Assignment 2

Emery Dittmer

2023-03-09

Excercise #2

For this exercise we are investigating the idea of centrality in networks. We will look at what seat or position that we should sit on for the bus ride to Fakebook from downtown San Francisco! We can network with the people in our immediate area but not outside of that.

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.

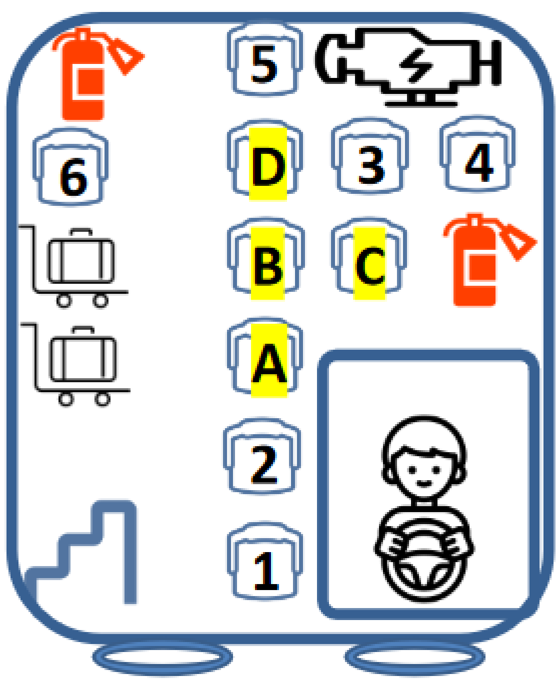


Figure 1: Bus Network Illustartion.

1.2 Assumptions

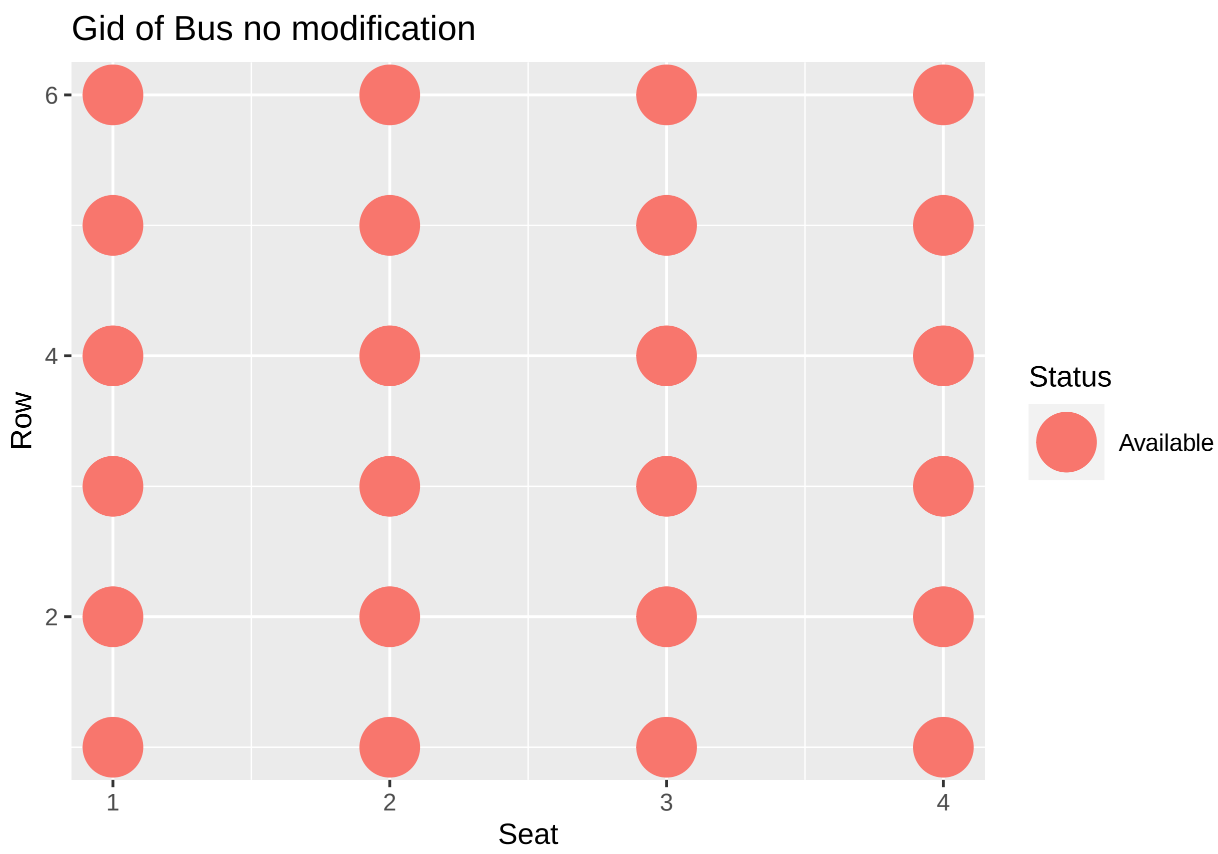
To simplify the problem, we will use several assumptions that are listed here: \ -1. We will use a grid to model this problem with 3 types of seats: \a. No Seat: defines a position on the bus that is not a seat. The cabin for the driver or engine take up these spots. \b. Occupied: a seat that exists 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.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8     v purrr   0.3.4
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'igraph'
##
##
## The following objects are masked from 'package:purrr':
##
##     compose, simplify
##
##
## The following object is masked from 'package:tidyr':
##
##     crossing
##
##
## The following object is masked from 'package:tibble':
##
##     as_data_frame
##
##
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
##
##
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
##
```

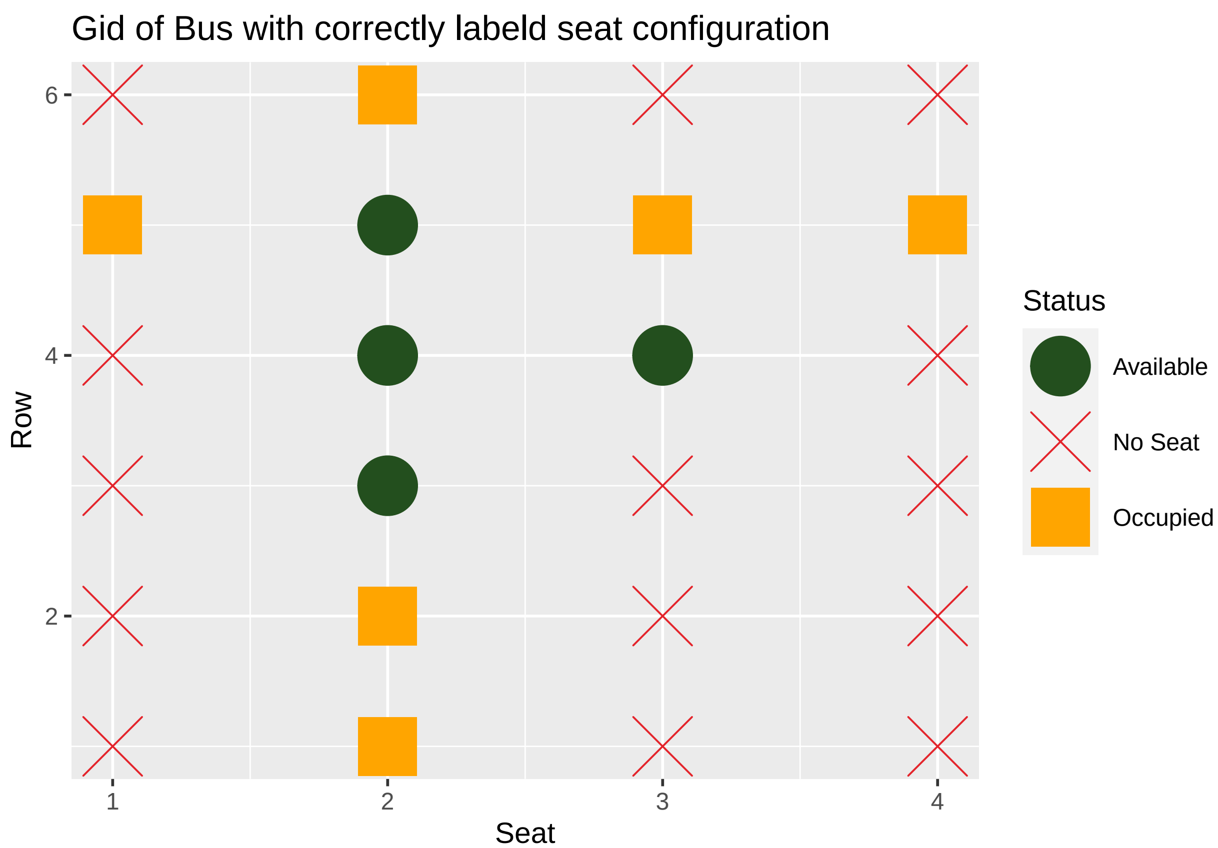
```
##
## The following object is masked from 'package:base':
##
##     union
##
##
##
## Attaching package: 'tidygraph'
##
##
## The following object is masked from 'package:igraph':
##
##     groups
##
##
## The following object is masked from 'package:stats':
##
##     filter
```

1.3 Data Collection: Making a bus coordinate system

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.

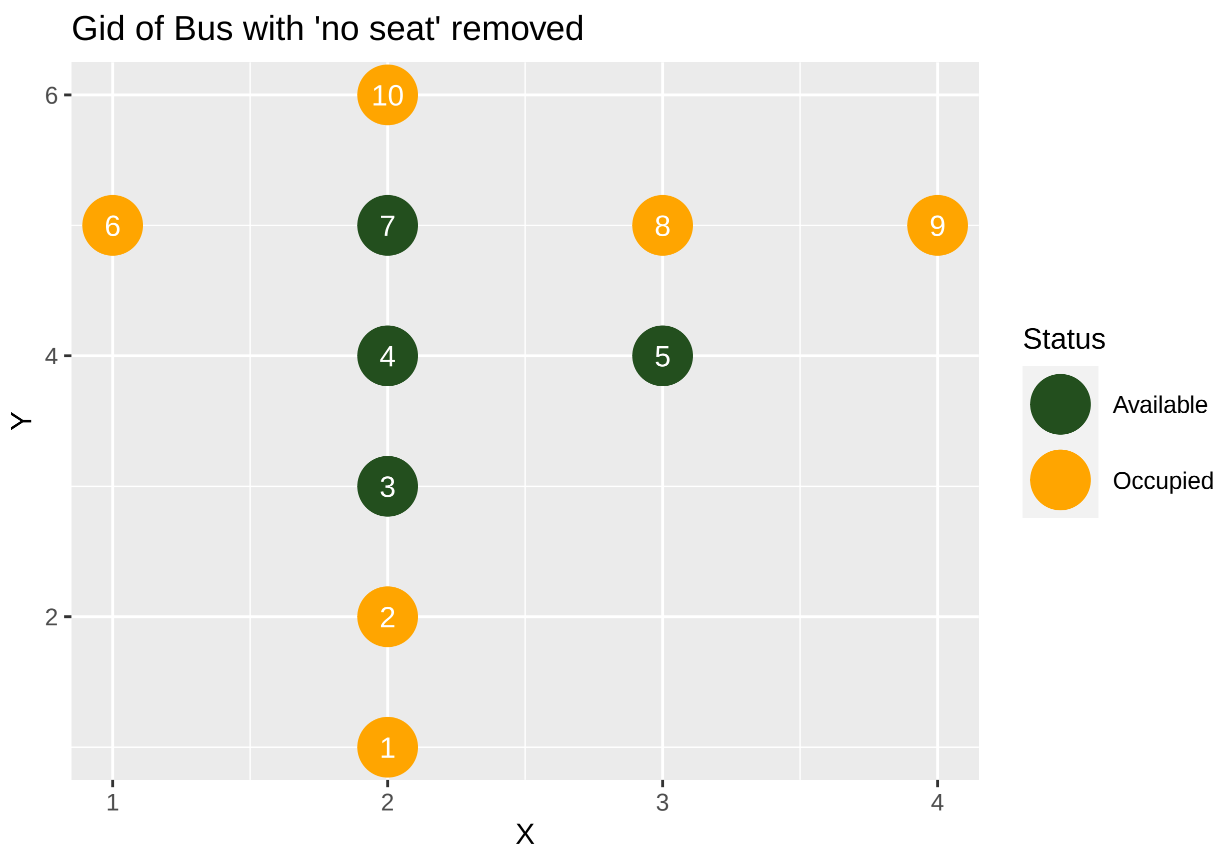


Let's take a look at our data!



Looks like we have our bus, the seats available and taken! Now lets filter our data frame to have only the useful coordinates, or seats that exist.

let's remove the seats that do not actually exist.



We have a simplified coordinate system with the existing seats. We will need all of this information to compute the degree of centrality for each seat, which we can then filter out.

1.4 Transformations to Data

Centrality indicates the influence of a node in a network. Higher centrality means higher influence. Therefore for this problem we would want higher centrality. Special considerations and data transformation

Distance Matrix Transformations

We will need to look at the distance between each seat to see which seats can form connection with others. Ultimately we will find the most central in our network.

1	2	3	4	5	6	7	8	9	10
0.000000	1.000000	2.000000	3.000000	3.162278	4.123106	4.000000	4.123106	4.472136	5.000000
1.000000	0.000000	1.000000	2.000000	2.236068	3.162278	3.000000	3.162278	3.605551	4.000000
2.000000	1.000000	0.000000	1.000000	1.414214	2.236068	2.000000	2.236068	2.828427	3.000000
3.000000	2.000000	1.000000	0.000000	1.000000	1.414214	1.000000	1.414214	2.236068	2.000000
3.162278	2.236068	1.414214	1.000000	0.000000	2.236068	1.414214	1.000000	1.414214	2.236068
4.123106	3.162278	2.236068	1.414214	2.236068	0.000000	1.000000	2.000000	3.000000	1.414214
4.000000	3.000000	2.000000	1.000000	1.414214	1.000000	0.000000	1.000000	2.000000	1.000000
4.123106	3.162278	2.236068	1.414214	1.000000	2.000000	1.000000	0.000000	1.000000	1.414214
4.472136	3.605551	2.828427	2.236068	1.414214	3.000000	2.000000	1.000000	0.000000	2.236068
5.000000	4.000000	3.000000	2.000000	2.236068	1.414214	1.000000	1.414214	2.236068	0.000000

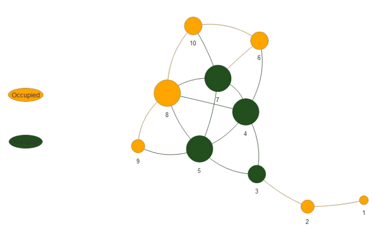
from_seat_id	X	Y	Status	to_seat_id	Distance
1	2	1	Occupied	2	1.000000
1	2	1	Occupied	3	2.000000
1	2	1	Occupied	4	3.000000
1	2	1	Occupied	5	3.162278
1	2	1	Occupied	6	4.123106
1	2	1	Occupied	7	4.000000

This is just a sample, but the table overall contains all the distances between seats.

Now we have the distance between each of the available seats and the taken or occupied seats. We just need to apply the rules of connections (diagonal, front,back, ect) and we will be able to summarize the table to get the strength of each seat based on the connections. We will filter all of the connections who are further than sqrt 2 away from the current seat

from	X	Y	Status	to	Distance
1	2	1	Occupied	2	1
2	2	2	Occupied	1	1
2	2	2	Occupied	3	1
3	2	3	Available	2	1
3	2	3	Available	4	1

Now we have the final distance tables. We can begin the transformation of data to edges and nodes.



Now let's try to find the measures for each seat ## 2. Centrality Measures

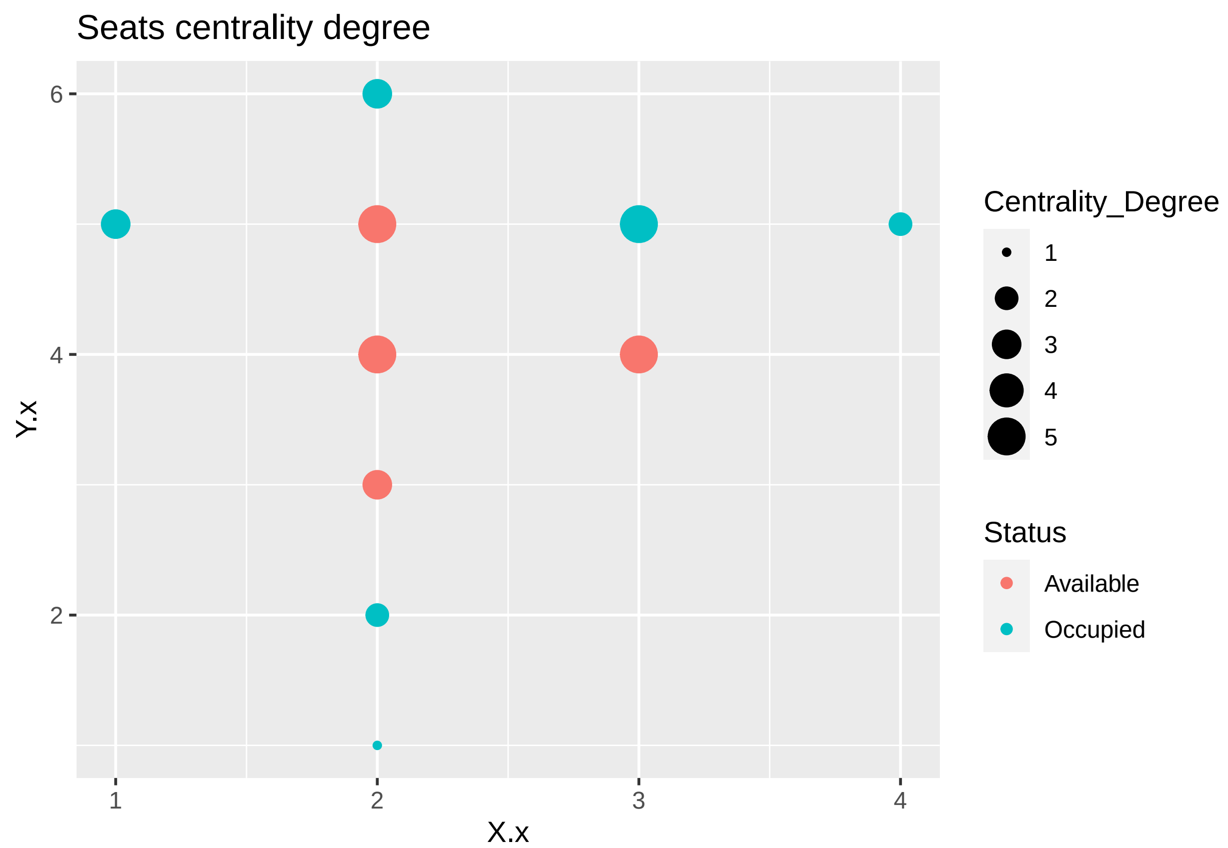
2.1 Degree Centrality

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)

Seat	Degree_Centrality
B	5
C	5
D	5
A	3

We can validate this with the igraph package wich has a built in functionality for centrality degree

Seat	Centrality_Degree
B	5
C	5
D	5
A	3



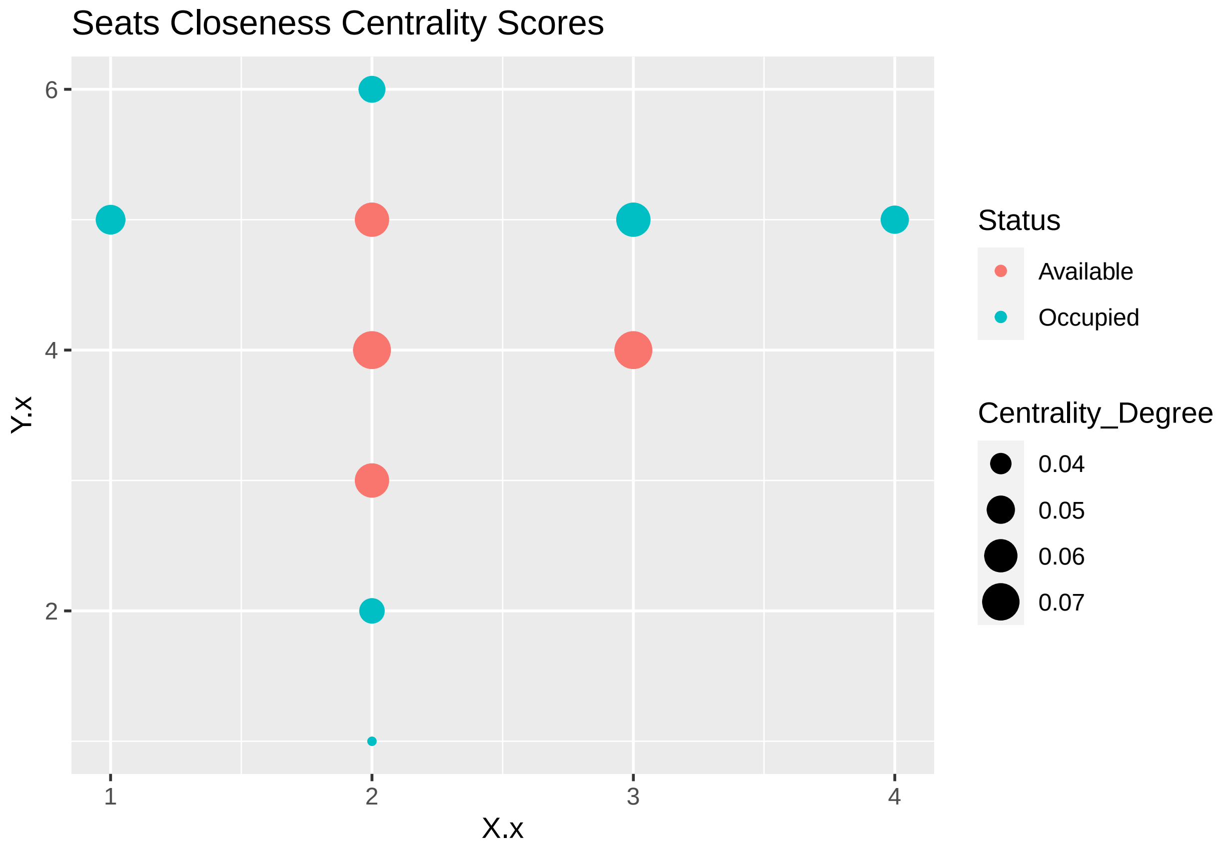
There is agreement between our calculations and the calculations for the package therefore we can use them!

2.2 Closeness centrality

```
A measure that calculates the ability to spread information efficiently via the edges the node is conne
```

For instance, for node A (labelled 3), the closeness is 1/((1+2+1+1+2+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

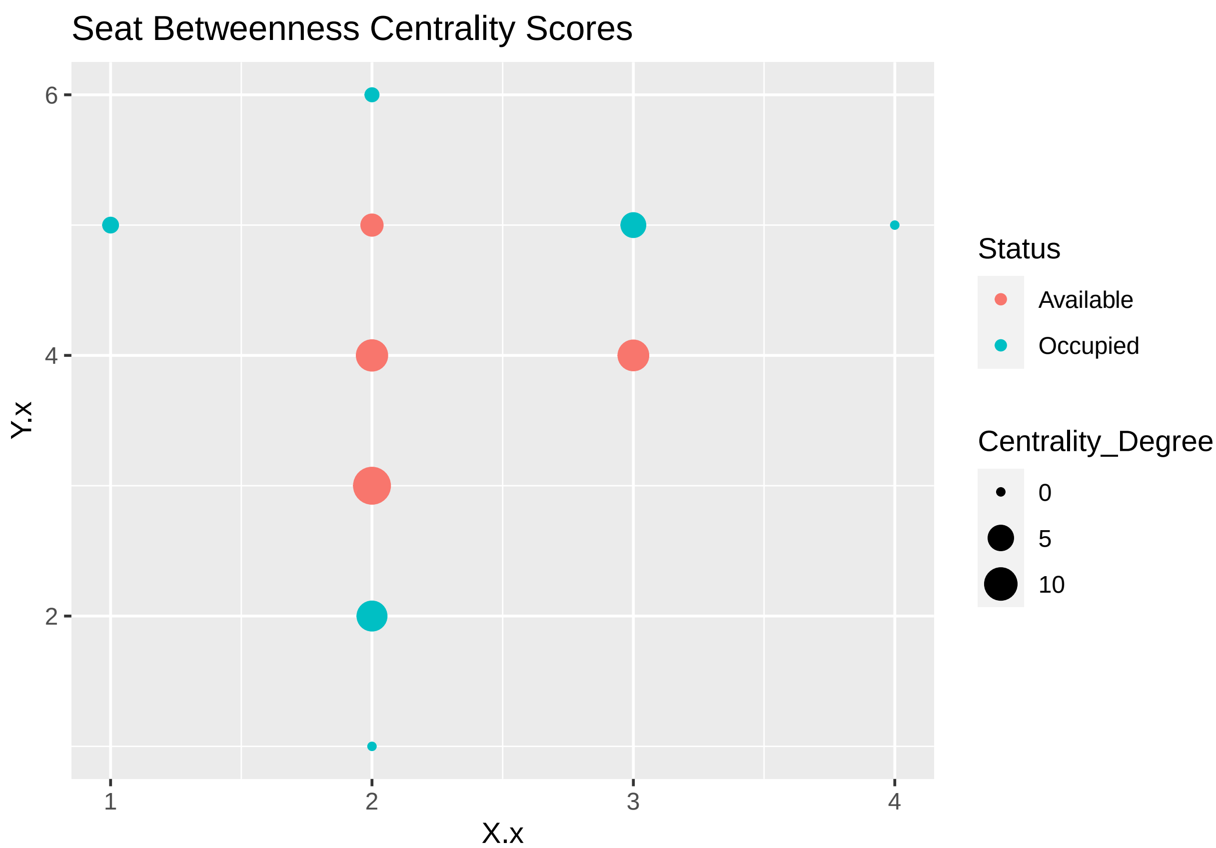
Seat	Closeness_Degree
B	0.07142857
C	0.07142857
A	0.06250000
D	0.06250000



2.3 Betweenness centrality

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	Betweenness_Degree
A	14.000000
B	9.033333
C	8.600000
D	3.266667



Comparison between all 3!

Let's compare the centrality of all 3 measures

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

It looks like Seat B may be the best

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

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.

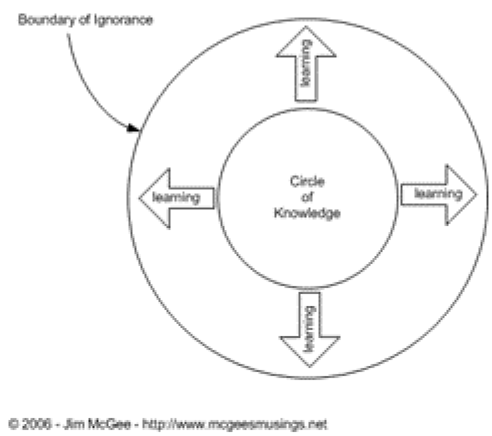


Figure 2: circle of knowledge and boundry of ignorance.

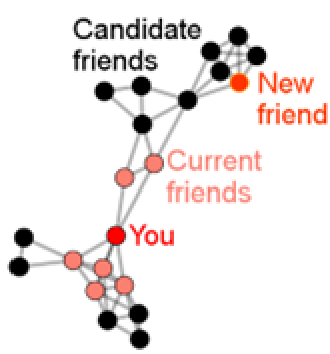


Figure 3: Growth of network illustration

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. Final Visulizations

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