Global Footprint Network - Analysis

Aditya Sharma
A20516668

Raj J. Shah
A20524266

1 Introduction

The ever-growing impact of human activity on our planet requires a deeper understanding of how different behaviors influence environmental footprints across nations and how can we evaluate such measures. This project delves into this crucial issue, aiming to understand how different activities across countries influence their "ecological footprints."

We started by gathering a rich dataset encompassing various metrics for a multitude of countries. This data included Biocapacity per person, ecological footprint, population size, and the Human Development Index (HDI). We weren't just interested in analyzing these metrics individually; we wanted to see how they played off each other and compared to global trends. By uncovering potential correlations between these factors and carbon emission footprints, we hoped to gain a deeper understanding of how human activity shapes environmental impact.

After getting to know the data inside and out, we embarked on the process – building a network. Imagine a map where countries (nodes) are cities, and metrics like ecological footprint, HDI, and Biocapacity are the roads (edges) connecting them. We created a "weighted" network, where the thickness of these roads reflects how a country's performance on a specific metric compares to global or US benchmarks. Thicker roads meant a country was exceeding those benchmarks, while thinner ones indicated they were falling short. We even developed additional network models, like an "ego graph," to understand a single country's connections to others in more detail. Our primary goal is to draw valid and meaningful conclusions from our analysis. We're particularly interested in exploring several areas like how metrics like ecological footprint and Biocapacity deviate from the global average across different countries over time. Similar analyses for continents will offer a broader perspective. How does a nation's economic progress, measured by GDP relative to the US, affect its footprint and Biocapacity?. By visualizing resource consumption patterns through a network of ecological footprints and Biocapacity, we can identify countries with resource stress or abundance, informing sustainable development strategies. We aim to potentially group countries with similar metrics and even identify "influencers" within these communities across different periods.

Throughout this research, we'll prioritize the ethical considerations inherent to topics like human development and environmental sustainability. Additionally, we anticipate potential challenges with incomplete data for certain countries. We'll address this limitation thoughtfully to ensure the integrity of our analysis. This project represents a significant step towards understanding the complex relationship between human activity and ecological footprints. Through rigorous analysis and exploration, we hope to shed light on the factors contributing to these footprints and provide valuable insights for a more sustainable future.

2 Dataset

The Global Footprint Network (GFN) dataset, a publicly available dataset stands out as an important and powerful utility for building network or graph models to explore sustainability and understanding measures regarding our natural resources. This comprehensive dataset offers ecological footprint and biocapacity data for many countries. Ecological footprint refers to the amount of nature a country uses to produce the resources it consumes and absorb its waste. Biocapacity, on the other hand, represents the Earth's ability to regenerate those resources.

The dataset utilizes standardized metrics, ensuring all countries are represented equally regardless of their economic structure or development level. We can understand networks to understand resource dependencies by analyzing the connections between countries with high ecological footprints and those with low biocapacity.

The GFN dataset goes beyond simply providing a static snapshot. The data spans several decades, all the way back to 1961. This allows researchers to create dynamic network models. The models can serve as timelapse, revealing how resource use patterns have shifted over time. By analyzing these shifts, we can anticipate future challenges and develop strategies for promoting sustainable practices. To create the National Footprint and Biocapacity Accounts, a wealth of data is collected from a wide range of sources. Since 2019, the Global Footprint Network, York University, and the Footprint Data Foundation have collaborated to leverage information from the Food and Agriculture Organization, the UN Comtrade database, the International Energy Agency, and over 20 additional sources.

Data Acquisition

Our initial search for ecological and environmental data led us to the GFN. Recognizing the valuable insights the dataset offered, we explored their resources further. We stumbled upon a sampled dataset of the GFN dataset from data.world, a platform for sharing research data. (https://data.world/makeovermonday/2018w17-ecological-footprint-per-capita/) This sample proved instrumental in familiarizing ourselves with the data structure and format. After exploring the sample and recognizing its potential for our project, we proceeded to utilize the GFN's official API to acquire the full dataset relevant to our needs. The API provided a more streamlined and efficient way to access the comprehensive data required for our project's analysis.

Importantly, for this project, we relied solely on the data acquired directly from the GFN API. There were no modifications or alterations as mentioned in the previous report regarding 'Data Deliverable'.

3 Project Process

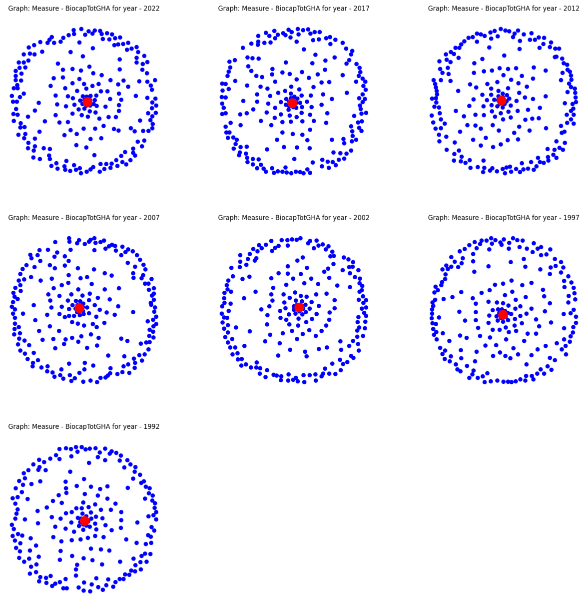
Our project process encompassed several key stages, each vital to the successful completion of our analysis. We began with data acquisition, obtaining the necessary datasets from the Global Footprint Network (GFN) through their official API. This provided us with comprehensive ecological footprint and biocapacity data for numerous countries, spanning several decades.

Once we acquired the data, we proceeded with preprocessing. This involved cleaning the data, handling missing values, and ensuring consistency in formatting to prepare it for analysis. It required consolidation from different APIs for various countries. Additionally, we conducted exploratory data analysis (EDA) to gain insights into the structure and characteristics of the dataset, identifying any potential patterns or anomalies. It mostly included generating graphs depicting different metrics with respect to countries and examining network measures for each.

With the preprocessed data in hand, we moved on to constructing network models. We utilized tools and techniques from network science to represent the relationships between countries based on ecological footprint, biocapacity, and other relevant metrics. This involved creating various types of networks, including relative graphs and ego graphs, to explore different aspects of the data.

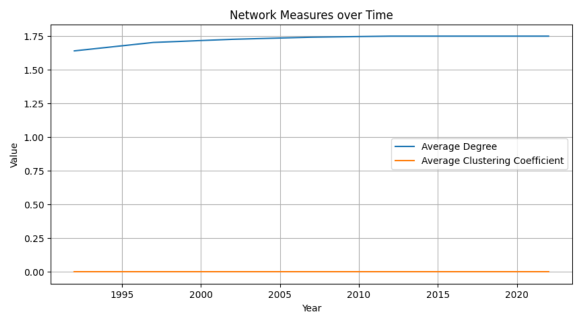
Next, we applied analysis techniques to ex-

Figure 1: Graphs of BiocapTotGHA over the years



tract meaningful insights from the constructed networks. This included measures such as average degree, clustering coefficient, and community detection to characterize the network structure and identify patterns of resource use among countries. We also employed network modeling approaches, such as small-world and preferential attachment models, to simulate the behavior of the networks and evaluate their properties.

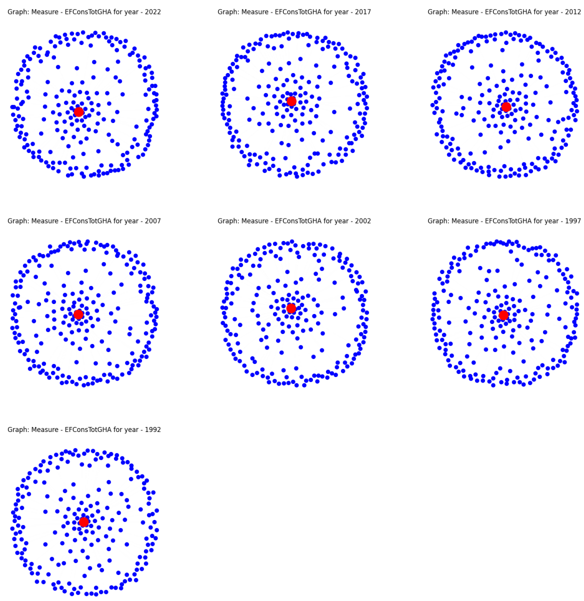
Figure 2: Graphs of 'BiocapTotGHA' Metrics over the years



The above graph illustrates the Average Clustering Coefficient and Average Degree over time. While the Average Clustering Coefficient remains consistent, there is a substantial increase in the Average Degree. This suggests that, despite the addition of new nodes (countries) over time, they are connected to a similar number of nodes in the network, indicating a well-connected network structure retained over time. The closeness centrality also remains relatively constant, starting at 0.41, increasing to 0.43, and then returning to 0.41 over time. This consistency suggests that, despite changes in the network structure and the addition of new nodes, the overall closeness centrality of the network remains stable. This stability indicates efficient communication pathways between nodes, facilitating the exchange of information and resources. The combination of a consistent clustering coefficient, increasing average degree, and stable closeness centrality reflects the robustness and connectivity of the network over time.

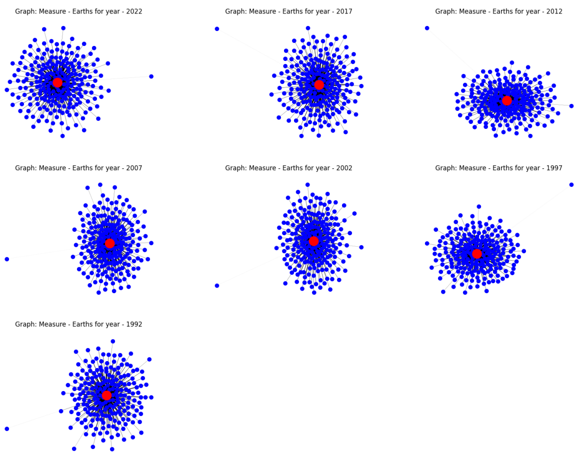
Similarly, Other Measure that we discovered are as follows:

Figure 3: Graphs of 'EFConsTotGHA' over the years



The above graphs exhibited trends similar to the one discussed above, leading us to believe that the network behaves similarly to real-world networks in theory. Employing network models would further solidify this proof as we

Figure 4: Graphs of 'Earths' over the years



proceed.

Throughout the project process, we encountered several challenges that required careful consideration and problem-solving. These included selecting appropriate parameters for network modeling, and interpreting the results of our analysis accurately.

4 Results

The primary outcome of our analysis is the construction and visualization of network models representing the relationships between countries based on ecological footprint, biocapacity, and other relevant metrics. We created a relative graph for our dataset. In the graph as mentioned earlier, nodes were countries and edges were the difference between the 'BiocapTotGHA' of the countries. This helped us to identify the usage pattern of the countries. Most of the countries were using more resources than the average resources used by the world (as a country). There were also some outliers that had no edges, indicating that these countries used resources exactly equal to the world average. Another possibility is that, due to computational complexity and rendering issues, we utilized a small subset of nodes to create the graphs. Consequently, it's possible that some data was lost during this process, resulting in the graph not being connected. The measures for these graph were as follows:

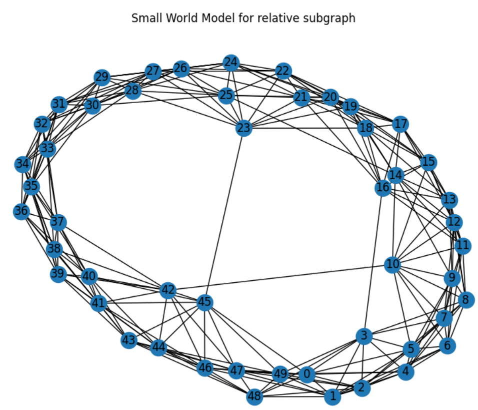
- Number of nodes: 50
- Number of edges: 946
- Average degree: 37.84
- Density: 0.77
- Average clustering coefficient: 0.88
- Is the graph connected? False

These results suggest that the graph represents a complex network of interactions between countries based on their Biocapacity ('BiocapTotGHA') footprint. The high average degree and clustering coefficient indicate a densely connected network with strong clustering tendencies. However, the presence of isolated components highlights the existence of separate subgroups within the network, potentially reflecting different regional or geopolitical dynamics.

Next, we tried to simulate our model using the small world modelling and preferiantial attachment modelling. We decided to go against the random modelling as it is random in nature and would not be of any significance here.

Below is the graph we obtained when used the small world modelling.

Figure 5: Small world Model for Relative Subgraph



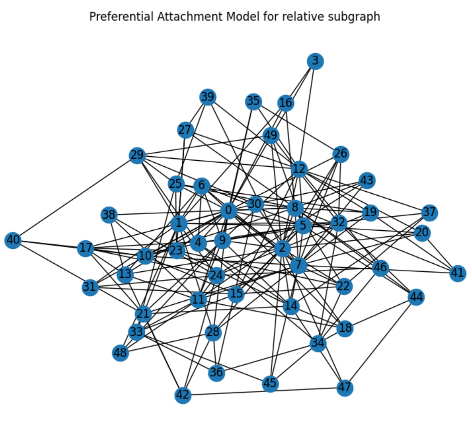
The graphs suggest the presence of significant nodes connected to many others, indicating the formation of several different communities. However, when weight was used as a measure to identify communities, only two communities emerged.

The results demonstrate a high clustering coefficient of 0.63 and a short average path length of 2.4375. These findings suggest that the small-world model used to generate the graph exhibits characteristic properties of small-world networks, including relatively high clustering, short average path lengths, and connectedness. We opted to maintain the average degree (k) as 10. For different values of k, similar results can be observed, consistent with those exhibited by small-world networks. These properties render small-world networks efficient for information transfer and are commonly observed in various real-world networks, such as social networks and certain types of biological networks.

The emerged communities are as follows: **Community 1:** [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 46, 47, 48, 49] **Community 2:** [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]

These findings indicate the partitioning of nodes into two distinct communities, with each community exhibiting cohesive internal connections.

Figure 6: Preferential Attachment Model for Relative Subgraph



We experimented with another network model known as the preferential attachment model, which resulted in a graph noticeably denser compared to the previous model. Upon analyzing the parameters of the network model, we observed that the average degree of the graph was 7.36, indicating that, on average, each node is connected to approximately 7 other nodes. The density of the graph, calculated as 0.15, represents the ratio of actual connections to the total possible connections in the graph. This suggest against the visual analysis that the preferiantial attachment model yielded a denser graph compared to small world model. The density of Small world model was 0.2.

Density serves as a crucial metric indicating the ease of information dissemination within a network. We found that the small-world model exhibited higher density and average degree compared to the preferential attachment model. However, it suffered from a higher average path length and lower closeness centrality (0.42).

These results present a challenging decision when selecting models for real-world scenarios and analysis deployment. Both models offer distinct advantages and drawbacks, making it difficult to determine the most suitable option. While the small-world model facilitates efficient information spread due to its high density and average degree, its higher average path length and lower closeness centrality pose limitations.

In contrast, the preferential attachment model may provide more efficient communication pathways with its shorter average path length and higher closeness centrality. However, its lower density and average degree may impact the overall spread of information.

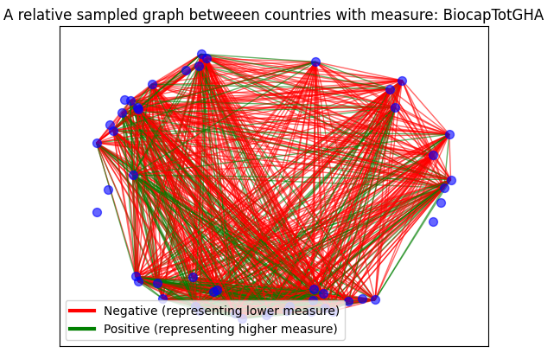
Ultimately, the selection of the appropriate model depends on the specific characteristics of the network and the objectives of the analysis and deployment.

The average clustering coefficient was determined to be 0.227, suggesting a relatively low level of clustering in the network. This coefficient quantifies the tendency of nodes to form clusters or communities. Furthermore, the closeness centrality values provided insights into the centrality of nodes within the network. Notably, the node with the highest closeness centrality was found to be 0.63.

Additionally, the average path length in the graph was calculated as 2.091, indicating the average number of steps required to navigate from one node to another. These results collectively suggest that the preferential attachment model used to generate the graph exhibits characteristic properties of such networks.

These properties include a relatively low average clustering coefficient, indicative of a network with some level of hierarchy, and a short average path length, suggesting efficient communication between nodes. Moreover, the graph was found to be connected, implying that there is a path between every pair of nodes, which aligns with the typical characteristics observed in preferential attachment networks.

Figure 7: Relative Graph



Relative graph communities:

- **Community 1:** [3, 7, 8, 18, 32, 45, 52, 57, 59, 66, 68, 79, 80, 89, 99, 102, 103, 109, 121, 126, 129, 135, 138, 145, 155, 157, 167, 181, 193, 203, 212, 223, 234, 1007]
- **Community 2:** [351, 1004, 1005, 1006, 1016, 1018, 1021, 2004]
- **Other Communities:** [258], [36], [187], [65], [82], [1008], [239], [1013]

The relative graph revealed two main communities, indicating large groups either using above or below average resources compared to the world. Other communities consisted of single nodes either not connected to the graph or not fitting into the main communities.

We know that the real-world network exhibits high clustering and short average path lengths, resembling characteristics of small-world networks in our case. Additionally, the preferential attachment model yielded a denser graph with shorter average path lengths compared to the small-world model, but showed deviations from real-world network properties. While both models captured certain aspects of real-world networks, neither fully replicated their complexity. Real-world networks typically exhibit a combination of characteristics from various models, highlighting the need for nuanced modeling approaches. Hence, we could conclude that our network exhibited characteristics resembling a small-world model, which aligns with expectations considering it was a sample rather than the entire dataset.

The analysis also yielded insights into various aspects of countries' resource use and environmental impact, including the detection of communities with similar resource consumption patterns and temporal trends analysis to anticipate future challenges.

5 Discussion

Examining ecological footprints through network analysis proved to be a rewarding experience, highlighting that the approach to uncover hidden patterns and relationships between countries' resource use and biocapacity. However, the project also revealed some areas for improvement. Data limitations, particularly missing data points for certain countries, with the need for the importance of exploring alternative data sources and potentially employing data imputation techniques in future projects. Additionally, while the project utilized existing metrics effectively, crafting new features from the data (e.g., resource consumption per capita) could have yielded even deeper insights.

Looking ahead, several exciting avenues for future exploration have emerged. Constructing multi-layered networks that incorporate factors like trade flows, political alliances, and energy consumption could reveal hidden dependencies and influences on resource use. Furthermore, incorporating time series forecasting techniques could allow us to predict future ecological footprint trajectories under various scenarios, informing policy decisions for a more sustainable future. Delving deeper into specific sectors within countries (e.g., agriculture, industry) could provide a more granular understanding of how resource consumption patterns differ across industries and their contribution to a nation's overall footprint. Finally, integrating data on deforestation rates, renewable energy production, and other relevant factors could enrich the dataset and potentially reveal new correlations between human activities and environmental impact.

Ultimately, this project has been a valuable learning experience, not just in terms of the insights gained on ecological footprints, but also in identifying areas for further exploration and refinement. By incorporating these lessons learned and pursuing the potential of these future directions, we can refine our knowledge and develop more effective strategies for promoting sustainability.

6 Future Work

This project helps us build the groundwork for a deeper understanding of ecological footprints and understanding its relevance and how using different measures, we can analyze the resource capability and its trends. Future endeavors can delve into specific sectors within countries (e.g., agriculture, industry) for a more granular analysis of resource consumption patterns and their contribution to footprint generation. Time series forecasting techniques can be employed to not only identify past trends but also predict future ecological footprint trajectories under various scenarios (e.g., population growth, and technological advancements). This predictive power can help in informing policy decisions and resource management strategies aimed at mitigating environmental impact.

Beyond the core network, constructing multi-layered networks encompassing measures like trade flows, and other relevant factors alongside the ecological footprint network could reveal hidden patterns and dependencies that influence resource consumption. Additionally, pinpointing "influencer" countries within these networks could shed light on how nations learn and adapt from each other's environmental practices. This knowledge sharing could foster collaborative efforts towards sustainability.

By crafting new features from existing data (e.g., resource consumption per capita), we can gain a deeper understanding. Additionally, integrating external data like deforestation rates and renewable energy production can create a more comprehensive picture. This enriched dataset will fuel even more robust analyses and potentially reveal new relationships between human activities and their environmental impact. By pursuing these avenues of future work, we can refine our understanding of the factors shaping ecological footprints and help in formulating strategies for promoting sustainability.

References

[1] https://www.footprintnetwork.org/

[2] https://data.footprintnetwork.org/#/

[3] https://data.footprintnetwork.org/#/api

[4] https://data.footprintnetwork.org/#/abouttheData

[5] https://en.wikipedia.org/wiki/Global_Footprint_Network

[6] https://networkx.org/documentation/stable/

[7] https://networkx.org/documentation/stable/reference/generated/networkx.generators.ego.ego_graph.html

[8] https://networkx.org/documentation/stable/auto_examples/drawing/plot_ego_graph.html
[9] https://networkx.org/documentation/stable/auto_examples/algorithms/index.html
[10] https://networkx.org/documentation/stable/reference/classes/generated/networkx.Graph.subgraph.html
[11] https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.filter.html
[12] https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.query.html
[13] https://stackoverflow.com/questions/39500265/how-to-manually-create-a-legend