Entropy Maximization for Range Addition Stations

Aaron I. Rabinowitz

Contents

1 Introduction

Road vehicles require a network of roads and a network of Range Addition Stations (RASs) whose nodes are coincident with a subset of the nodes of the road network. Allocation of RASs to road network nodes is a central concern in transportation planning. To a great extent, the topology of the RAS network will be determined by the topology of the road network, itself determined by and determinative of population and economic topology. In other words, roads connect places, RASs serve vehicles on roads. It is unlikely that, for the sake of the RAS network, new roads will be built or population concentrations will shift. Thus, the problem of optimal resource allocation, with respect to RASs should aim to propose locations for RASs such that the performance of the road network is maximized with respect to the demand generated by population and economic topology.

When designing a system, the initial stages of the design should make explicit the relative importance of objectives in conflict. A common conflict is efficiency vs. resilience. There are many definitions of resilience but, in general, systems which ar more redundant are more resilient. It is intuitive to view efficiency and redundancy as conflicting objectives. From the point of view of a RAS network, the most efficient system is one wherein every RAS is at full capacity at all times.

For such a network, user delays due to queuing are guaranteed. Although more efficient and, thus, able to charge less per unit energy delivered, consumers will find the experience annoying and the market will correct providing some redundancy. From the perspective of the road network, significant RAS redundancy may be called for in certain areas where individual stations are very inefficient. In order to make the road network functional, inefficient RASs must be funded via incentives, the effectiveness of which will depend on how well station locations are optimized.

Consider the scenario shown in Figure 1. Cities 1 and 2 are connected by a single road with one RAS (A) located halfway between the cities. A vehicle with unlimited range could make the trip between the cities without stopping at A but the maximum range of the relevant vehicle is greater than the distance from either city to A but less than the distance between the cities.

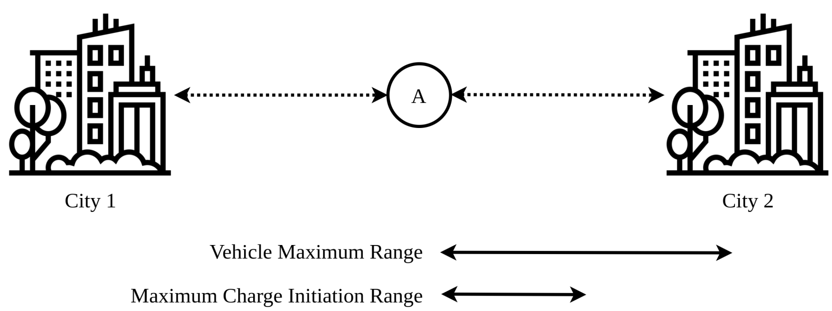

Figure 1: Cities connected by one road with one RAS. Solid lines indicate a link is sufficient to lead to a range addition event after traversal, dashed lines indicate the opposite, and dotted lines indicate a link which originates and/or terminates in a city.

Drivers heading from city 1 to city 2 and vice versa will have to utilize A. There are two consequences to this reliance: (1) should A go offline then no traffic can make it from city 1 to city 2 and (2) since all traffic must utilize A all vehicles will have to add range at equivalent stages of their trip. The second consequence is of greater importance if traffic between the cities is concentrated at certain time slots. Consider now a second set of cities, Cities 3 and 4 as in Figure 2 which are also connected by just one road but this road hosts three RASs (B, C, and D).

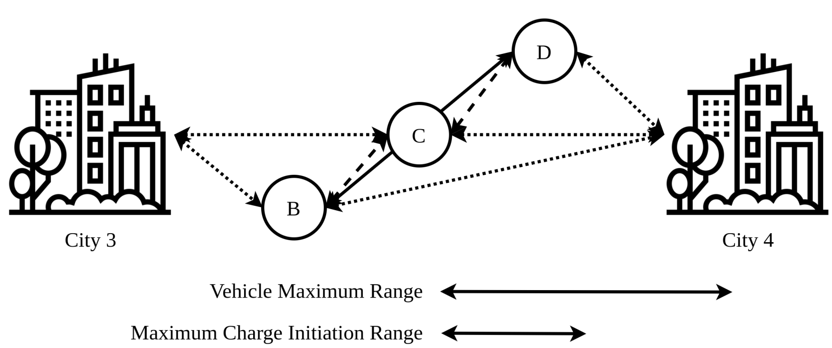

Figure 2: Cities connected by one road with multiple RASs. Solid lines indicate a link is sufficient to lead to a range addition event after traversal, dashed lines indicate the opposite, and dotted lines indicate a link which originates and/or terminates in a city.

Limited range vehicles have several options of, nominally, equal cost in this scenario. Consider the graph topology; B, C, and D are reachable from City 4, B and C are reachable from City 3, and the RASs are mutually within range. The 3 - 4 direction has two equivalent paths (3 - B - 4 and 3 - C - 4) and one higher cost path (3 - B - D -4). The 4 - 3 direction has three equivalent paths (through each station) and one higher cost path (4 - D - B - 3). In the first scenario there were only two shortest paths for travelers in cities 1 and 2, a number which increases to five for cities 3 and 4 with two additional viable paths. One might expect that, over time, drivers will tend to distribute themselves among the various paths thus reducing the expected wait time and the reliance on any one station. It should be noted, however, that the redundancy was not increased by adding stations but, rather, by adding stations in the locations in which they were added. If most cars begin their trips at roughly the same time then all will reach C at roughly the same time but by the time cars from 3 reach D most cars from 4 will be close to B. The optionality created by adding new stations as opposed to new chargers at the same station reduces demand coordination allowing expected queuing time to decrease without altering the vehicles per charger ratio. When traveling in the 3 - 4 direction one may observe that the choice boils down to "charge at B" or "charge at C". One is very unlikely to charge at B and again at C because the distance between the nodes and from C to 4 is not great enough to justify it. Before node B the driver had three viable paths, after passing node B the driver has only one: much information is gained by passing node B.

Finally, consider the situation of cities 5 and 6 as seen in Figure 3. Cities 5 and 6 are farther apart than the previous examples and are connected by two main roads.

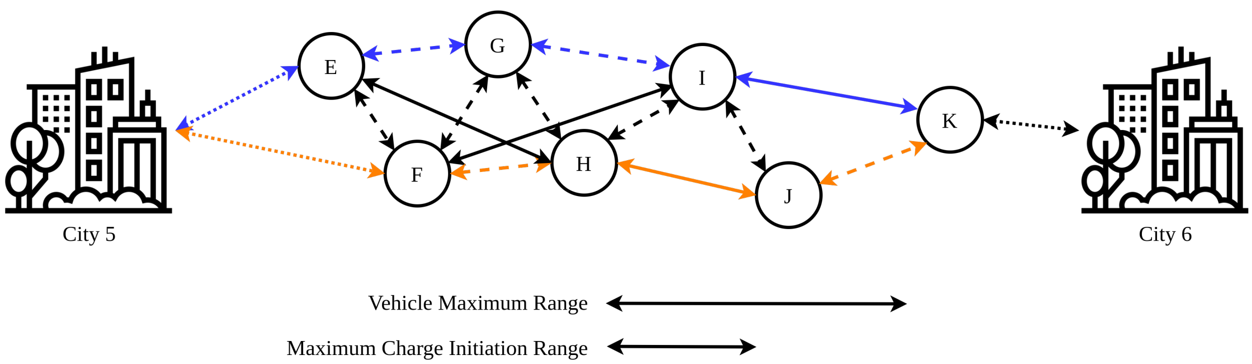

Figure 3: Cities connected by multiple roads with multiple RASs. Solid lines indicate a link is sufficient to lead to a range addition event after traversal, dashed lines indicate the opposite, and dotted lines indicate a link which originates and/or terminates in a city.

The upper main road (blue links) hosts stations E, G, I, and K and the lower road (orange links) hosts stations F, H, and J. There are many secondary roads which connect the nodes of the main road and both main roads converge at K before heading to city 6.

2 Information Theory

The core concept of information theory is the quantification of the informational value of a given system state. The degree to which a state provides information is the degree to which that state is unexpected. As an example, a lottery draw selects just one sequence from a massive set of possible sequences each one of which is equally likely to be the winner. Learning that a given sequence is not a winner provides little information concerning the winning sequence. As the draw progresses each number eliminates the vast majority of remaining possible sequences making all remaining possible winners much more likely to win. Most lottery players will be eliminated by a given number draw and yet the announcement will state the number selected instead of the numbers eliminated because it is a much more efficient way to provide the same information.

The information content of an event E in inversely proportional to the likelihood of event E. The information content of event E is defined as

$$I(E) = log\left(\frac{1}{p(E)}\right) \tag{1}$$

where $p(E)$ is the probability of event E and the base of the logarithm corresponds to the size of the set of possible values for E and is often 2, e, or 10. Entropy (S) is the expectation of information content for a random trial. Entropy is defined as

$$S = \sum_{E \in \hat{E}} p(E)I(E) = \sum_{i=1}^{N} p(E)log\left(\frac{1}{p(E)}\right) \tag{2}$$

where $\hat{E}$ is the set of possible events. Let A be a sequence of 2 coin flips where $p(h) = 1 - p(t) = 0.5$. The possible sequences for A are hh, ht, th, and tt all with a probability of 0.25. Using base 2, the entropy for A is 2 bits. In other words, prior to the two coin flips one has no information on what sequence will come up. Similarly, the information content for any possible sequence in A is 2 bits meaning that, after two flips, one is completely certain what sequence has come up. If, however, $p(h) = 1 - p(t) = 0.75$ then $p(hh) = 0.5625$, $p(ht) = p(th) = 0.375$, and $p(tt) = 0.0625$. In this case $I(hh) = 0.83$, $I(ht) = I(th) = 1.42$, and $I(tt) = 4.0$ and the entropy of A is 0.864. Because of the difference in odds, one can be more certain about the outcome of the latter scenario than of the former scenario before any flips have taken place. Similarly, the uncertainty of two dice rolls is greater than that of two coin flips even if both are "fair" because there are more possibilities for each dice roll.

Selection of the logarithmic base is somewhat arbitrary and, as long as consistently applied, will not effect the order of entropy among sets. However, there is physical meaning to entropy and the logarithmic base generally reflects what type of problem is being solved. In communications the base is usually 2 and the units of entropy are bits. The physical interpretation of this is that in order to attain relative certainty about the content of a message a certain number of bits are required. English words are largely composed of common sequences of letters and thus, if one knows that a sequence of ASCII characters represent English text then the entropy of the sequence is much lower than if one has to assume the sequence is random. English sentences also mostly follow a common structure described by subject-verb-object word ordering and preceding articles and adjectives which allows for sequence-trained large language models to generate highly plausible text sequences. In statistical thermodynamics Shannon Entropy is computed using natural logarithms where the unit of entropy is Joules per Kelvin, the same units as heat capacity, allowing entropy to be used in thermodynamic equations.

3 Graph Theory

4 Centrality

Consider the following graphs, which nodes are central?

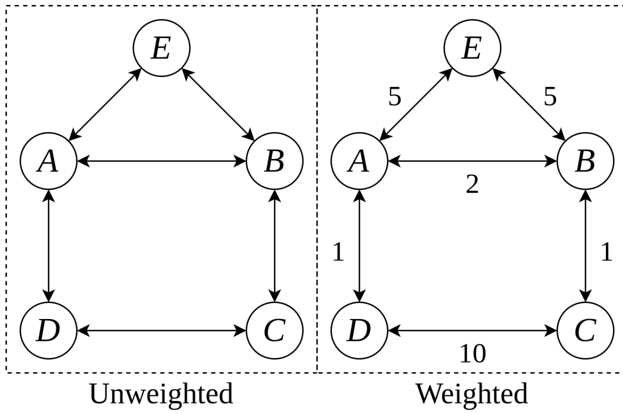

Figure 4: Simple example undirected graph unweighted and weighted isomorphs (from [1])

From Figure 4 it is apparent that A and B are more central than the other nodes. Graphs, however, need not have a geometric center nor any plainly obvious 2-D geometry. Take, for example, a network of communications satellites in low Earth orbit. Although the state of the network could, at any point, be projected onto a 2-D surface it would be arbitrary where the corners of the map should be located. In 3-D space all satellites would be approximately the same distance from the geometric center, the center of the Earth, a location where there will be no proximate satellites. Centrality, thus, must have a more abstract definition. There are four commonly used metrics for centrality, degree centrality, closeness centrality, betweenness centrality, and information centrality. These will be defined in the below subsections.

4.1 Degree Centrality

A computationally simple way to compute adjacency-focused centrality is degree centrality. Degree centrality is defined as

$$\Omega_d(v) = \frac{\eta(v)}{n-1} \tag{3}$$

where $\eta(v)$ is a function returning a value computed from the adjacency of node v. Degree is the number of adjacent nodes adjacent to a given node. A cell tower has high degree centrality as part of the cellular communications network.

4.2 Closeness Centrality

For a graph G, the sets of shortest paths from each node v to all other nodes $R_{v,U}$ and vice versa $R_{U,v}$ are known. Closeness centrality is defined as

$$\Omega_c(v) = \frac{n-1}{\sum_{u \in U} \eta(v,u)} \tag{4}$$

where n is the cardinality of V and $\eta(v,u)$ is the cost of the shortest path from v to u. Closeness centrality reflects how close a given node is to those nodes reachable from it via shortest paths. St. Louis has high closeness centrality as part of the US freeway system.

4.3 Betweenness Centrality

For a graph G, the sets of shortest paths from each node v to all other nodes $R_{v,U}$ and vice versa $R_{U,v}$ are known. Betweenness centrality is defined as

$$\Omega_b(v) = \sum_{o,d \in V} \frac{\sigma(o,d \mid v)}{\sigma(o,d)} \tag{5}$$

where $\sigma(o,d \mid v)$ is the weighted sum of the set of shortest paths between o and d which contain v and $\sigma(o,d)$ is the weighted sum of the set of shortest paths between o and d. Betweenness centrality reflects the relative volume of use for a node in a graph (and may also be computed for edges). The Denver airport has high betweenness centrality as part of the US domestic aviation network.

4.4 Information Centrality

Conceptually the information content of a node on a graph is the path information gained after leaving that node. There may be many paths on a graph which pass through a given node and thus an expectation must be computed. The following derivation is based on [1] which is the original publication on information centrality.

For each O/D pair in $o,d \in V$ there is a set of paths $P_{od} = \{P_{od}(1), P_{od}(2), \ldots, P_{od}(n_{od})\}$ with associated costs C_{od}. Picking $o = A$ and $d = B$ from the graph in Figure 4, P_{AB} contains $P_{AB}(1) = A-B$, $P_{AB}(2) = A-E-B$, and $P_{AB}(3) = A-D-C-B$. The costs of the paths will be $C_{AB}(1) = 1$, $C_{AB}(2) = 2$, and $C_{AB}(3) = 3$ for the unweighted graph and $C_{AB}(1) = 2$, $C_{AB}(2) = 10$, and $C_{AB}(3) = 12$ for the weighted graph. Under the assumption that lower-cost paths are more likely to be taken and knowing that information is inversely proportional to probability it should be apparent that the information content of the indirect paths is greater than that of the direct path. It should also be apparent that the edge weights increase the information disparity. Consider $I_{od}(i) = C_{od}(i)^{-1}$ as an analogue for information content (this is used rather than (1) for computational simplicity). Entropy for node o with respect to node d can thus be computed as

$$S_{od} = \sum_{i=1}^{n_{od}} W_{od}(i) C_{od}(i) \tag{6}$$

where

$$W_{od}(i) = \frac{I_{od}(i)}{\sum_{i=1}^{n_{od}} I_{od}(i)} \tag{7}$$

Generalizing path entropy to node entropy means considering how node o interacts with all possible O/D pairs which. For the sake of computational efficiency, this generalization can be formulated into a matrix-operations based process as in [1]. Information centrality for the nodes in $G = \{V, E\}$ is based on the inverse incidence matrix B whose elements are

$$b_{ij} = \begin{cases} 1 & (i,j) \notin E \\ 1 - \eta(i,j) & otherwise \end{cases} \tag{8}$$

$$b_{ii} = 1 + \sum_{(i,j) \in E_i} \eta(i,j) \tag{9}$$

where $\eta(i,j)$ is a function returning edge costs. The diagonal elements of B are the degrees of V plus self-connection and the non-diagonal elements are 1 if no direct connection exists and less than 1 where one does exist. Rows and columns of B all have the same sum if G is reciprocal. The information for each node can be computed form the incidence matrix $C = D^{-1}$ as

$$\Omega_i(v) = \frac{n}{nc_{vv} + \sum_{u=1}^{n} c_{uu} - 2\sum_{u=1}^{n} c_{vu}} \tag{10}$$

Information centrality for a node should be interpreted as being, roughly, the degree to which that node is determinative of future nodes on a path between an unknown origin and destination. A node with high information content is a node which both appears on many lowest-cost paths and one which, generally, limits the set of paths after it. As an example, the phrase "medium-rare" may be preceded by many words but will usually be followed by a type of red meat. Information centrality should be thought of as the opposite of network entropy.

4.5 Comparison

The methods of computing centrality are compared for the simple graph shown in Figure 4.

Table 1: Centrality for simple undirected and unweighted example graph

Node	Degree Centrality	Betweenness Centrality	Closeness Centrality	Information Centrality
A	0.75	0.25	0.8	0.355
B	0.75	0.25	0.8	0.355
C	0.5	0.083	0.667	0.282
D	0.5	0.083	0.667	0.282
E	0.5	0	0.667	0.275

Table 2: Centrality for simple undirected and weighted example graph

Node	Degree Centrality	Betweenness Centrality	Closeness Centrality	Information Centrality
A	0.75	0.5	0.364	0.667
B	0.75	0.5	0.364	0.667
C	0.5	0	0.286	0.535
D	0.5	0	0.286	0.535
E	0.5	0	0.182	0.646

Concerning the isomorphs in Figure 4, intuition would say that nodes A and B are most central and this is supported by all centrality measures presented. The principle effects of considering edge weights are that the shortest paths between C and D no longer utilize (C,D), all shortest paths include A or B, and 3 include (A,B). The differences are not attested to by degree centrality which is identical among isomorphs. The increased reliance of the network on A and B is reflected in the remaining centrality metrics. The lower diversity of paths for the weighted graph is reflected in lower nodal information content.

The manner in which the example graph edges are weighted leads to an interesting result. Without explicit weights all edges are weighted equal and, thus, the likelihood of a path is modeled as the inverse of the number of edges weighted. Numerically, the unweighted case is equivalent to a weighted case where all weights are 1. When the weights of greater than 1 are added the graph entropy is seen to increase. This is a result of the assumption that path likelihood is inversely correlated with path distance. Because the average path distance increased with the explicit weights the system entropy went up even though the weights serve to limit the viability of alternate paths. In order to avoid this issue weights can be normalized by the sum of weights in the graph. Using normalized edge weights a more apt comparison between the graphs can be made as shown in Table 3.

Table 3: Normalized-weight information centrality for simple undirected and weighted example graph

Node	Unweighted graph	Weighted graph
A	0.296	0.139
B	0.296	0.139
C	0.235	0.112
D	0.235	0.112
E	0.229	0.135
Sum	1.291	0.635

5 Topographical Entropy Maximization

Entropy maximization is any system parameter optimization which has the goal of maximizing system entropy. As an example, the entropy of a series of coin flips is maximized if the coin is "fair" with $p(h) = 1 - p(t) = 0.5$. For a graph, entropy is maximized where information centrality is minimized. As shown in Tables 1 and 2, changing the edge weights of Figure 4 results in different levels of entropy. In theory, a graph may have no identifiable most-central node. The most obvious form of such a graph would be one where only the "outer" links are present (a triangle for example). The end result would be equal information centrality for each node. Note, however, that the nodes of a triangle would have higher information centrality than those of a square etc. due to the greater optionality of the higher-sided shapes.

Entropy maximization can be readily applied to transportation planning. Take, for example, the road network diagrammed in Figure 5.

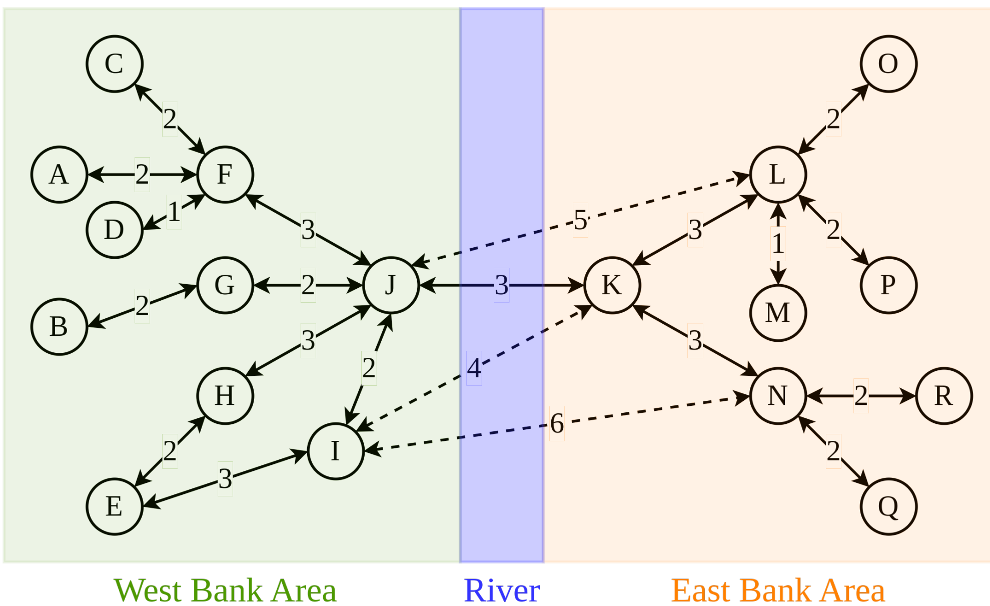

Figure 5: Example road network for city divided by a river. The solid links are present and dashed links are proposed additional links.

The example city is, roughly, bifurcated by a river and the east and west bank areas are connected by a solitary bridge (J, K). It is evident that this bridge will feature in all shortest paths for O/D pairs on opposite sides of the river and is, thus both congested and a single point of failure for the network. Planners, thus, want to add at least one more bridge and have proposed the dashed links (I, N), (I, K), and (J, L). Information centrality for the network subject to the proposed additional bridges is listed in Table 4.

Table 4: Information centrality for example road network subject to bridge additions

Node	Baseline	(I, N)	(I, K)	(J, L)
A	0.0089	0.0083	0.0085	0.0088
F	0.0131	0.0127	0.0128	0.0137
B	0.0078	0.0072	0.0074	0.0076
G	0.0108	0.0102	0.0104	0.0109
C	0.0089	0.0083	0.0085	0.0088
D	0.0068	0.0062	0.0064	0.0065
E	0.0114	0.0124	0.0126	0.0117
H	0.0128	0.0129	0.0129	0.0133
I	0.0120	0.0146	0.0149	0.0123
J	**0.0162**	**0.0162**	**0.0161**	**0.0177**
K	**0.0155**	**0.0153**	**0.0162**	**0.0157**
L	0.0126	0.0122	0.0128	0.0164
N	0.0121	0.0147	0.0123	0.0119
O	0.0087	0.0081	0.0085	0.0099
P	0.0087	0.0081	0.0085	0.0099
M	0.0066	0.0061	0.0064	0.0071
R	0.0085	0.0092	0.0083	0.0081
Q	0.0085	0.0092	0.0083	0.0081
Sum	**0.1899**	**0.1920**	**0.1917**	**0.1984**

Besides the obvious conclusion that adding a second bridge increases entropy, a few interesting observations can be made. First, that the (J, L) bridge increases overall entropy more than the (I, N) bridge both for the network in general and for node J in particular. Second that bridge (I, N) similarly outperforms bridge (I, K). Both can be explained by looking at the structure of the network in the east bank area. Because node I is fairly isolated compared to node J, extra bridges from node I provide less optionality than those from node J. The same can be said for nodes N and K.

Recall that entropy is the expectation of information for a set of events. Entropy is greater for a larger number of less likely events compared to a smaller number of more likely events. For example, the entropy of a "fair" coin flip is 1 bit while the entropy of a "fair" dice roll is 2.585 bits. The intuition of the results of the road network example is that bridge (J, L) was most effective in creating sub-optimal but plausible alternative routes thus reducing the load on (J, K). The maximal case of entropy in the network would occur if each node was incident to each other node by the same distance but this is not possible because the network nodes are at physical locations.

6 EVSE Entropy Maximization

EVSE location optimization is a topic which has produced, continues to produce, and will continue to produce ludicrous volumes of research very little of which informs deployment in any real way. This situation is the result of the fact that EVSE infrastructure is insufficient in many ways and is perceived to serve as a impediment to Plug-in Electric Vehicle (PEV) sales growth meaning that the large body of optimization and data analysis focused researchers are able to fund projects and publish papers related to the topic. In the most general sense, optimization of EVSE locations is not an especially unique problem and most seek to optimize the allocation of scarce development resources to minimize a given cost function. A common theme of these papers is the goal of optimizing efficiency in metrics which are either business or consumer oriented. Robustness is often treated as a constraint if considered at all. This line of thinking has led to large concentrations of EVSE in those areas most dense with PEV ownership and a number of subsidized charging stations located at arbitrary intervals along major transportation corridors. Evenly spaced chargers along corridors is not necessarily a bad solution but it is an arbitrary one.

Consider, instead, how redundancy might be computed from and informational perspective. As a driver embarks on a long trip (one that requires charging), the driver will have a charging window. The charging window starts at a distance where a substantial portion of the vehicle's range is used and will end when the driver is close enough to the destination that additional range is not required. For example, if a car has 300 km of range and is undertaking a 400 km trip at least one charge is required. However there will be little region to charge in the first 100 km because doing so will mean having to charge twice. For the same reason the last 100 km is unlikely to see a charge event. Thus, the charging window for the trip is the 200 km in the middle. If chargers were evenly spaced in 50 km intervals along the driver's route then of the 7 chargers available the middle 5 would be viable options. Since the driver only needs to charge once there is real redundancy built in. At the beginning of the trip there are 5 shortest paths that the driver must choose between. If the driver passes the first middle charger without charging then the number of shortest paths goes from 5 to 4, if the driver charges at the first middle charger then the number of shortest paths goes from 5 to 1. In this scenario charging provides more information than not charging and this is not surprising, stopping to charge will be less common than passing a charger.

In the above example the trip origin and destination as well as the vehicle range are known allowing for the definition of a charging window. In practice, these quantities can only be imputed stochastically. The basic principle, however, can be generalized:

Lemma 1 (Charger-Charger Edge Energy Cost Upper Limit)**.** *Under normal circumstances, vehicles will not travel from charger u to charger v if the edge energy cost $d(u,v)$ is greater than a given upper limit L^U which is less than or equal to the vehicle's maximum energy storage capability.*

Corollary 1.1. *Chargers u and v are incident if and only if $d(u,v) < L^U$.*

Lemma 2 (Charger-Charger Edge Energy Cost Lower Limit)**.** *Under normal circumstances, vehicles will not travel from charger u to charger v if the edge energy cost $d(u,v)$ is less than a given lower limit L^L which is greater than or equal to the vehicle's minimum energy storage capability.*

Corollary 2.1. *Chargers u and v are incident if and only if $d(u,v) > L^L$.*

Lemma 3 (Charger-Place Edge Energy Cost Upper Limit)**.** *Under normal circumstances, vehicles will not travel from charger u to place v if the edge energy cost $d(u,v)$ is greater than a given upper limit L^L which is less than or equal to the vehicle's minimum energy storage capability. The same holds for place-charger edges.*

Corollary 3.1. *Charger u and place v are incident if and only if $d(u,v) < L^U$.*

Lemma 4 (Place-Place Edge Energy Cost Upper Limit)**.** *Under normal circumstances, vehicles will not travel from place u to place v if the edge energy cost $d(u,v)$ is greater than a given upper limit L^L which is less than or equal to the vehicle's minimum energy storage capability.*

Corollary 4.1. *Place u and place v are incident if and only if $d(u,v) < L^U$.*

In other words, chargers relationships with each other are determined by distance. If two chargers are very far apart then it is unlikely that a car will travel from one to the other without having to charge at a third location. If two chargers are close together a vehicle may use either on a given trip but is unlikely to use both. The constants L^L and L^U define the span of incidence. Chargers themselves serve to allow for trips between places and, thus, will only be visited for the purposes of charging. Places can be visited at any State of Charge (SOC) so there is no minimum edge cost for incidence if the edge connects at least one place node.

References

[1] Karen Stephenson and Marvin Zelen. "Rethinking centrality: Methods and examples". In: *Social Networks* 11.1 (1989), pp. 1–37. ISSN: 0378-8733. DOI: https://doi.org/10.1016/0378-8733(89)90016-6. URL: https://www.sciencedirect.com/science/article/pii/0378873389900166.