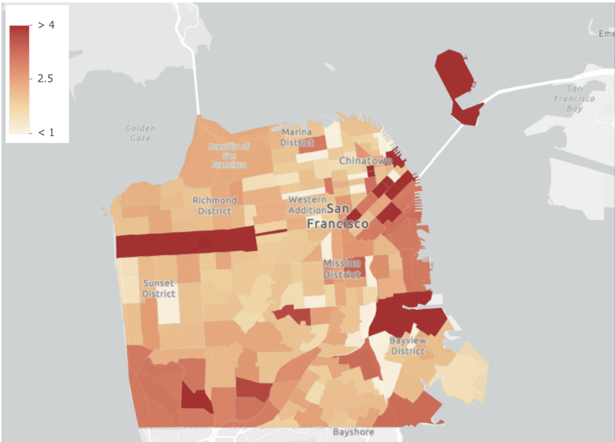

Fig. 9. Iterative classification prediction on the number of chargers in each transportation analysis zone .

placement across methods. We hypothesize that the similarity of the results for each of the methods is due to using the traffic flow between nodes as the primary edge weight, and therefore charger placement is roughly correlated with traffic flow in all methods regardless of the community assignment. The interpretation of structural roles, namely high placement in role 0 and lower in role 1 and 2 of each location outlined in Section IV-A also support charge placements

There are some subtle differences between the charger placement across the three methods. Examining the communities identified by spectral clustering further, we can see that Treasure Island is not identified as a ideal place to put chargers, while both iterative classification and the Louvain algorithm place many chargers on Treasure Island. This difference could be attributed to the spectral clustering algorithm excluding pass through zones in the analysis, such that the relative importance of Treasure Island as being connected to the Bay Bridge pass through zone is diminished. The iterative classification algorithm also places more chargers in the Golden Gate Park zone above the Sunset District and some TAZs in the Lake Merced area/southwest corner of San Francisco, whereas the Louvain and spectral clustering algorithms do not place as many chargers there. We can attribute this difference to iterative classification having the baseline of present day chargers to train on, whereas the community detection algorithms do not.

Finally, in comparison to our baseline of the current placement of Level 2 and fast EV chargers across San Francisco (Figure 3), we see fewer charges placed in the Presidio in our models. While still among the top 50% in all of our models, the Presidio is no longer the top zone. Our models placement of stations in Lake Merced, downtown, Fisherman's Wharf, and Golden Gate Park match up fairly well with the current distribution. Additionally, the areas without many nodes, specifically the interior and residential regions, are very similar between our placement and the baseline.

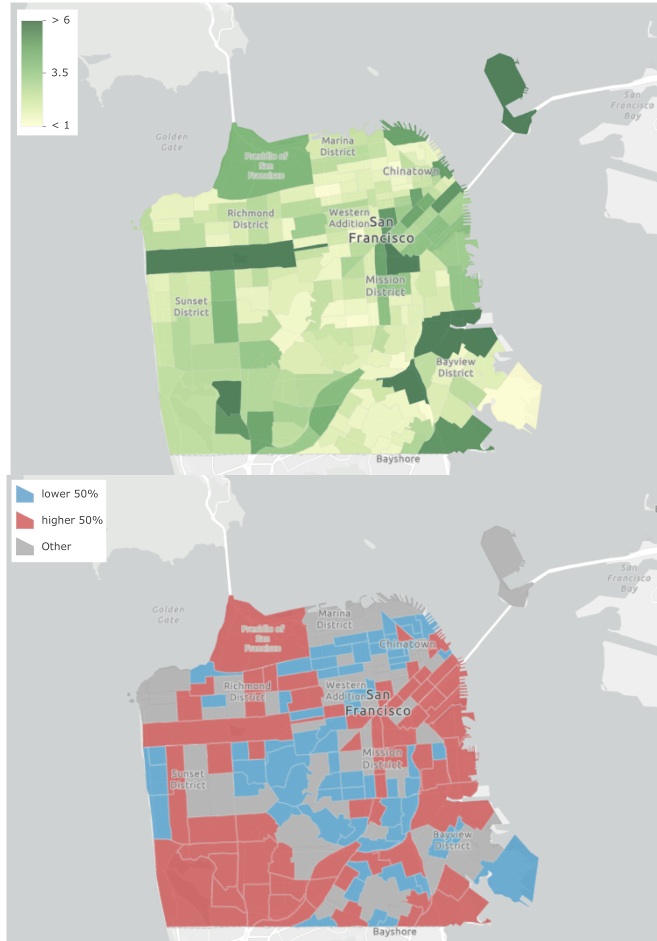

Fig. 10. Comparison of methods. (Top) Shows average number of chargers predicted by three methods in each TAZ, all scaled to a total of 500 chargers and with all methods weighted equally. (Bottom) Shows TAZs that were either given a number of chargers below the median in all three methods (colored in blue), or given a number of chargers above the median placed by all three methods (colored in red).

V. Conclusion

For our project, we were able to complete four different methods to analyze the transportation network and place EV chargers across San Francisco, which was discretized into 193 traffic zone nodes with traffic volume represented by edge weight. From our role extraction algorithm, we were able to learn more features about the nodes that shared similar structure in the graph to inform our qualitative analysis of the chargers placement. The two community detection methods, the Louvain algorithm and spectral clustering algorithm, both detected similar communities, with the exception that the Louvain algorithm had the pass through zones in nearly their own communities and the spectral clustering algorithm split the western half of San Francisco into two communities rather than one. These discovered communities notably correspond to major districts of San Francisco. The two methods of community detection and iterative classification placed similar charger distributions across San Francisco, with some differences that can be attributed to using pass through zones in analysis or whether the baseline of current chargers was utilized. Overall, we were able to obtain overall similar EV charger placement with varied methods of community detection and iterative classification. We hypothesize that this result was because we all used traffic flow as the dominant attribute to determine communities and charger placement, and as a result, the EV charger placement through our algorithms shared correlation to placing EV chargers at the nodes with the highest weights.

For future work, we would like to implement another method of EV charger placement using clustering of graph neural net (GNN) feature, node embeddings, and PageRank. This would allow us to directly compare iterative node classification with a more sophisticated, machine learning-based method of node embeddings. Furthermore, we used aggregated data across all hours of the day and days of the week to construct the graph representation of San Francisco TAZs. It would be interesting to study how the EV charger placement would change if we were to only use weekdays, or take the maximum traffic flow between nodes rather than aggregating it to predict EV charger placement. We additionally would like to use excess solar generation as a time-varying feature to inform where it would be best to place EV chargers to use this energy that would otherwise be curtailed. This may require modeling the network as a temporal graph to capture electric load profile and solar generation's diurnal patterns. Finally, we used a relatively small graph to model traffic of San Francisco. It would be interesting to study how these models change if we were to make the graph of the whole Bay Area, as this would increase the number of nodes and similarly the complexity of determining charger placement in a larger area. Modeling EV fast charger placement for long distance travel (for example, along highways in California rather than within a city) would also be an productive direction for future research to see how graph-based methods fare in a long-distance commuting context.

Contributions

All authors agree that we have contributed equally to all deliverables of the project and request to be graded equally.

Acknowledgment

The authors would like to acknowledge all of the CS 224W teaching staff for their helpful advice and support throughout the quarter during office hours and over Piazza. In particular, Michele was very patient and thoughtful in providing suggestions for potential methods in our initial project meetings. Justin would also like to acknowledge Streetlight Inc. for providing the travel demand data under an academic license.

References

[1] C. W. Team, R. Pachauri, and L. Meyer, "Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change," IPCC, Tech. Rep., 2014.

[2] "Transport," 2018. [Online]. Available: https://www.iea.org/tcep/transport/

[3] F. Rossi, R. Iglesias, M. Alizadeh, and M. Pavone, "On the interaction between autonomous mobility-on-demand systems and the power network: models and coordination algorithms," *IEEE Transactions on Control of Network Systems*, 2019.

[4] A. Estandia, M. Schiffer, F. Rossi, E. C. Kara, R. Rajagopal, and M. Pavone, "On the interaction between autonomous mobility on demand systems and power distribution networks - an optimal power flow approach," *arXiv*, 2019.

[5] J. Coignard, S. Saxena, J. Greenblatt, and D. Wang, "Clean vehicles as an enabler for a clean electricity grid," *Environmental Research Letters*, vol. 13, no. 5, 2018.

[6] J. Hackl and B. T. Adey, "Estimation of traffic flow changes using networks in networks approaches," *Applied Network Science*, vol. 4, no. 1, p. 28, May 2019. [Online]. Available: https://doi.org/10.1007/s41109-019-0139-y

[7] M. M. Vazifeh, H. Zhang, P. Santi, and C. Ratti, "Optimizing the deployment of electric vehicle charging stations using pervasive mobility data," *Transportation Research Part A: Policy and Practice*, vol. 121, pp. 75 – 91, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856417300010

[8] B. Csonka and C. Csiszár, "Determination of charging infrastructure location for electric vehicles," *Transportation Research Procedia*, vol. 27, pp. 768 – 775, 2017, 20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146517310128

[9] "Streetlight data: — transportation metrics," 2019. [Online]. Available: https://www.streetlightdata.com/transportation-metrics/

[10] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li, "Rolx: structural role extraction & mining in large graphs," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1231–1239.

[11] N. Dugué and A. Perez, "Directed Louvain : maximizing modularity in directed networks," Université d'Orléans, Research Report, Nov. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01231784

[12] F. D. Malliaros and M. Vazirgiannis, "Clustering and community detection in directed networks: A survey," *ArXiv*, vol. abs/1308.0971, 2013.

[13] O. C. Map, "The open charge map api," 2016.