# CNT 5805 Final Project Report
# Green Economy of Texas & Florida: Trend, Growth and Opportunities


Submitted By
Team 07

Team Members:
Divya Selvaraj
Shalini Sneha
Likhita Pula


11/14/2021

## **Section 1:**

### *Part 1: Project Summary*

A Green Economy promotes the transition to low-carbon, resource-efficient, and socially inclusive economies, which in today's world is of foremost importance. It is very important for all world economies to lower their carbon footprint while moving towards an economically stable and flourishing future. It is based on the concept that if economic expansion is the primary goal, economic growth must be decoupled from resource use and negative environmental consequences. The increase in the population and their growing needs has led to climate change and environmental risk that is encouraging policymakers to take steps towards green development. Through the insights and inferences drawn from this project, we can enable the economies (of Texas and Florida) to develop strategies that can ensure the development of green production that brings a sustainable future. This was one of the key factors that motivated us to analyze this data.

For this project, we have analyzed the data from the USA Trade online website that is provided by the U.S. Census Bureau. The data includes export values of each green product for all the states of the United States (US) that are exported over a period of 5 years (2016 to 2020). The export value of each green product for each year is based on US dollars, and the dataset contains a 6-digit Harmonized System (HS) code for each green product along with its name. Additionally, a list of green products that is constructed using a combined list of environmental goods developed by the Organization for Economic Cooperation and Development (OCED) is used to identify green products from the comprehensive list of all products (commodities).

Below is the link to the data source:

https://usatrade.census.gov/

### *Part 2: Project Purpose*

The purpose of analyzing this network is to analyze the green production space of two key states in the US and propose strategies for these states that can enable them to thrive economically while remaining environmentally sustainable. While Texas has the highest export rank with major oil and gas resources, Florida is rich in its agricultural production. This has inspired us to focus on these two US states for our analysis. By examining the expansion of the green production baskets in Texas and Florida over a span of 5 years (2016 to 2020), we can understand whether their economic development has helped in mitigating environmental risks or not. According to existing research, countries expand their production basket by adding products that require similar capabilities to those that they already manufacture, which is referred to as "path dependency". Through this project, we would like to find whether Texas and Florida's green product space follows path dependency or not, which can enable us to further research and propose new economic development policies.

### *Part 3: Research Questions*

As a first step towards the project, we filtered out data containing export values of all green products from the complete list of products for each US state. Using this data, we calculated the

Revealed Comparative Advantage (RCA), which measures the relative advantage of a certain country in a certain class of products based on its trade flow. $RCA_{Ci}$ for a country c and product i can be calculated as:

$$RCA_{ci} = \frac{\frac{X_{ci}}{\sum_j X_{cj}}}{\frac{\sum_c X_{ci}}{\sum_{cj} X_{cj}}} \geq 1$$

Where $x_{c_j}$ shows the country's export value for product i, $\Sigma_j x_{c_j}$ shows the country's total export value for all products j, $\Sigma_c x_{c_i}$ shows the total export value of product i in the world, and $\Sigma_{cj} x_{c_j}$ shows the total export values of all products for all countries in the world. For our project since wee want to analyze Green Product Space (GPS) network for Texas and Florida, in place of countries we have considered states and our world in the United States (US).

Given that a country exports product i competitively, the proximity between i and j is the smallest number of pairwise conditional probabilities which can be calculated as:

$$\varphi_{i,j} = \min\left\{P\left(x_i/x_j\right), P\left(x_j/x_i\right)\right\}$$

Where $\varphi_{i,j}$ is the proximity value between products i and j, and $P(x_i|x_j)$ can be defined as the probability that i is exported competitively, given that j is also competitively exported. Similarly, $P(x_j|x_i)$ can be defined as the probability that i is exported competitively, given that j is also competitively exported.

Using these RCA and proximity values, the final Green Product Space (GPS) Network was built, which was used to answer the below research questions:

1. How different are the two states in green product space?

   The Green Product Space is a network that maps all green products depending on how comparable their required skills are. The green products in the GPS network are the nodes, and the proximity values that connect the nodes are the links. The GPS network of Texas and Florida enables us to visualize the number of green products with a high revealed comparative advantage (RCA) (RCA > 1) that these states can produce in base year 2020.

2. Do the states show path dependency in green product evaluation?

   To evaluate whether Texas and Florida followed the path dependency (which means a state's green production basket diversifies based on its current capabilities, including technology, capital, institutions, and population skills), a list of new green products that were not part of a state's production basket earlier (say in 2016) and had entered state's production basket at a later period (say in 2018) were identified. To identify the new green products, two years, say 2016 and 2018, are considered, and all green products (with RCA value > 1) from a state that were not part of the green production basket at base year 2016 but had entered the state green production basket at year 2018 were considered. By assessing these new green products, deductions were made regarding whether these states follow path dependency or not.

3. How have these states evolved over the years in their green economy?

   A GPS network for each year (2016 to 2020) was constructed to analyze the evaluation of the green product baskets of Texas and Florida over the years. This has enabled us to assess these states based on their improvement in the expansion of their green production baskets.

4. What can be a future strategy for these states to thrive economically while remaining environmentally sustainable?

   Supplementary data points like pollution levels (from the United States Environmental Protection Agency (US EPA) and Gross Domestic Product (GDP) estimated by the United States Bureau of Economic Analysis (US BEA) were collected to analyze the changes over the years and relate the insights drawn to the GPS network. This has enabled us to suggest strategies that Texas and Florida can undertake to experience economic growth while mitigating environmental risks.

**<u>Section 2:</u>**

*Part 1: Network Details*

Using the RCA and Proximity values for the green products in the year 2020 nodes and edges of the network were extracted. The nodes and edges were imported into Gephi, and it was noted that there are about 244 nodes and 26450 edges. The network built is an undirected network and the weights are assigned to the edges based on the calculated proximity values.

*Part 2: Network Overview & Statistics*

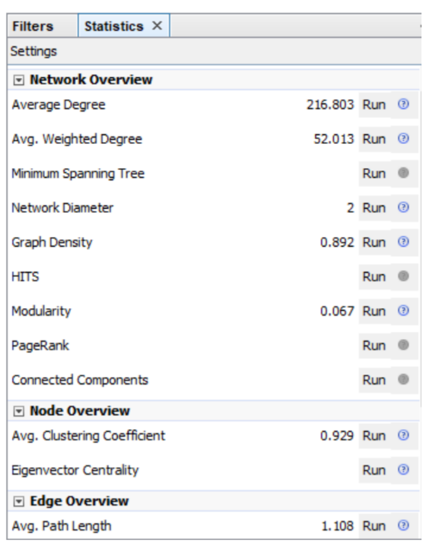


Fig 1: Screenshot of Network Overview statistics of the GPS network in Gephi

Fig. 1 shows an overview of network statistics for the given GPS network. It is observed that graph density is about 0.892, suggesting the network to be very dense. Even so, the average clustering coefficient is about 0.929, suggesting that connections in the network are dense. The statistic of average degree shows that, on an average, every node has approximately 217 edges connected to it, which tells us that the green products are well connected to each other. By analyzing the degree distribution plot from Fig 2, we can say that the distribution shows exponential growth, suggesting it is not scale-free but rather random.

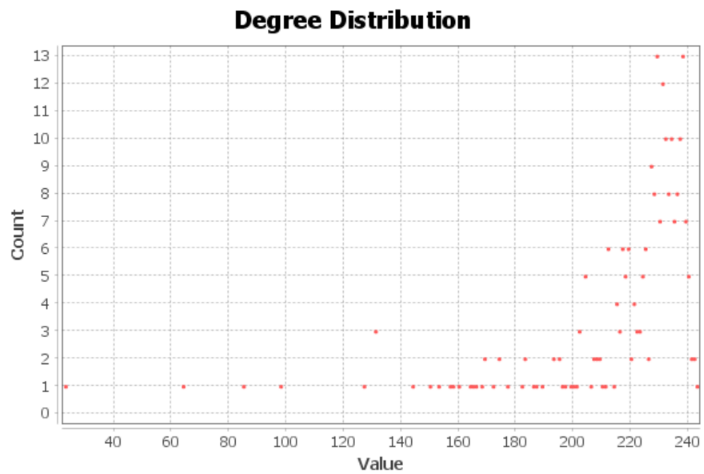


Fig 2: Degree Distribution Plot

*Part 3: Initial Network*

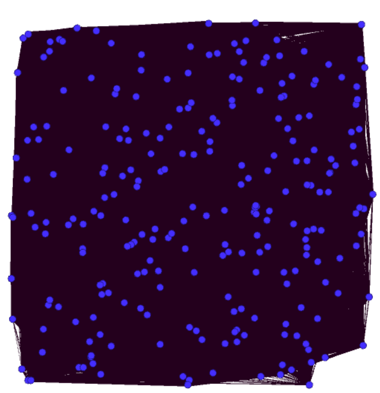

Fig 3: Initial GPS network in Gephi

From Fig 3, we cannot make extensive deductions, but the network looks very dense with a high number of edges and nodes. Since there are about 244 nodes and 26450 edges, we cannot identify if any hubs or clusters have developed among the nodes from the initial graph. The nodes in the network represent green products, and the edges of the network are calculated based on proximity measure that tells that two products can be produced using similar capabilities.

**Section 3:**

*Part 1: Different Network Layouts*

On our initial Network we ran below layouts (All the layouts shown below are built in python using networkx package):

- Fruchterman Reingold Layout
- Kamada Kawai Layout
- Circular Layout
- Maximum Spanning Tree Layout

**Fruchterman Reingold Layout:**

Fruchterman Reingold Layout is a force-directed algorithm that simulates a force-directed representation of the network, with edges acting as springs that keep nodes close together and nodes acting as repelling objects, a force known as anti-gravity. The simulation will continue until the locations have reached a state of near equilibrium. It is also known as the "Spring layout" in Python network graphs. Fig 4 shows the GPS network for the year 2020, where we can see that few nodes are stretched apart from each other while most of the nodes are concentrated at the center, suggesting most of the green products have closer proximity or relatedness to the other green products.

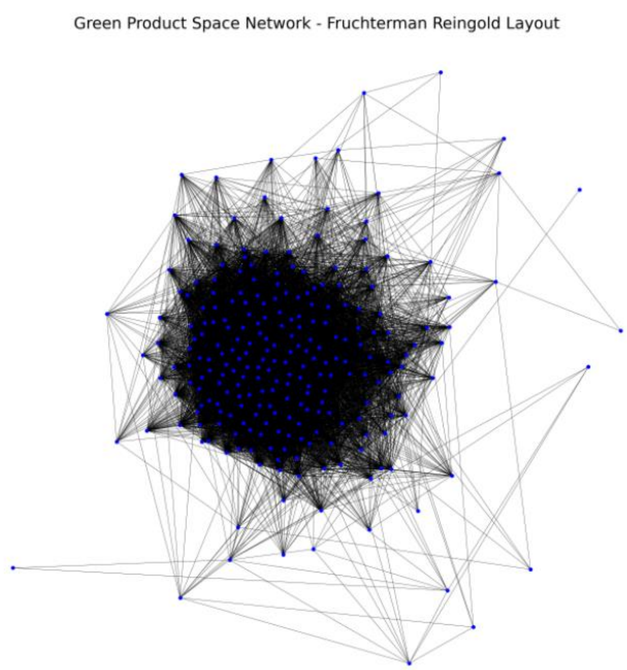


Fig 4: GPS network build using Fruchterman Reingold Layout for the year 2020

## Kamada Kawai Layout:

Kamada Kawai Layout positions nodes using Kamada-Kawai path-length cost-function. An advantage of this method is that it can be applied straightforwardly to drawing edge-weighted graphs, assuming that edge lengths have to reflect their weights. Fig 5 shows GPS network built using kamada kawai layout for the year 2020 and here we can see that compared to Fruchterman Reingold Layout, the nodes (green products) seem to be more spaced out in the middle. However, for our analysis Fruchterman Reingold Layout would be better to examine as it shows that the nodes in the middle are placed closer to each other due to their similar capabilities for production.

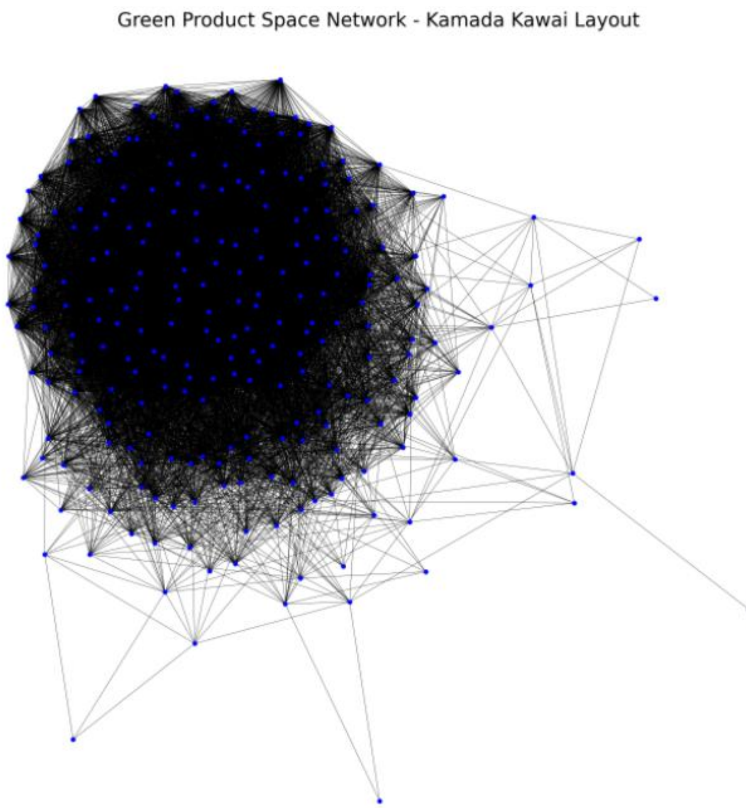


Fig 5: GPS network build using Kamada Kawai Layout for the year 2020

## Circular Layout:

The circular layout algorithm draws nodes in a circle by node, a metric (degree, clustering coefficient, betweenness centrality etc.) or by an attribute. It is used to show a distribution of nodes with their links. In Fig 6, GPS network build using circular layout is shown and we can see that the middle part of the network which illustrates the edges of the network appear to be black suggesting each node has many edges connecting them to the other nodes (green products). However, compared to the earlier layout this layout does not provide deeper insights on where the nodes are located and how close one green product is with another.

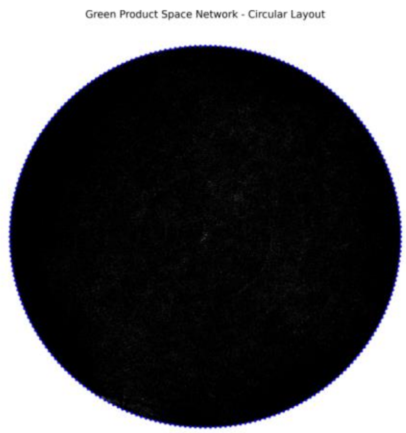


Fig 6: GPS network build using Circular Layout for the year 2020

**Maximum Spanning Tree Layout:**

A maximum spanning tree (MST) is a subgraph of a graph (a tree) that has the largest feasible sum of edge weights. The spanning trees for each connected component of the graph are combined to form a spanning forest.

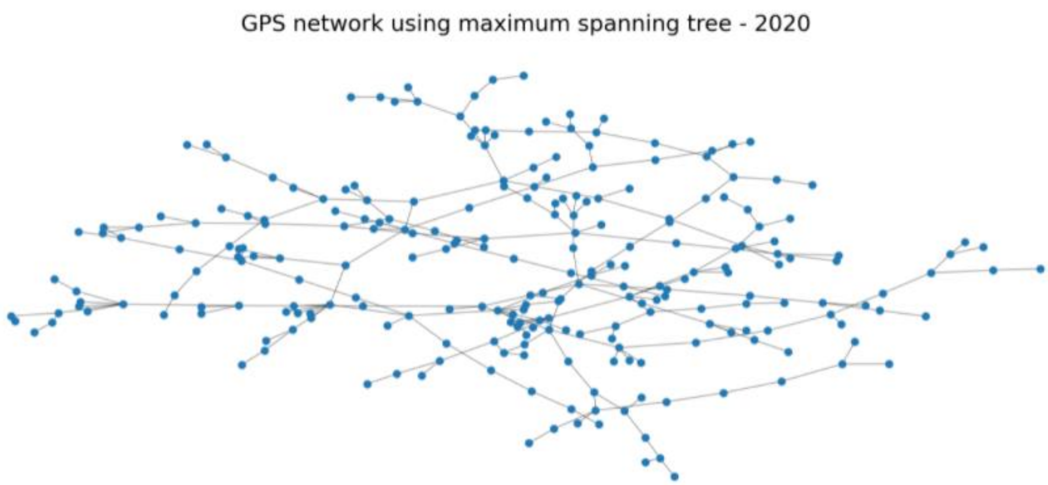


Fig 7: GPS Network build using Maximum Spanning Tree Layout for the year 2020

From the Fig 7, we can see that the complexity of the graph is reduced with the decreased number of edges. Few of the nodes (green products) appear to be closely connected to others suggesting that these products might be having higher relatedness and need similar capabilities for their production. This layout shows few hubs (few nodes that are close to each other) but at the same time most of the nodes are scattered across the network.

*Part 2: Final Network Layouts*

The GPS network built using the Maximum Spanning Tree (MST) layout clearly shows how few nodes (green products) are closely placed in the network, suggesting the formation of green product communities that require similar capabilities for their production. Since there is less overlap between the edges in the MST layout as compared to other layouts, we can even identify some hubs in the network. However, the network built using the Fruchterman Reingold Layout could be enhanced to generate better results as it also shows how few products (nodes) are stretched apart from the other products (nodes). Fruchterman Reingold Layout would provide greater insights by selecting a threshold for the proximity values and coloring the nodes with RCA > 1, so we chose Fruchterman Reingold Layout as our final layout.

**Section 4:**

*Part 1: Network Construction Actions & Analysis*

Since we want to analyze the GPS network of Texas and Florida, we first calculated the RCA and proximity value for each year and as 2020 is the year that we are focusing, the final GPS network was build using the nodes and edges list of the year 2020. Fig 8 shows the GPS network for 2020 that was filtered considering proximity threshold of 0.25 ($\varphi_0$) which has enabled us to improve the visual representation of network. This GPS network with $\varphi_0 > 0.25$ has about 244 nodes (green products) and 12161 edges (calculated based on proximity values).

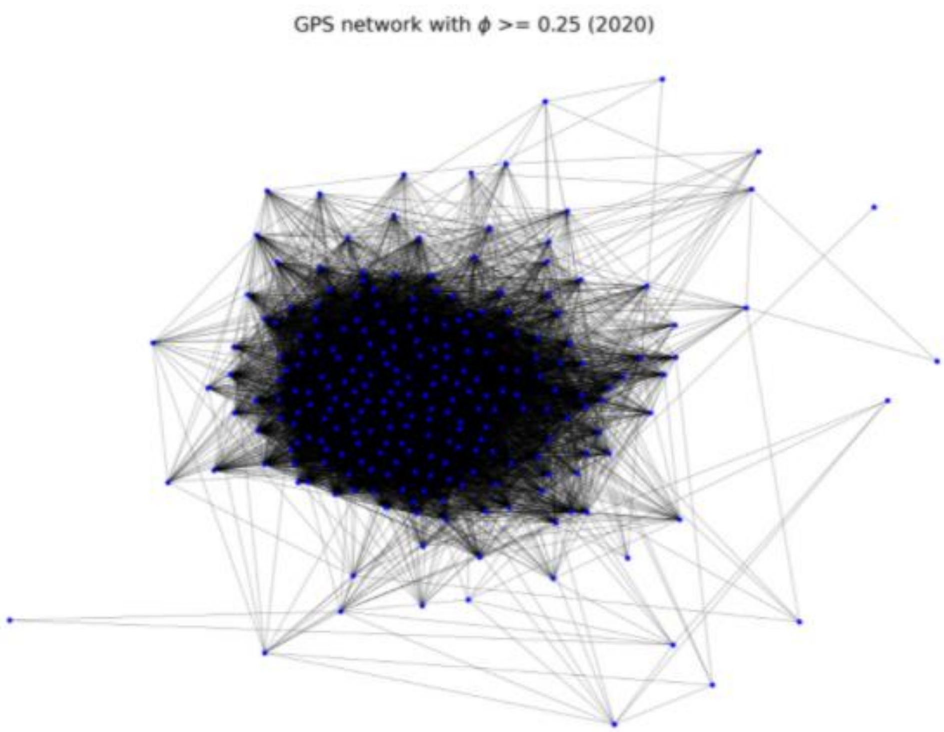


Fig 8: GPS network for the year 2020 with $\varphi_0 > 0.25$

To visualize the network of Texas and Florida, we examined the RCA values for the green products of the two states and highlighted the nodes of the GPS network (of 2020) based on their RCA values. Since RCA value > 1 would suggests that the green product is competitively produced by the state all nodes in GPS network with an RCA value > 1 are colored as green and the nodes with RCA value < 1 are colored as red which has enabled us to visualize the difference in the GPS network of Texas and Florida. From Fig 9 which shows the GPS network of Texas and Florida for the year 2020 we can clearly see that Florida has a large green production basket when compared to Texas but at the same time the green production basket of Texas is more diversified when compared to Florida which appears to be concentrated at a place in the network.

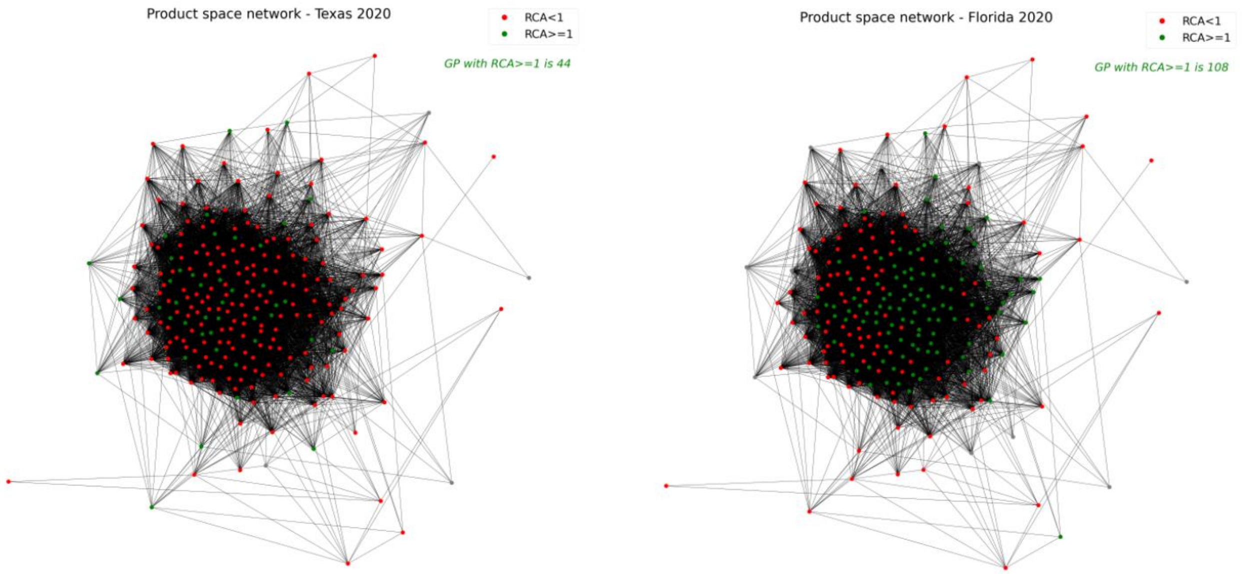


Fig 9: GPS network of Texas and Florida for the year 2020

To further analyze the network, we have colored each node in the base GPS network (for the year 2020) based on 11 green product categories [1,4]. Fig 10, shows the GPS network with green products colored based on below green product categories:

- APC - Air pollution control
- CRE - Cleaner or more resource efficient technologies and products
- EPP - Environmentally preferable products based on end use or disposal characteristics
- HEM - Heat and energy management
- MON - Environmental monitoring, analysis and assessment equipment
- NRP - Natural resources protection
- NVA - Noise and vibration abatement
- REP - Renewable energy plant
- SWM - Management of solid and hazardous waste and recycling systems
- SWR - Clean up or remediation of soil and water
- WAT - Waste water management and potable water treatment

It is evident that the GPS network has comparatively higher number of green products that fall in the category of Renewable Energy Plant (REP). This visualization helps us to detect few communities as few of the nodes that are falling in same category are appearing to be clustered together in the network (for example the red colored nodes).

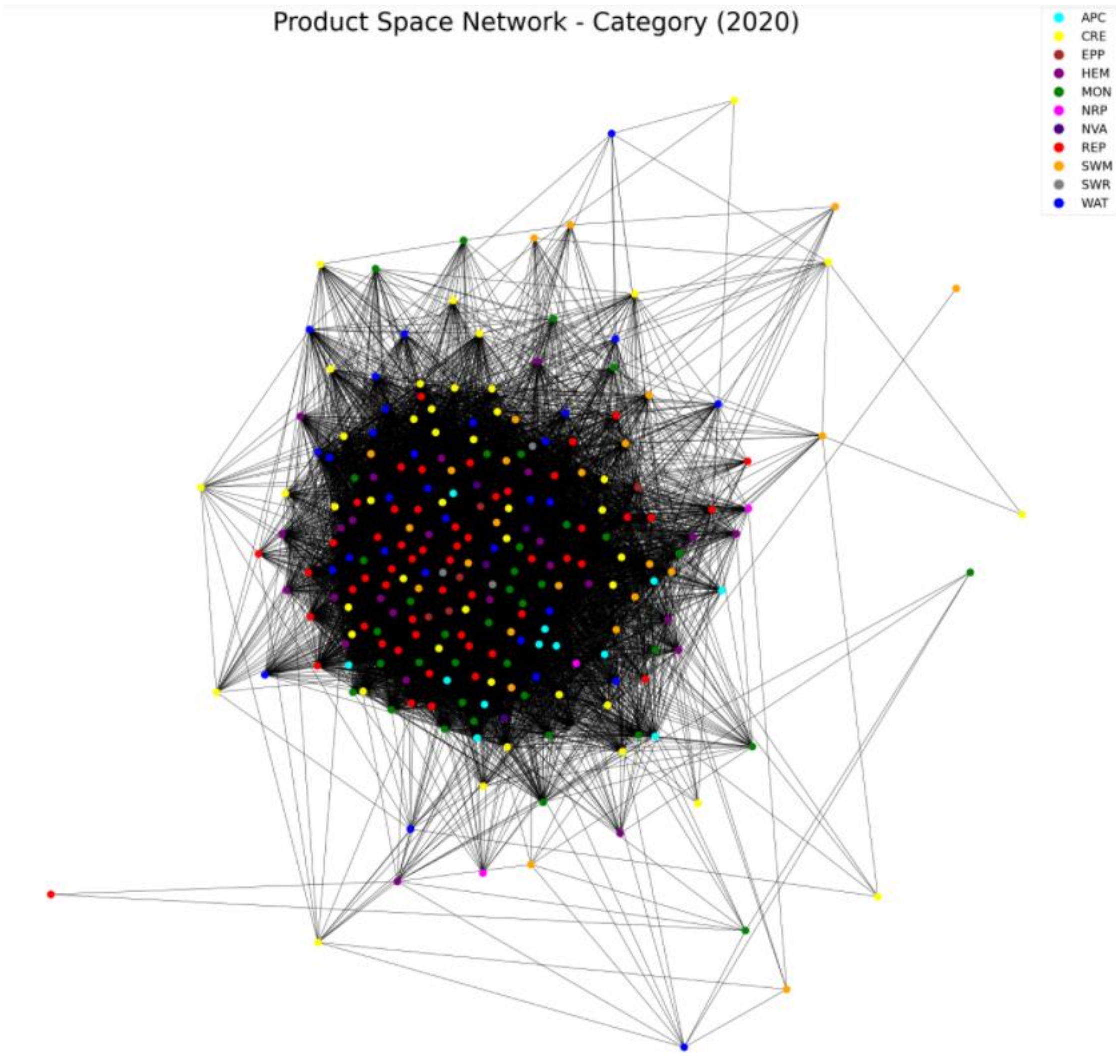


Fig 10: GPS network (for the year 2020) based on 11 green product categories

Although Florida has a larger green production basket it would be interesting to visualize the green products that have high export value for both the states. Thus, in Fig 11 the nodes of the GPS network have been resized based on the export value of the green product. We can see that the green products that fall in the category of REP, WAT and NVA are showing high values of export in Florida.

| HSCode | Green Product | Category | state export |
|---|---|---|---|
| 841199 | Gas Turbine Parts Nesoi | REP | 545803447 |
| 853710 | Controls Etc W Elect Appr F Elect Cont Nov 1000 V | REP | 202728185 |
| 848180 | Taps Cocks Etc F Pipe Vat Inc Thermo Control Nesoi | WAT | 192238731 |
| 840999 | Spark-ignition Reciprocating Int Com Pistn Eng Pts | NVA | 158920138 |
| 903289 | Auto Regulating Ins & Appr Ex Throstat, mnstat, Etc | REP | 130736825 |
| 850440 | Static Converters; Adp Power Supplies | REP | 118062000 |
| 902750 | Instruments Etc Using Optical Radiations Nesoi | MON | 102772793 |

| | | | |
|---|---|---|---|
| 870390 | Passenger Motor Vehicles, Nesoi | CRE | 96635581 |
| 841430 | Compressors Used In Refrigerating Equipment | APC | 94819565 |
| 841480 | Air/gas Pumps, Compressors And Fans Etc, Nesoi | APC | 86367307 |

Table 1: Top 10 green products of Florida with high export value in year 2020

Table 1 clearly shows that Florida has high exports of renewable energy plant products While Texas shows high exports in WAT (Waste water management and potable water treatment) category products.

| HSCode | Green Product | Category | state export |
|---|---|---|---|
| 848180 | Taps Cocks Etc F Pipe Vat Inc Thermo Control Nesoi | WAT | 1207606133 |
| 732690 | Articles Of Iron Or Steel Nesoi | WAT | 1083201676 |
| 847989 | Mach & Mechanical Appl W Individual Function Nesoi | SWM | 988177744 |
| 841182 | Gas Turbines Of A Power Exceeding 5,000 Kw | REP | 898804132 |
| 853710 | Controls Etc W Elect Appr F Elect Cont Nov 1000 V | REP | 885189485 |
| 848190 | Pts F Taps Etc F Pipe Vat Inc Press & Thermo Cntrl | WAT | 716351818 |
| 850440 | Static Converters; Adp Power Supplies | REP | 714547456 |
| 841199 | Gas Turbine Parts Nesoi | REP | 666864559 |
| 854370 | Elec Mach And App, Having Indiv Functions, Nesoi | WAT | 660443542 |
| 842139 | Filter/purify Machine & Apparatus For Gases Nesoi | APC | 587798900 |

Table 2: Top 10 green products of Texas with high export value in year 2020

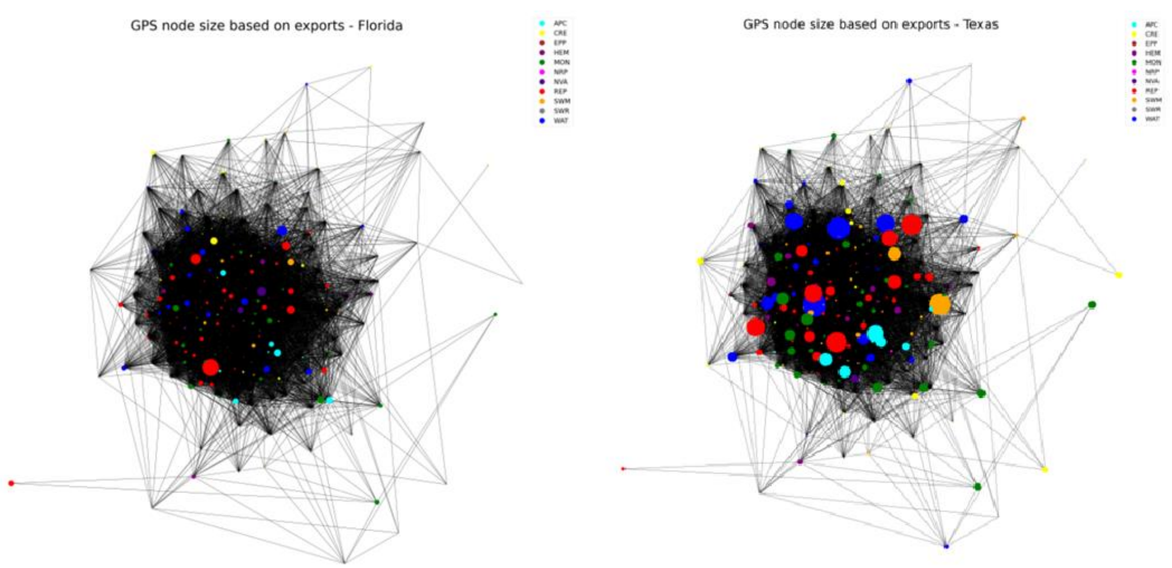


Fig 11: GPS Network (for the year 2020) with nodes resized based on export value of the products in Florida (left) and Texas (right)

## *Part 2: Visualization Summary*

Below are the key deductions made after analyzing the GPS network for Florida and Texas through sizing, coloring, and labelling the nodes:

- Florida has about 108 products having RCA value > 1 while Texas has only 44 products with RCA >1. This indicates that the green production basket of Florida is larger when compared to Texas which is also evident in Fig 9. However, Texas shows diversified green production basket suggesting that it has the capability to expand its green production through structural changes.
- GPS network has comparatively higher number of green products that fall in the category of Renewable Energy Plant (REP)
- Green products that fall in the category of REP, WAT and NVA are showing high values of export in Florida while Texas shows high exports in the WAT category products.

**<u>Section 5:</u>**

*Part 1: Statistics*

Below statistics are computed after filtering edges based on the proximity threshold which is greater than 0.25 ($\varphi_0 > 0.25$)

<u>Average Degree:</u> 99.68

Average Degree is the average number of edges per node in the graph. The number of nodes is represented on the x-axis, while the degree value is represented on the y-axis. The number of edges a node has, or the total of edges each node has, determines its degree. From the degree distribution plot, we can say that GPS network is a random network and does not seem to follow power law distribution.

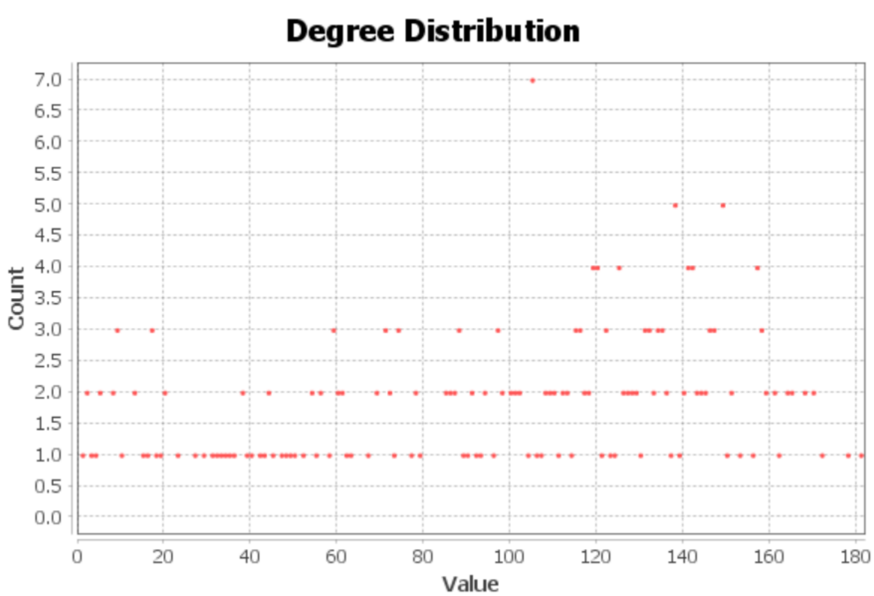


Fig 12: Degree Distribution Plot for GPS Network with $\varphi_0 > 0.25$

The average degree distribution suggests that in the Green Product Space (GPS) network, on an average every green product has a proximity with approximately 99 other green products. Also, the degree distribution of the network which suggests that the network is random, tells us that unlike scale free network the GPS network would not be having closely clustered hubs (communities). This is evident in Fig 10, where we can see the possibility of the green products that belong to one category being clustered far apart.

Average Clustering Coefficient: 0.645

Average clustering coefficient is the average of all individual clustering coefficients of the network. It measures the degree to which nodes in a network tend to cluster together and the clustering coefficient distribution shows the nodes in the network do not have very high or very low clustering coefficient.

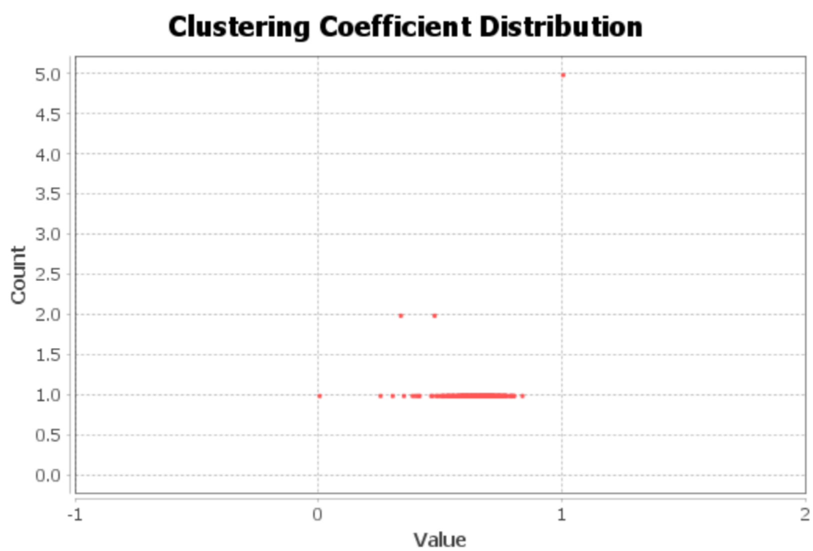


Fig 13: Clustering Coefficient Distribution Plot

The average clustering coefficient is about 0.645 and the clustering coefficient distribution suggests that most of the nodes have a clustering coefficient between 0.3 to 0.7. From the clustering coefficient distribution plot, we can say that there are not too many nodes that are very tightly connected forming close community in the GPS network. We can see that in the GPS network 5 nodes have very high clustering coefficient of 1.

Density:0.41

The density of the network is about 41% which is indicating that the network is not very dense after filter the GPS network with $\varphi_0 > 0.25$. Density is a measure of how tightly interconnected a network is and is calculated by examining the proportion of edges relative to the possible number of connections.

Network Diameter: 4

There should be at least 4 number of connections required to traverse the graph. Diameter of the graph is the longest shortest path between all sets of nodes in the network graph.

<u>Relative Comparative Advantage (RCA) Distribution:</u>

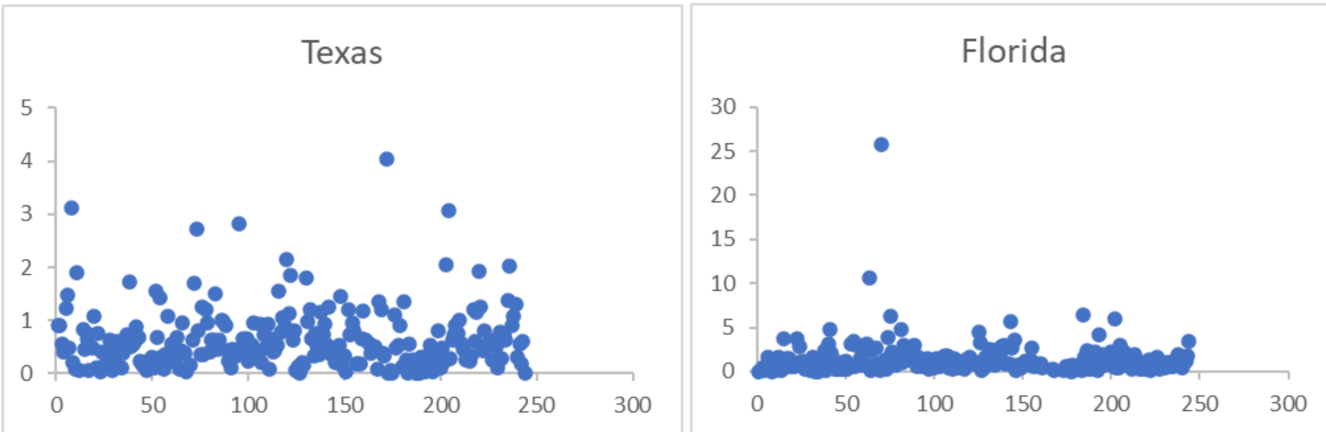


Fig 14: RCA distribution for Texas and Florida for the year 2020

From the distribution of RCA values for Texas and Florida we can see that Florida shows high values of RCA ranging up to 25.67 whereas the distribution of Texas RCA values lies between 0 to 4.03. This suggests that Florida has some products which have very highly competitive value (about 25.67).

## Section 6:

To better understand the network, it was important to consider a threshold value for the proximity which can be used to filter the number of edges in the green product network. Using the proximity threshold $\varphi_0 > 0.25$ it is seen that the number of edges decreased from 26450 edges to 12161 edges which were important to analyze.

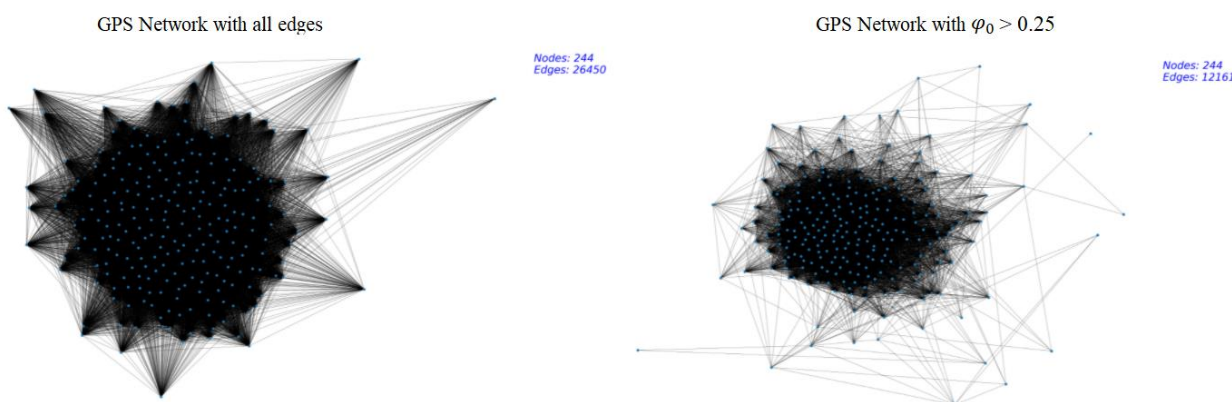


Fig 15: GPS Network (for the year 2020) before (left) and after (right) filtering the edges with proximity value less than 0.25

Since it was observed that GPS network for the year 2020 has more green products that fall in the category of REP in the green production basket of Florida and Texas, It would be interesting to filter green products that fall in REP category. Thus, we have filtered the REP category nodes to study their placement in network.

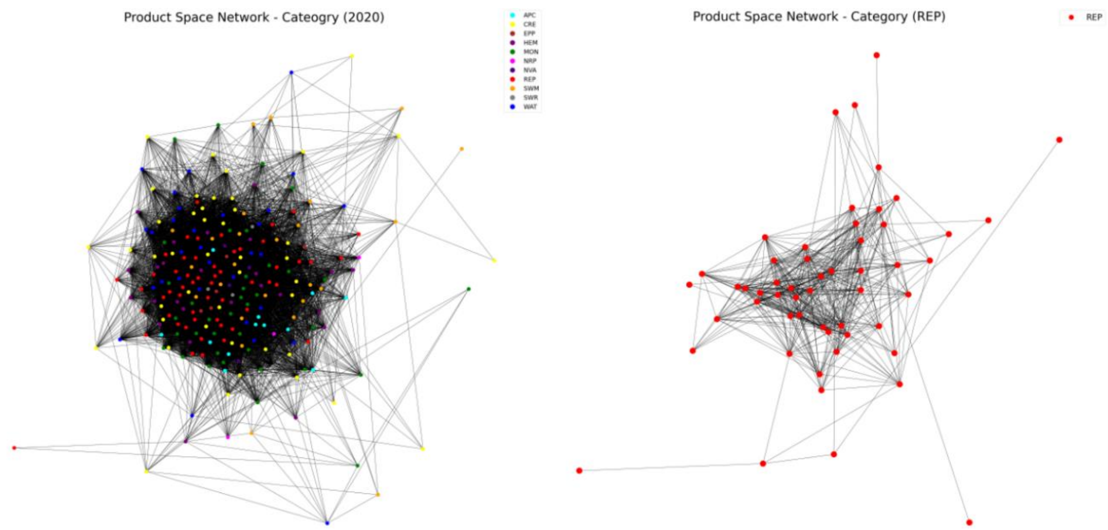


Fig 16: GPS Network (for the year 2020) with nodes colored based on 11 green product categories (left) and nodes filtered based of REP category

For better understanding we added labels to the filtered REP category green products and added HScode label to all the REP category green products (as shown in Fig 17)

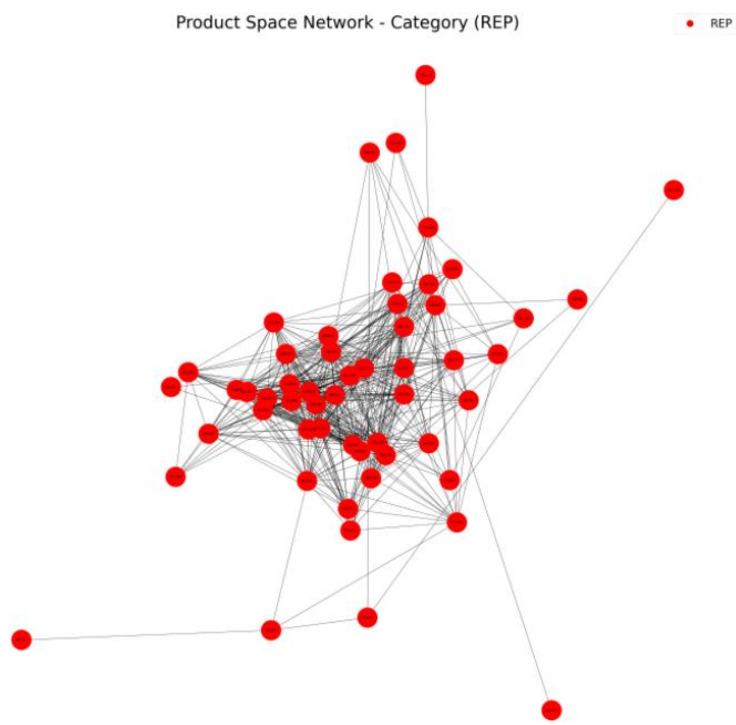


Fig 17: REP Category Green Products labeled based on their HScode

Based on the HScode label for REP category products, we can observe Table 2 which shows few REP category products that appear far from the center of the network.

| Hscode | Category | Green Product |
|---|---|---|
| 700991 | REP | Glass Mirrors Unframed Not Vehicle Rearview Mirror |
| 732119 | REP | Cook App & Plate Warmrs, Nonelec, Of Irn/stl Nesoi |
| 841011 | REP | Hydraulic Turbines, water Wheels, Not Ov 1,000 Kw |
| 841012 | REP | Hydraulic Turb & Wtr Wheels Power>1,000kw<10,000kw |
| 841013 | REP | Hydraulic Turbines And Water Wheels Power>10,000kw |
| 850163 | REP | Ac Generators (alternator) > 375 Kva But =< 750kva |

Table 3: Green Products that appear far from the center of the filtered network based on REP category

Hydraulic turbines production capabilities differ from the production capabilities of other renewable energy plant products. In fact, in the Fig 9 they are seen closer to the products of WAT (Waste water management and potable water treatment) category which is justified.

Based on the GPS network shown in Fig 11, it would be interesting to filter out the green products with RCA value > 1 for both Texas and Florida and see how they differ from each other. Thus, we built a visualization that shows only green products with RCA value > 1 while the nodes are sized based on their export value.

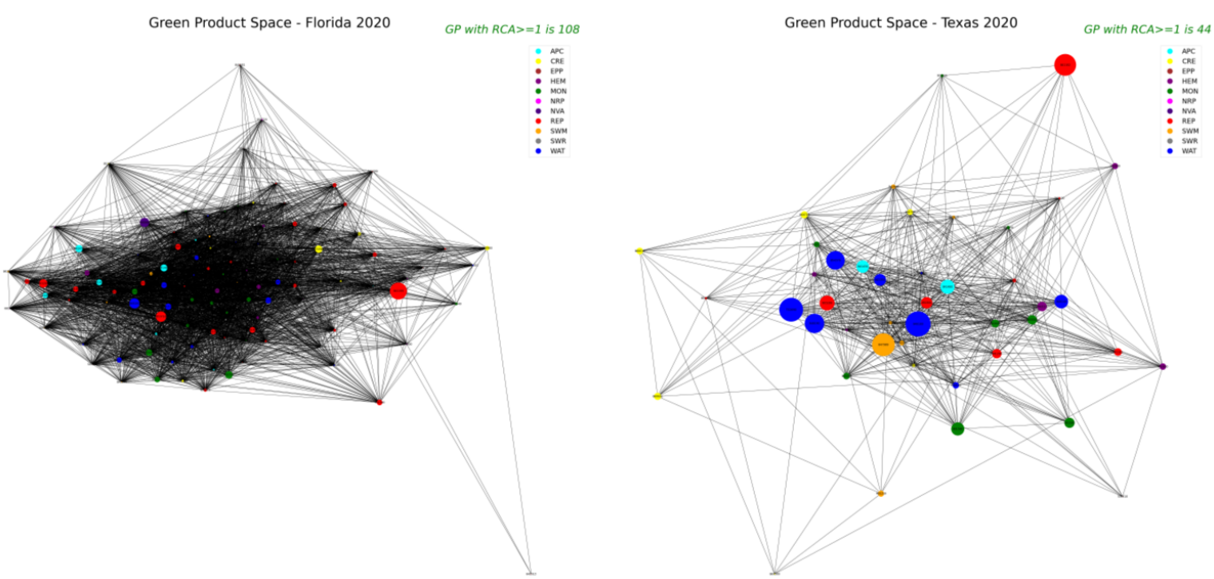


Fig 18: GPS network (for the year 2020) of Florida (left) and Texas (right) showing only products with RCA>1 and nodes resized based on export value

We can clearly see in Fig 18 that the GPS network of Florida is dense as there are higher number of green products when compared to that of Texas. Both states show diversification in their green production, but Texas appears to be more diversified than Florida with nodes far from the center of the network (not in proximity) but still capable of production.

## Section 7:

*Network Features*

From Fig 10, which highlights the GPS network based on green product category, we can see that most of the nodes that fall into the same category appear to be clustered together in the network. While we can see that all green products that fall into the category of REP (renewable energy plant) mostly fall into the center of the network, the green products falling into the CER (cleaner or more resource efficient technologies and products) category appear to be placed at the outer edges of the network. This difference in placements suggests that states that are competitively producing REP category products would find it difficult to transition to CER category products that are placed far from them in the GPS network.

One of the key research questions we had was to determine whether GPS network of Florida and Texas follow path dependency or not. To analyze the path-dependency we considered 3 time intervals – [2016-2018], [2017-2019], [2018-2020] years. All green products (with RCA value > 1) of a state that were not part of green production basket at base year but had entered the states green production basket at later target year were filtered as new green products.

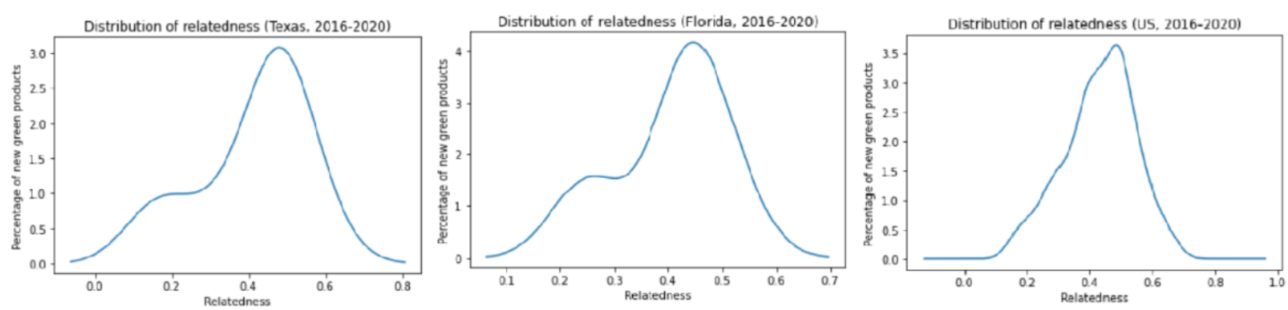


Fig 19: Kernel distribution relatedness for Texas, Florida and US based on the GPS network

Based on earlier research [2,3], For each new green product maximum proximity (relatedness) is calculated and these maximum proximity values for Texas, Florida and US are used to plot the kernel distribution of relatedness. The Fig 19 shows the relatedness distributions of Texas, Florida, and US where we can clearly see that the Florida, Texas and US follow both path-dependency and non-path-dependency as they have relatedness values between 0.2 to 0.6. However, one thing that we can interpret from the distributions is that they all are more inclined towards following non-path-dependency also which allows them to make high frequency jumps in the network to expand their green production.

## Section 8:

*Part 1: Results*

Through the detailed analysis executed on the GPS network of Texas and Florida, we were able to answer the research questions that were listed at the beginning of our project. Below are the detailed deductions made based on our analysis.

GPS Network of Florida and Texas:

From Fig 9, we can say that green production basket of Florida is larger when compared to Texas. However, Texas shows that its green products with high RCA value (RCA>1) are scattered across the network suggesting that it has the capability to expand its green production through structural jumps without following path dependency in the network. However, Florida shows most of its green products with high RCA values concentrated at one place in the center of the network suggesting they can grow their green production basket using capabilities of similar products [5].

Evolution of GPS Network of Florida and Texas:

From Fig 20, we can see that the nodes (green products) having RCA value > 1 are randomly placed in the network which suggests that over the years the green production basket of Texas does not show much of path-dependency. It is observed that the green production basket of Texas has decreased over the years.

If we observe Fig 21 in detail, we can see that the green production basket of Florida in 2016 was more diversified when compared to 2020 which appears to be more concentrated. GPS network of Florida shows how over the years its green production basket has been larger (more diversified) than the green production basket of Texas.

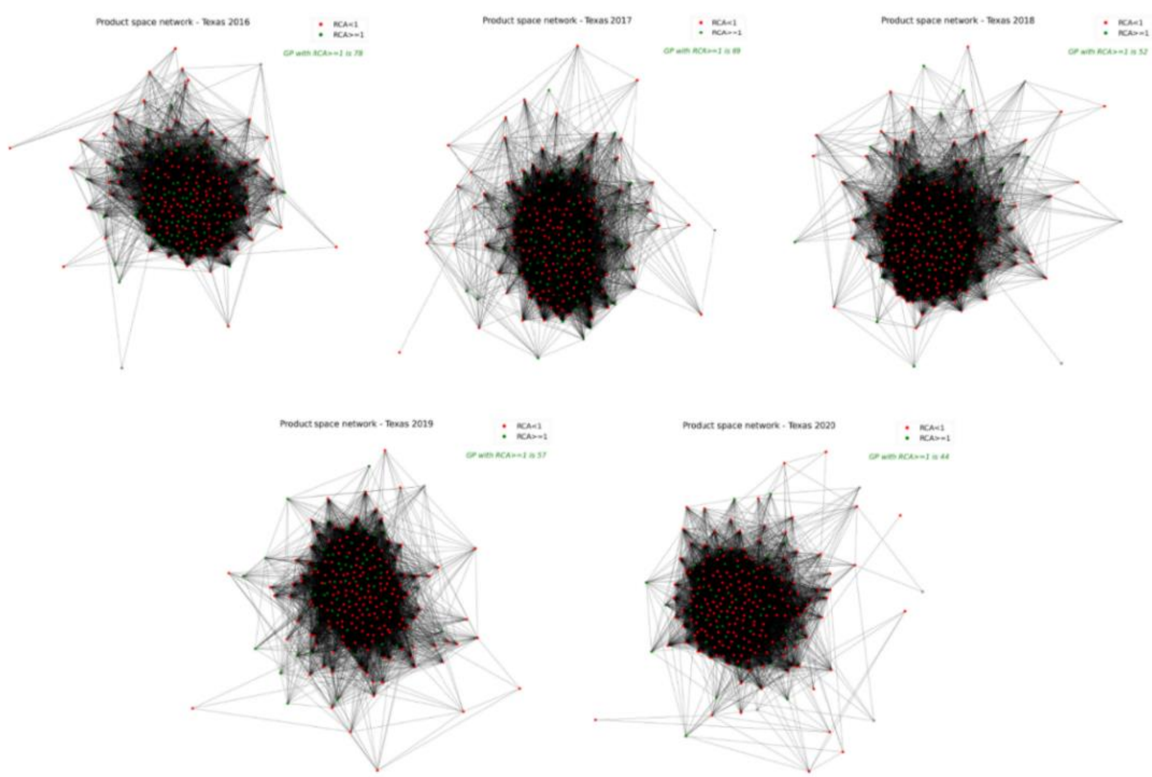


Fig 20: Evolution of GPS network of Texas over the years (2016 to 2020)

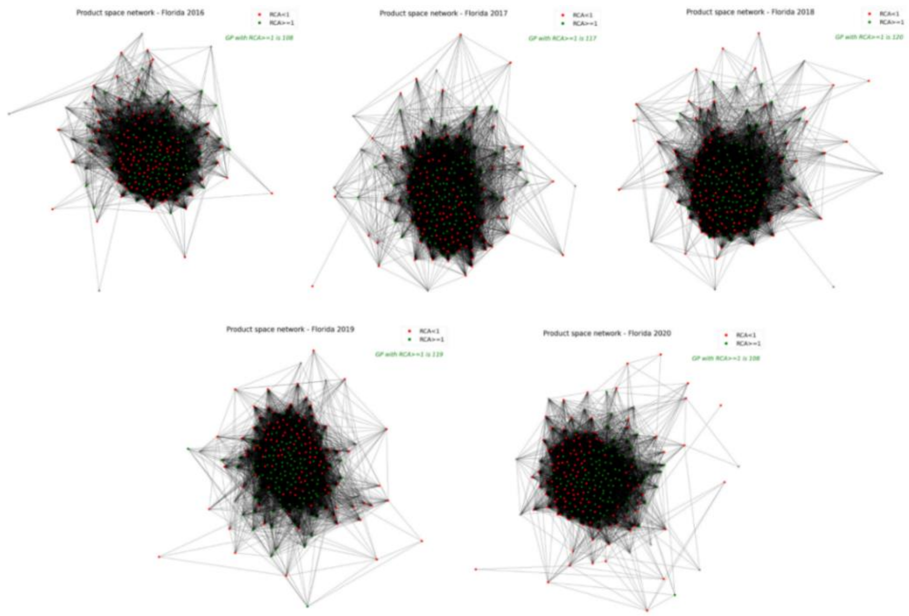

Fig 21: Evolution of GPS network of Florida over the years (2016 to 2020)

Path-dependency in Florida & Texas GPS Network:

From Fig 22, which shows distribution of relatedness (2016 to 2020) for GPS network of Texas, Florida, and US, we can deduce that Florida & Texas both follow path-dependency, as well as non-path dependency in their green production.

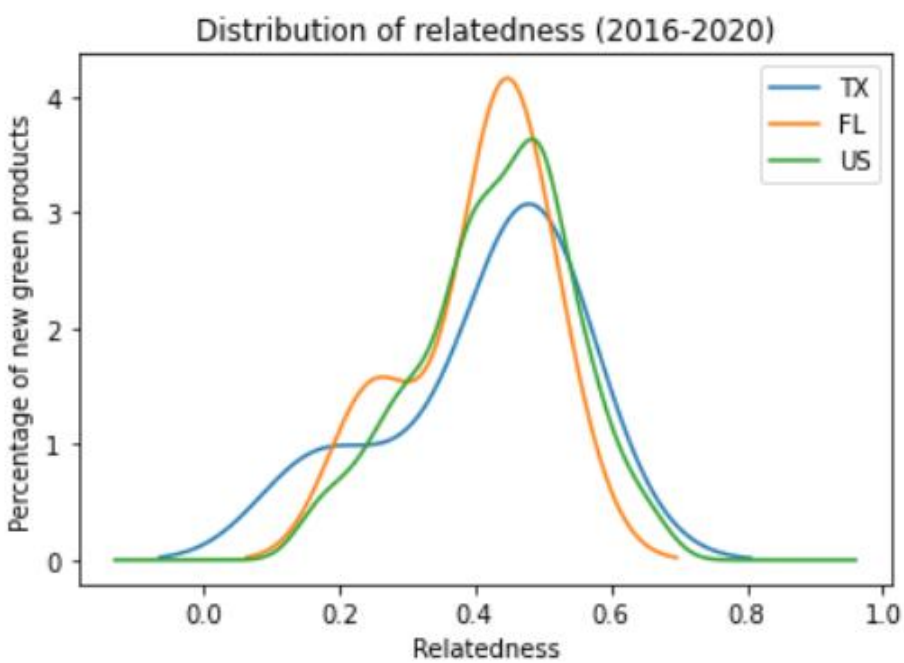


Fig 22: Distribution of relatedness (2016 to 2020) for GPS network of Texas, Florida, and US

Since we considered maximum proximity for the products that are highly similar together and majority of them fall between 0.2 to 0.6, it shows that Texas, Florida, and US follow both path-dependency and non-path-dependency. But overall, they show more non-path-dependency suggesting they are capable of making high frequency jumps in the network to start producing new green products that are not related to their existing capabilities. From the plot, we can say that Texas is following more non-path-dependency when compared to Florida and it is even above US showing that they are taking huge jumps in the GPS to be able to expand their green product basket.

<u>Future Strategy & Recommendations:</u>

From Fig 23, we can clearly see that the pollutant emissions of Texas are quite high compared to those of Florida. The difference in the green production baskets of the two states could be one of the key reasons for this difference. Since Texas shows more non-path dependency compared to Florida, we are suggesting that it can take structural jumps to expand its green production basket, we recommend policy makers use their innovation and technological advancements to enable these structural jumps in the GPS network and expand their green production.

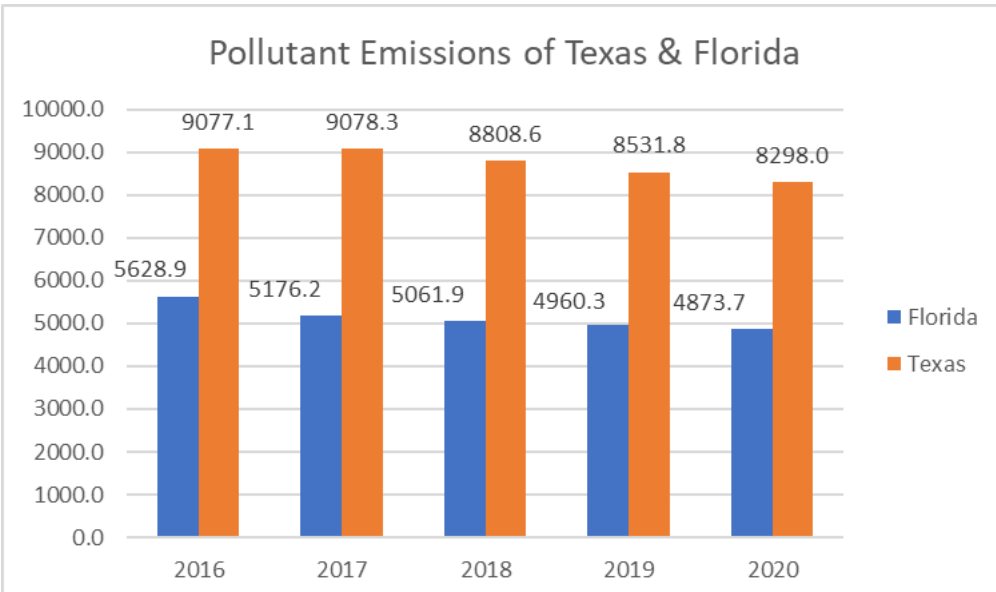


Fig 23: Overall Pollutant Emissions of Texas & Florida over the years (2016 to 2020)

This will ensure a sustainable future for the state of Texas. Though we see decrease in the pollutant emissions over the years there can be other factors that influence this change and to further decrease the trend Florida can try to develop new strategies to go beyond REP (Renewable Energy Plant) category green products to include more green products of MON (Environmental monitoring, analysis and assessment equipment) & WAT (Waste water management and potable water treatment) category using path dependency (based on the GPS Network for the year 2020 in Fig 10).

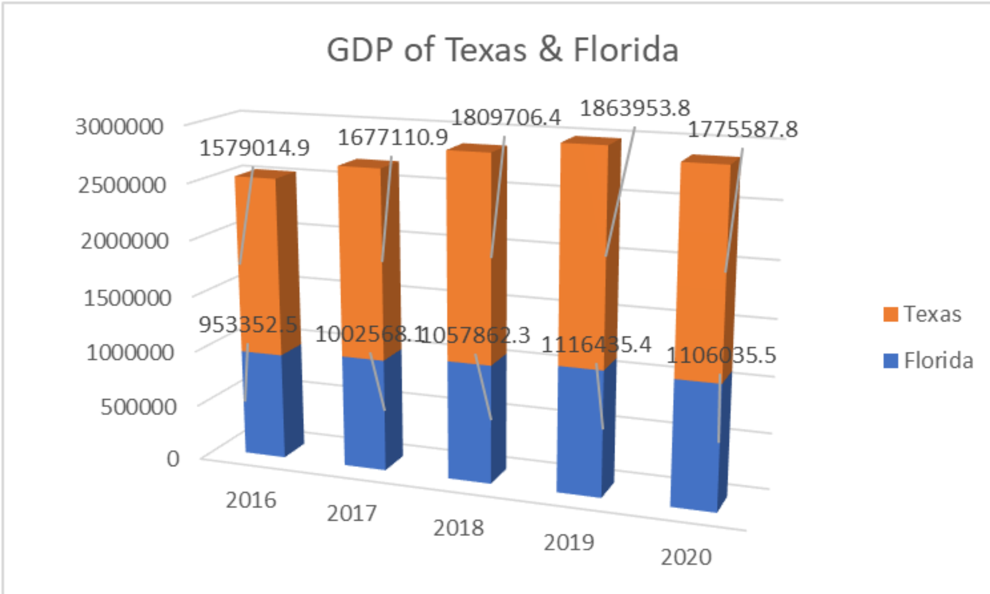


Fig 24: Gross Domestic Production (GDP) of Texas and Florida over the years (2016 to 2020)

From Fig 24, we observe that the GDP for Texas and Florida has decreased between the years 2019 to 2020. There are several factors that would have contributed to this decrease (like COVID-19 pandemic) but among these factors we can say that the decrease in green production basket of the two states could also be one of the key contributors. Thus, we suggest that employing techniques that lead to increase in green production will enable greater economic growth for the two states. Policy makers can build new strategies that decrease tariffs and provides subsidies to organizations that promote green production.

### *Part 2: Final Contemplation*

To further enhance our study, it would be interesting to examine the technological innovation of the two states over the years, which can help us to further assess the reason for the difference in the GPS network between these two states. Technological innovation can be measured based on the number of patents filed by each state in the US in the field of green development. Other points of data that can be examined to further improve our analysis are the industrial tariffs and subsidies of the US states and how they change over time.

Through this project, we were able to understand some key concepts in economics and network science, which has been a fruitful journey. One challenge that we encountered while building the network was to compute the correct proximity values that lie between 0 and 1. To compute the proximities correctly, we had to analyze our definition of the world in the RCA calculation, which led us to download data for all states in the US. For each network visualization that was built as part of the project, we equally divided individual time to make insightful deductions that could increase the quality of our research. Since the project required both technical and analytical expertise, we decided to brainstorm together to understand the concept through different research papers and then assigned individual tasks for each project member.

The GPS network built for this project was developed using Python so that relevant customizations could be added to the network. As part of data preparation, multiple csv files that covered information around the export value of all commodities for each state in the US were downloaded and concatenated in a python notebook. Next, the green product list was used to filter the green products for each US state, which could be used to calculate the RCA and proximity values. Custom python functions were developed to build the RCA and proximity matrix for all years (2016 to 2020). Using the calculated RCA and proximity values, an edges and nodes list were created and loaded into Gephi to run some statistical analysis on the network. Later, Python's networkX library was used to run different layouts on the network and deductions were made to build the final GPS network for the year 2020. As a next step, RCA values greater than 1 were considered to be color nodes (green products) for Texas and Florida. Since our focus areas for the project were Texas and Florida, we initially decided to consider our world definition in RCA to be the combined green products of Texas and Florida. But as we progressed in the project, we realized the importance of considering all US states when building the GPS network, which made us download data for all states that were not part of the initial project plan.

Finally, we want to express our gratitude to Dr. Seyyedmilad Talebzadehhosseini, for his constant support and mentorship that has helped us successfully complete the project and to Dr. Edwin Nassif for introducing us to the network science concepts and providing us the opportunity to work towards this project.

**References:**

1. Sauvage, J. (2014), "The Stringency of Environmental Regulations and Trade in Environmental Goods", *OECD Trade and Environment Working Papers*, No. 2014/03, OECD Publishing, Paris, https://doi.org/10.1787/5jxrjn7xsnmq-en.
2. Coniglio, N.D., Vurchio, D., Cantore, N., Clara, M. (2018). On the Evolution of Comparative Advantage: Path-Dependent Versus Path-Defying Changes. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3136471 . Accessed 15 April 2019.
3. Talebzadehhosseini, S., Scheinert, S.R., Rajabi, A., Saeidi, M., Garibay, I.: The pattern of economies green growth: the role of path dependency in green economy expansion. arXiv: 1906.05269 (2019)
4. Hidalgo, C.A., Klinger, B., Barabási, A.L., Hausmann, R.: The product space conditions the development of nations. Science 317(5837), 482–487 (2007)
5. Dosia, G., Trancherob, M.: The Role of Comparative Advantage, Endowments and Technology in Structural Transformation (2019)