Analyzing the Best Starting Country in Plague Inc Using Graph and Breadth First Search

Satya Radhityan Yahya - 13525098
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail: satyaradhit1@gmail.com, 13525098@std.stei.itb.ac.id

***Abstract*—This research of a video game Plague Inc to answer what is the best country to start your plague in by representing country as nodes and their respective connection with other countries as edges .**

Keywords—Plague Inc; country, transmission, graph theory, distance.

I. Introduction (*Heading 1*)

Plague Inc is a video game where the goals to winning the game is to infect the whole world with a plague and eventually eradicate them. There is 58 different countries in the game that the player need to infect in order to win the game. At the start of the game, player will choose their starting country to start their plague. After that, the player will need to infect other country with 3 different transmission to use (airport, seaport and land borders).

Each country have a different ways to transmit the plague, some country have an airport while other don't, some country border other countries while island countries don't, etc. This differences makes choosing the starting country a crucial part of the game.

Fig 1. Plague Inc world map, Source: Official Steam page.

By representing countries as nodes and their respective transmission as edges, then we can determine what is the best starting country by calculating total distance from the starting country to all other countries.

II. Theoretical Framework

A. Graph Theory

The theoretical foundation of this study is rooted in network epidemiology, where a global landscape is modeled as an undirected, unweighted graph denoted as $G = (V, E)$. Here, the vertex set V represents the 58 country within the game, while the edge set E encompasses the union of a land border transmission, plane transmission and ship transmission between two different countries.

B. Breadth First Search

Breadth First Search (BFS) is used to calculate the structural proximity of each starting country to the rest of the world. BFS is uniquely suited for unweighted graph as it systematically explores adjacent vertices layer-by-layer, guaranteeing the calculation of the absolute shortest path length, or geodesic distance $d(u,v)$, between a source node u and a target node v.

C. Closeness Centrality and Eccentricity

Closeness Centrality (C_c) and Eccentricity (ϵ) is the parameters used to determine which starting country is the best one. Closeness Centrality denotes how close a country is with all other countries. A country with higher C_c need less total distance to infect the world.

$$C_C(\mathrm{u}) = \frac{|\mathrm{V}| - 1}{\sum_{\mathrm{v} \in \mathrm{V}} \mathrm{d}(\mathrm{u}, \mathrm{v})}$$

While Eccentricity is the max distance from the starting country to all other countries. Best starting country need to have as less ϵ as possible.

$$\epsilon(u) = max(d(u, v)), v \in V$$

III. Methodology

This research will be divided into 3 main method. Data collection and mapping, distance measurement and parameters evaluation.

A. *Data Collection and Adjacency Matrix Mapping*

The set of all countries and their respective transmission is extracted from the game and then mapped in the form of adjacency binary matrix 58×58. This matrix is symmetric because all form of transmission is bidirectional, where 0 (colored red) denotes no form of transmission and 1(colored green) denotes either one of three form of transmission is exist between two countries.

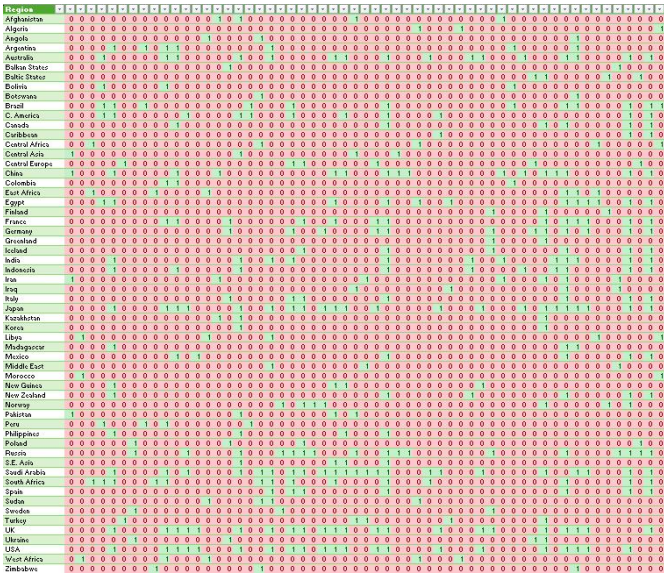

Fig 2.Adjacency binary matrix of connectivity between two countries, Source: Author.

B. *Distance Measurement (BFS)*

These adjacency matrix then will be converted into .csv files to then be inserted into python code with BFS algorithm to get the distance matrix.

```python
import pandas as pd
import networkx as nx

df = pd.read_csv('CountryConnections.csv', index_col=0)
G = nx.from_pandas_adjacency(df)
all_pairs_shortest_path = dict(nx.all_pairs_shortest_path_length(G))
distance_matrix = pd.DataFrame(all_pairs_shortest_path).sort_index().T.sort_index()
distance_matrix.to_csv('TotalDistance.csv')
```

Fig 3. Python code, Source: Author.

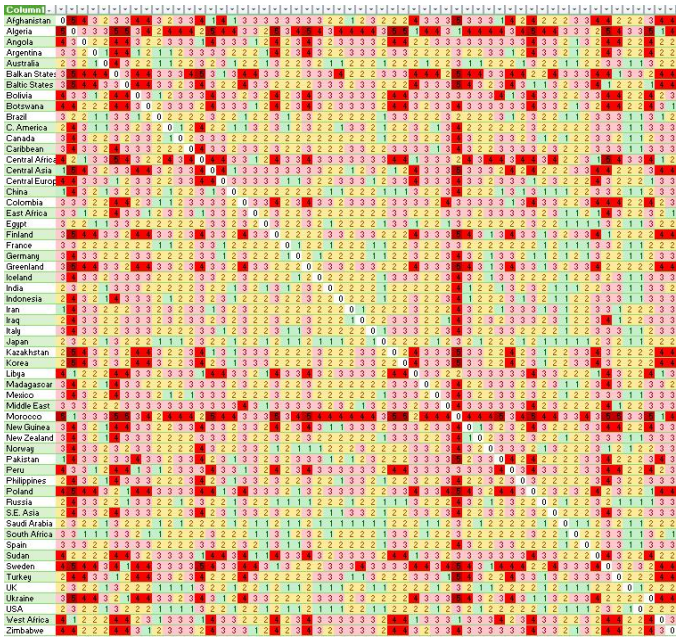

Fig 4.Adjacency matrix of the distance between two countries, Source: Author.

With:

$d(u, v)$ = 1 colored green

$d(u, v)$ = 2 colored yellow

$d(u, v)$ = 3 colored light red

$d(u, v)$ = 4 colored red

$d(u, v)$ = 5 colored dark red

C. *Parameters evaluation*

After we get all the distance between all the countries, its time to determine which country is truly the best. Using two parameters earlier:

Rank	Country	C_C	ϵ
1	Saudi Arabia	0,6	3
2	UK	0,5938	3
3	Japan	0,5876	3
4	USA	0,5816	3
5	France	0,5182	3
6	South Africa	0,5044	3
7	Egypt	0,4914	3
8	Brazil	0,4872	4
9	Australia	0,5135	4
10	Russia	0,5044	4
…	…	…	…
49	Afghanistan	0,3476	5
50	Central Asia	0,3476	5

51	Finland	0,3455	5
52	Greenland	0,3434	5
53	Poland	0,3333	5
54	Central Africa	0,3257	5
55	Balkan States	0,3081	5
56	Sweden	0,3081	5
57	Algeria	0,2767	5
58	Morocco	0,2664	5

Fig 5. Top 10 and Bottom 10 country.

IV. Results and Discussions

A. Result

From Fig 5. we get that the best starting country is Saudi Arabia with $C_C = 0{,}6$ and $\epsilon = 3$, while the worst starting country is Morocco with $C_C = 0{,}2664$ and $\epsilon = 5$. Also, good to point out that all top 10 countries have an air transmission.

Some cases like Australia and Brazil where Australia got a higher C_C than Brazil, but Brazil have less ϵ than Australia makes Brazil better than Australia, because even when Australia have a better C_C it still needs more iteration of transmission to infect all countries.

Eccentricity only ranges between 3-5, which means that no country is truly an outlier. Also, shocking to know that only one island type countries rank on bottom 10 which is Greenland. Sweden is the only country that have more than one type of transmission ranking on bottom 10.

B. Discussions

It makes sense that countries with all three types of transmission will make it to the top, especially those with higher connection like Saudi Arabia. Inversely, country like Morocco where it only have 2 connections is more struggling to infect all other countries.

Airport and seaport have a different connectivity for each country. Case on Sweden where it does have a seaport, but only connect to one country makes it become bottom 10 on ranking.

V. Conclusion

Pick a starting country with good connectivity in airport and seaport such as Saudi Arabia, UK, USA and Japan. See which country can be infected by which country to ensure that all countries got infected. Avoid land transmission-only country to start infecting a good number of countries at the start of the game.

VI. Suggestion

For the next research, consider country demography such as population, climate (hot/cold), rural/urban, arid/humid and handle double connection between two countries. Integrate DNA upgrades, also consider the chance of countries entering quarantine and shutting down their airport, seaport and land borders.

VII. Thank You Note

First of all, I would like to express my gratitude to God Almighty because with His grace and presence, this paper can be completed well without any significant obstacles. Thank you to Dr. Rinaldi Munir as the lecturer in charge of the IF1220 Discrete Mathematics course, who has provided guidance, inspiration, and direction throughout the lecture.

References

[1] R. Munir, "Graf (Bagian 1)," IF1220 Matematika Diskrit lecture notes, Program Studi Teknik Informatika, STEI-ITB, 2024. [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/20242025/20-Graf-Bagian1-2024.pdf.

[2] R. Munir, "Graf (Bagian 2)," IF1220 Matematika Diskrit lecture notes, Program Studi Teknik Informatika, STEI-ITB, 2024. [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/20242025/21-Graf-Bagian2-2024.pdf.

[3] Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495).

[4] Sihotang, J. and Simamora, S., 2020. Analysis of shortest path determination by utilizing breadth first search algorithm.. *Science Info Journal: Informatics and Science*, *10*(2), pp.1-5.

[5] Cohen, E., Delling, D., Pajor, T. and Werneck, R.F., 2014, October. Computing classic closeness centrality, at scale. In *Proceedings of the second ACM conference on Online social networks* (pp. 37-50).

STATEMENT

I hereby declare that the paper I wrote is my own writing, not an adaptation or translation of someone else paper, and is not plagiarized.

Bandung, 19 Juni 2026

Satya Radhityan Yahya - 13525098