Graph Analysis of the Game Atlas

Manas Mittal

1 Introduction

The objective of this analysis is to leverage Graph Theory concepts to explore the structure of the game Atlas. The game involves players taking turns to name places, with the condition that each subsequent name must start with the last letter of the previous one. Using datasets of officially recognized countries and the most populated cities, we constructed three graphs:

- **Country-Only Graph**: Nodes represent countries.
- **City-Only Graph**: Nodes represent the top 500 most populated cities. Assumption : I changed the problem statement from 500 most densely populated to 500 most populated because there was no good dataset available for the prior. The assumption as made as this does not effect the further analysis.
- **Mixed Graph**: Nodes include both countries and cities.

We analyzed these graphs using distinct Graph Theory properties and just human observations to derive strategic insights and assess the difficulty of formulating an optimal game strategy as the domain space increases.

2 Dataset Creation

- **Country Data**: Extracted from the US Government's recognized list (197 countries). Source : Link
- **City Data**: Extracted from World Population Review's top 500 most populated cities. Source : Link
- **Storage Format**: Three graphs were stored in `.pkl` files for inference:
  - `country_graph.pkl` - Graph with country-level nodes.
  - `cities_graph.pkl` - Graph with city-level nodes.
  - `mixed_graph.pkl` - Graph combining both country and city nodes.

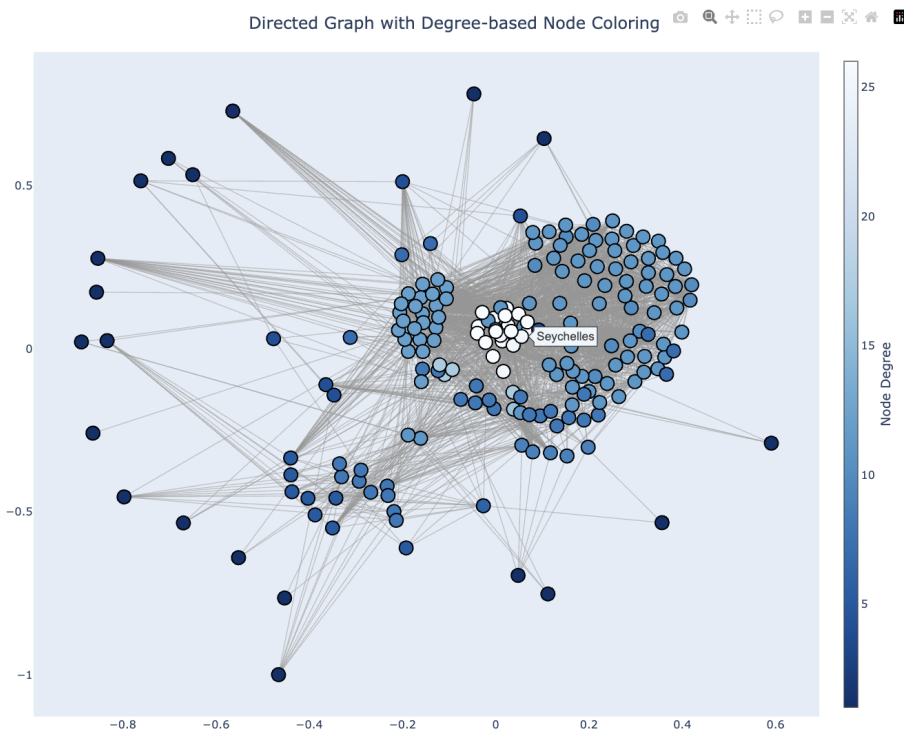

Figure 1: Country Graph with node degrees

3 Graph Visualizations

Other graphs can be found in:

- `country_graph.gexf` - Country graph representation in GEXF format.
- `country_graph.html` - Interactive visualization of the country graph.
- `cities_graph.html` - Interactive visualization of the cities graph.
- `cities_graph.gexf` - Cities graph representation in GEXF format.

4 Graph Properties and Findings on Country graph

The exploratory code for this section is available in `Task1_Inference.ipynb`.

4.1 Degree Analysis

Indegree = 0 Nodes: Letters from which no country starts: {b, v, f, j, z, p} (Total: 38 countries).

- *Insight*: These countries never appear in the middle of the game, making them irrelevant in forming strategies.

Outdegree = 0 Nodes: No such country exists.

- *Insight*: No "trap" countries that end the game immediately at the start.

Outdegree = 1 Nodes: Countries ending in 'o', 'y', or 'q'.

- *Insight*: These countries create potential traps since only one country starts with each of these letters.
- *Optimal Strategy*: If a country ending in 'o', 'y', or 'q' has already been mentioned, strategically respond with another country that also ends with the already mentioned starting character. This traps the opponent in a loss.

Indegree = 1 Nodes: Countries starting with 'c', 'g', 'h', 'k', 'q' have only one incoming edge.

- **Example**: Iraq → Qatar, Luxembourg → Germany/Ghana, Bangladesh → Haiti/Hungary, etc.

4.2 Strongly Connected Components (SCCs)

- The graph is **not fully connected**, meaning certain letters serve as dead ends.
- Countries starting with 'H' can only be reached via 'B', meaning 'H' countries are safe to ignore strategically. This is the reason why countries starting with 'H' each had a different componenent.

4.3 Traps and Winning Strategies

The "AN-AA" Start Strategy:

- Countries starting with 'A' and ending in 'A': Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Australia, Austria.
- If a player starts with an "AA" country, they can force a win.
- **Escape Move**: Afghanistan, Azerbaijan start with 'N', which can break the cycle.
- **Recovery Move**: Using "NA" countries (North Macedonia, Nigeria, Nicaragua, Namibia, North Korea), players can loop back to the "AA" sequence.
- *Winning Insight*: Alternating between "AA" and "AN" transition can give a competitive edge.

The Yemen → Norway Trap:

- **Key Insight**: Yemen is the only country that starts with 'Y'.
- **Trap Strategy**: If an opponent says "Yemen," responding with "Norway" forces them into a losing position, as there are no other countries left that start with 'Y'.
- *Game Impact*: Knowing this trick ensures that if "Yemen" is ever mentioned, you can set up a near-certain victory.

4.4 Centrality Measures

- **Key Observation**: 75 countries end in 'a', and 24 end in 'n'.
- *Insight*: This means there is a very high probability of the opponent saying a country ending with 'a' after which we can use the "AN-AA" Start Strategy and force a win.

5 Degree Analysis of the Cities Graph

exploratory code for this section is available in `Task1_Inference_2.ipynb`.

5.1 Indegree = 0 Nodes

- Cities starting with the letters **Q, C, B, F** have no incoming edges.
- **Total Count**: 93 such cities exist.
- *Analysis*: These cities can never be reached during the game, meaning they hold no strategic significance. Players can safely ignore them while planning their moves.

5.2 Outdegree Analysis

- **No Nodes with Outdegree 0, 1, or 2:**
  - Unlike the country graph, no cities exist that have an outdegree of 0, 1, or 2.
  - *Analysis*: This means there are no easily identifiable "trap states" where a player is guaranteed to lose due to a dead-end.
- **Outdegree = 3 Nodes:**
  - Cities with exactly three outgoing edges start with the letters **E** and **U**.
  - **Cities:**
    - **E**: Ekurhuleni, Esfahan, Edmonton
    - **U**: Urumqi, Ulaanbaatar, Uyo
  - *Strategic Insight*: If a city ending in 'E' or 'U' has already been mentioned three times, strategically respond with another city that ends with the same letter. This forces the opponent into a difficult position, significantly limiting their possible moves and increasing the chances of winning.

6 Mixed Graph Analysis

No significant insights were found from the mixed graph analysis. The combination of country and city data did not reveal additional strategic advantages beyond those observed in the individual graphs.

7 Increasing Domain Space: Easier or Harder?

- **Country-Only Graph**: Allows more structured strategies since the dataset is smaller and more predictable.
- **City-Only Graph**: More complex as there are more possible transitions, increasing unpredictability.
- **Mixed Graph**: The hardest, as the number of possible nodes and paths is maximized, making it difficult to form a consistent strategy.

8 Conclusion

Through the use of Graph Theory, we identified key strategies, such as leveraging countries with high outdegree, avoiding dead ends, and manipulating SCCs to create winning loops. The complexity of forming an optimal strategy increases as we expand from countries to cities and further to a mixed dataset. However, by strategically selecting paths based on degree distribution and connectivity, players can gain a significant competitive advantage.