

ORGB 672:
Exercise 2 – Where to sit on the bus

ORGB 672

Dr. Roman Galperin

March 18th, 2023

Group 4:

Emery Dittmer 260658030

1 Context & Modeling

1.1 Problem Description

For this problem, we will be a summer intern at Fakebook. This intern takes a bus every morning from San Francisco to Menlo park. When he boards the bus, there are 4 empty seats (labelled A-D). However, not all of these seats are equal.

Anyone on the bus can form connections with their nearest neighbors who are in front, behind, to the side or diagonal from each other. Our goal is to sit in the seat that is the most advantageous to us. Let's assume that seats with a lot of contact or centrality will be the most advantageous. With this example, we will examine network centrality to determine which seats have the most prominent centrality.

We have the following image to base our network and centrality measures off of.

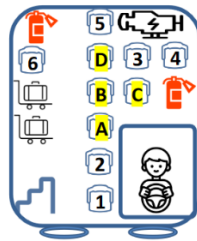


Figure 1: Problem setup: Bus seating configuration and potential seating options. Seats A-D (yellow highlighted) are available, while seats labelled 1-6 are taken.

1.2 Assumptions

To simplify the problem, we will use several assumptions that are listed here:

1. We will use a grid to model this problem with 3 types of seats:
 - a. No Seat: defines a position on the bus that is not a seat. The cabin for the driver or engine take up these spots.
 - b. Occupied: a seat that exists but is currently occupied
 - c. Available: an available seat to be sat in. These are the choice nodes we have.
2. We will assume that all the seats will be occupied. Therefore, we will set the weight of each edge from available seats to 100%, even if the seat is currently available.
3. We will assume that the alley for walking that divides the seats does not exist. We will not need to account for the small extra distance between seats D and 6.

1.3 Data Collection

No csv, or other data exists; however, based on the image above, we can form a coordinate grid to model the bus. A 4x6 grid will model the bus sufficiently for our purposes. However, some coordinates in this grid are not actual seats; worse still, some are already occupied.

To correctly represent the situation, we need to label the seats and remove unneeded seats. We will label the seats as either *available*, *occupied* or *no seat* to differentiate the seats (nodes) within the bus (network). Then we will remove the points that do not actually sit. Lastly, we will index or create arbitrary seat ids for modelling purposes.

After some data preparation, we can visualize our results.



Figure 2: Raw 4 X 6 grid of bus. No adjustments or labeling of seats

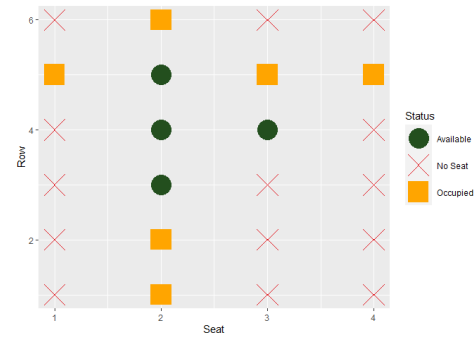


Figure 3: 4 X 6 grid of bus with labelled seats.

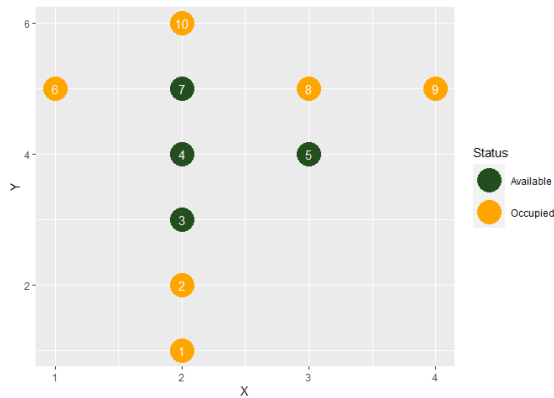


Figure 4: Raw 4 X 6 grid of bus with labelled seats and removed 'no seat' labels. Numbers are the seat IDs, these are arbitrarily assigned and needed for modelling purposes.

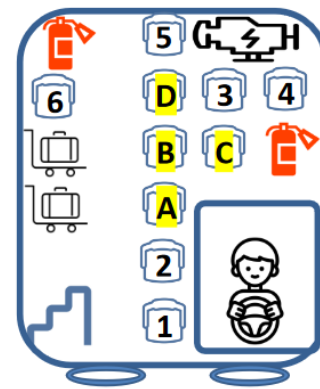


Figure 5: Problem setup: Bus seating configuration and potential seating options. Seats A-D (yellow highlighted) are available, while seats labelled 1-6 are taken.

1.4 Network Analysis Approach

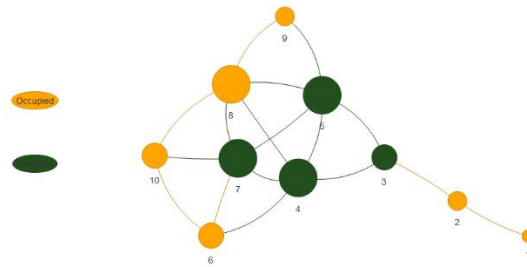
Based on the grid points defined above, the nodes are the seats, and the edges are the connections between each seat. The code transforms the data in several ways to make the code reusable and more dynamic. First, we calculate a distance matrix between each seat or node. For instance, we calculate the distance between seats 1 and 2 , then 1 and 3. These results are displayed in a matrix of distances.

	1	2	...	9	10
1	0.000000	1.000000	...	4.472136	5.000000
2	1.000000	0.000000	...	3.605551	4.000000
...
9	4.472136	3.605551	..	0.000000	2.236068
10	5.000000	4.000000	..	2.236068	0.000000

Table 1: Matrix of Distances Between Seats

The matrix is then converted to a table of distances. However, the table contains 3 redundancies: the self-distance, the out-of-scope distance, and reciprocal distances.

After removing the self-distances, out-of-scope and reciprocals, the nodes and edges were created and fed into the *Visnetwork* to verify the validity of the network.



A scatter plot showing the status of 10 people across 4 categories (X-axis) and 6 levels (Y-axis). The legend indicates 'Available' (green) and 'Occupied' (orange).

ID	X	Y	Status
1	2	1	Occupied
2	2	2	Occupied
3	2	3	Available
4	2	4	Available
5	3	4	Available
6	1	5	Occupied
7	2	5	Available
8	3	5	Occupied
9	4	5	Occupied
10	2	6	Occupied

1. Degree Centrality
 - a. The count of the number of links each node has to other nodes.

For instance, seat *A* (labelled as 3 above) has a degree centrality of 3 since it is connected to 3 other nodes: 2, *B* & *C* (*B* labelled as 4 and *C* labelled as 5 above)

2. Closeness Centrality

- a. A measure that calculates the ability to spread information efficiently via the edges the node is connected to. It is calculated as the inverse of the average shortest path between nodes.

For instance, for node *A* (labelled 3), the closeness is $\frac{1}{(1+2+1+1+2+2+2+3)} = 0.0625$. The higher the number, the closer the node is to the center based on distance. See appendix For details.

3. Betweenness Centrality

- a. A measure that detects a node's influence over the flow of information within a graph. This is the sum of the shortest paths between two points *i* and *j* divided by the number of shortest paths that pass-through node *v*.

Seat	Centrality Degree	Closeness Degree	Betweenness Degree
B	5	0.07142857	9.033333
C	5	0.07142857	8.600000
D	5	0.06250000	3.266667
A	3	0.06250000	14.000000

3 Discussion

3.1 Seat choice

While we have measured each seat's centrality and plotted the network diagram, we need to consider the consequences of the seat choice.

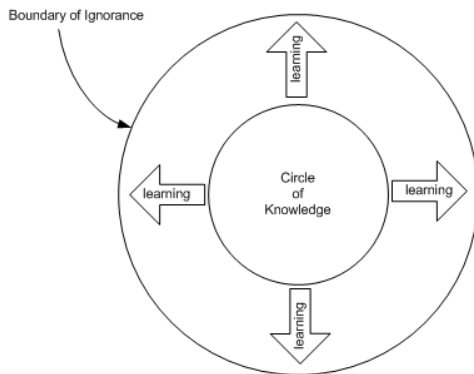
The primary goal is to leverage this opportunity to form connections. The connections will likely become valuable when looking for future employment, future progression or to have a colleague/friend you can rely on. We will aim to pick a seat that has connections with people. A seat without any links isolates us and removes us from the network.

The potential consequences of the seat selection are a network size that may be smaller or larger, a potential utility within the network (conveyor of information) and recognition. In other words, if a seat has more connections, your possible network is larger than other seats. If your seat is between two friends, you will be in the middle of their conversation or convey information and thus become associated with the network. For instance, seat 3 in the problem has side to side connection with seat *D* and 4, whereas seat 4 is only connected to seat 3

From this perspective, there are two intuitive solutions: create the most significant number of relationships or create a few strong connections. These two perspectives are equally valid.

We can be in a seat with the greatest number of connections, thereby becoming friendly with many people or choose a specific seat that allows us to make fewer connections. The benefit of picking a seat with fewer connections is that you grow the strength of your network. A strong network gives you access to a more intimate side of friends, who can help with roles, advocate for you or serve as mentors. From a growth perspective, these are valuable people in a network and are likely to help grow a network.

The tradeoff of a smaller network is that while your connections may be strong, there are fewer of them, and your network will be smaller. A larger network of “weak” connections is a tradeoff. Implicitly strong connections sound more desirable, but counterintuitively weak ties have more power for securing future roles, according to David Easley and Jon Kleinberg. We can borrow from the circles of knowledge and boundary of ignorance to explain this. As the circle of friends grows so does the boundary of friendship (friends of friends); assuming little overlap adding additional friends gives you access to a much larger network.



© 2006 - Jim McGee - <http://www.mcgeesmusings.net>

Figure 9: Circle of knowledge and boundary of ignorance. Growth is logarithmically proportional to learning. [Source](#)

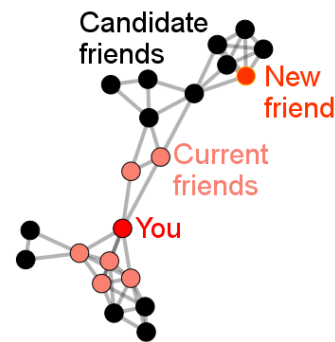


Figure 10: Growth of network illustration. [Source](#)

Therefore, based on the goal and the available seats, the best seat to take for this bus ride will be the one that maximizes the number of connections. Based on the centrality scores, **seat B is the best seat to take**. This seat is the most connected based on degree centrality and has a good balance of closeness and betweenness centrality. In this case, betweenness and connectedness centrality are not as important as degree centrality, as forming connections relies more on physical distance as represented by degree centrality.

This, however, relies on the assumption that seats *D*, *C* and *A* are filled. If this assumption does not hold, then seat *D* is the best to pick.

4 Graph

Finally, we can visualize all nodes in the network with their respective labels and centrality degrees.

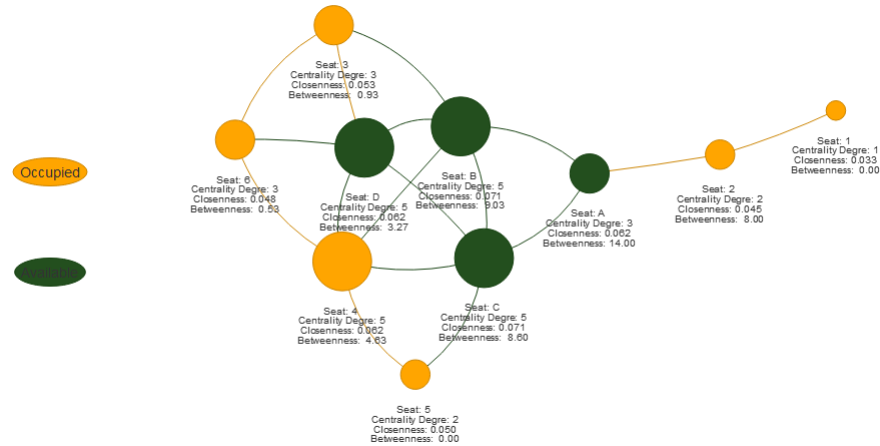


Figure 11: Network of bus seats visualized. Green are available seats, orange are taken seats.

5 Appendix

5.1 Centrality Calculations

5.1.1 Degree Centrality

The following equation defines the calculation of degree centrality.

5.1.2 Closeness Centrality

The following equation defines the calculation of closeness.

$$\frac{1}{\sum_{i \neq j} d_{ij}}$$

Where d_{ij} is the shortest distance between nodes i and j .¹

5.1.3 Betweenness Centrality

The following equation defines the calculation of betweenness.

$$\sum_{i \neq j} \sum_{i \neq v} \sum_{j \neq v} \frac{g_{ivj}}{g_{ij}}$$

Where g_{ij} is the total number of shortest paths between nodes i and j while g_{ivj} is the number of those shortest paths which pass through node v .²

5.2 Closeness Centrality

From	To	Distance
3	1	2
3	2	1
3	4	1
3	5	1

¹ <https://search.r-project.org/CRAN/refmans/igraph/html/closeness.html>

² <https://search.r-project.org/CRAN/refmans/igraph/html/betweenness.html>

	3	6	2
	3	7	2
	3	8	2
	3	9	2
	3	10	3
Total			16

Table 2: Distances between seat A (seat ID=3) and all other graph points. We are measuring closeness centrality.