Mapping the Shape of the U.S. Economy: A Topological Data Analysis Approach with BallMapper

Ryan Johnson[*]

[?] Department of Mathematics, University of Alaska Anchorage, Anchorage, AK

Student: *johnson.ryan-0@pm.me*[*]
Mentor: *scook25@alaska.edu*

KEYWORDS

Topological Data Analysis; BallMapper; Data Science; Economics; Macroeconomics; Topology;

ABSTRACT

Topological Data Analysis is (TDA) is a new data analysis method which gained popularity starting in the early 21st century. Currently, a large body of TDA research utilizes the traditional Mapper algorithm. We aim to fill an entry level gap into TDA as well as expand the body of literature on BallMapper, a new, Mapper adjacent algorithm. Using widely used U.S. macroeconomic data from the Bureau of Economic Analysis (BEA), the Federal Reserve Bank (FRB), and Bureau of Labor Statistics (BLS), we do this is three parts. We show an example of BallMapper's utility in exploratory data analysis which also provides an example of how to interpret BallMapper's output; analyze our topological graphs to notable historic economic events; and test the stability of BallMapper's output and the topological graph's features. Results show. [...]

INTRODUCTION

Topological Data Analysis (TDA) is a new and emerging field of data analysis that is increasing in popularity. Broadly, traditional TDA applications use two tools: an algorithm called Mapper that produces a network-like graph and and analysis technique called Persistence Homology. Mapper can be thought of the exploratory data analysis (EDA) and visual display of statistical analysis. That is, we may see patterns in our data or need a visual tool to help explain the results of an analysis. Persistence Homology on the other hand is more akin to statistical models like the Generalized Linear Model or Bayesian Statistics. They provide a mathematical framework and used as a tool to give a researcher confidence in where to look further or what they might be seeing in their data. Moreover, the combination of Mapper and Persistence Homology is what forms the central argument for TDA: data contains an underlying shape.[1,2]

For this analysis we will be focusing on the graph creation portion of TDA. Specifically, we are examining various macroeconomic indicators of the United States of America (U.S.) using a Mapper-adjacent algorithm called BallMapper(BM).[3] BallMapper is of particular interest to us because it significantly reduces the parameters to create a topological map. It simplifies traditional Mapper by reducing the need for the user to pass data through multiple functions, each with their own parameters.[A] One benefit to this reduction in parameters is that it removes some of the barriers to learning Mapper (or Mapper-like algorithms), and more generally, TDA. Those who have a mathematical background, introductory course in Topology, or conceptual knowledge of data science methods and algorithms will fare much easier. However, if you are not equipped with any of the aforementioned tools, TDA might seem unnecessarily complicated for data analysis.

Beyond the main objectives of this analysis, we hope that this paper is serve as an on-ramp for anyone interested in TDA but feels inundated with jargon upon early stage researching. Out objectives are to add to the small body of literature whose main focus is applications with BallMapper.[1] We go about this by first showing BallMapper's use in exploratory data analysis (EDA). Using our results from EDA, we then will compare BallMapper's outputs (topological graphs) to notable historic economic events. Lastly we will test the stability and consistency of BallMapper's output. The conclusion will talk broadly about results and other considerations of note, and the discussion will elaborate on future work and questions for research.

[A] add reference

Generally speaking, traditional TDA contains four broad steps, each with multiple user-defined parameters.

Conversely, TDABM has reduced this process to an the user selecting their data, a coloring variable and an epsilon value – details on this follow below. What should be noted, although TDABM reduces the number of steps needed to produce similar outcomes of traditional TDA methods, we lose control of being able fine tune out outputs.

We also find that interpreting results becomes more difficult.

However, TDABM is still in its infancy, so there is not a large body of research on interpretation.[?]

One major motivation for this paper is a paper written by[?] that applied TDABM to a global macroeconomic dataset to compare how countries have evolved over time, their transformation from the Great Depression Era, and various views on wealth and inequality. We could not find other TDA literature found specifically focused on the macroeconomic economy of singular countries. Hence, our topic of choice.

METHODS AND PROCEDURES

Data Selection & Preparation

This paper relied on three publicly available data from US government sources. The Bureau of Economic Analysis (BEA), The Federal Reserve Board (FRB), and The Bureau of Labor Statistics (BLS).[4,5] The data were gathered using R using two application programming interfaces (APIs): one for data from the BEA, and the other from the Federal Reserve Economic Data (FRED) API.[6,7] FRED aggregates data from national and international sources, as well as public and private sources. We additionally used recession dates based on business cycle contractions and expansions provided by the National Bureau of Economic Research (NBER). These agencies were selected because of they are authoritative sources for U.S. economic data. Their widespread use in both the private and public sector gives us high confidence in the accuracy and integrity of the data.[8]

Data Series	Series Abbreviation	Year Start	Year End	Data Frequency	Data Type	API
Gross Domestic Product	GDP	1930	2024	Annual	Flow	BEA
Personal Income & Its Disposition	PID	1948	2024	Annual	Flow	BEA
Foreign Direct Investment	FDI	2014	2024	Annual	Flow	BEA
Federal Funds Rate	FFR	1955	2024	Annual	Rate	FRED
Employment Cost Index	ECI	2001	2025	Quarterly	Index	FRED
Consumer Price Index	CPI	1947	2025	Monthly	Index	FRED
New Privately-Owned Housing Units Started	Housing Starts	1959	2025	Monthly	Flow*	FRED
Producer Price Index - All Commodities	PPI	1913	2025	Monthly	Index	FRED
Unemployment Rate	Unrate	1948	2025	Monthly	Rate	FRED

Table 1: *Housing Starts is considered a flow because it is provided as a Seasonally Adjusted Annual Rate (SAAR).

Table 1 shows all the data series we considered for this analysis. We excluded Employment Cost Index (ECI) and Foreign Direct Investment (FDI) due to their limited availability of years. Of the remaining data series, New Privately-Owned Housing Units Started (Housing Starts) has the smallest range and will provide the lower end for the years used in our analysis (1960-2024). Table 2 shows the final data series we will be using. These specific series were chosen because they represent different aspects of the macroeconomy, and are described in *Functional Description*. Additionally, we are focusing on an annual time frame for this analysis so some data transformation was needed. The final data series transformations follow.

Series Abbreviation	Economic Role	Business Cycle Timing	Functional Descriptioin	Transformation Applied
GDP	Growth Composition	Concurrent	Consumption, Investment, Gov., and Trade	None (Source in % Change)
PID	Income Structure	Concurrent	Wages, Entrepreneurship, Gov. Transfers	Real Adjustment & Log Difference
UnRate	Labor Market Dynamics	Lagging	Labor Distress Level and Trend	None (Annual Rate) & Simple Difference
FFR	Monetary Policy	Policy/Reactive	Cost of Capital and Borrowing Conditions	None (Annual Rate) & Simple Difference
PPI	Supply-Side Signal	Leading	Producer Costs & Measures Cost-Push Inflation	12-month Avg. & Log Difference
CPI	Demand-Side Signal	Lagging	Cost of Living & Measures Demand-Pull Inflation	12-month Avg. & Log Difference
Housing Starts	Leading Indicator	Leading	Physical Residential Production	12-month Avg. & Log Difference

Table 2: All years for analysis are 1960-2024.

The first transformation was Personal Income and Its Disposition (Personal Income).[B] Personal Income is reported in nominal dollars, so it does not account for inflation. Thus, we first adjusted it using the Consumer Price Index (CPI) to get Personal Income into Real Dollars.[9]

$$\text{Real Dollars} = \frac{\text{Nominal Dollars}}{\text{CPI}} * 100 \tag{1}$$

We then took the the log difference (Equation 2) from the output from Equation 1 to get a percentage change from the preceding year. This log difference helps to make our data more linear which aides BallMapper's use of the standard Euclidean Distance between each row of data. (CHECK - If BM did not use the Euclidean distance between points, we might not need to take log differences. However topic is for another paper.)[10]

$$\Delta \ln(\text{Level}) = [\ln(\text{Level}_t) - \ln(\text{Level}_{t-1})] * 100 \tag{2}$$

Our next transformation was on Housing Starts, Producer Price Index - All Commodities (PPI) and CPI. We should note that PPI tracks only physical goods such as: Farm Products and Feed; Industrial Commodities; and Other Commodities such as Aircraft, Steel, and Lumber.[11] All three of these sources were only provided on a monthly time frame, so to get an annual value we used a simple arithmetic mean. Once these data are in annual form, we then found the percent change from the previous year using Equation 2 for their final data in our analysis.

[Need to add parts more about the data itself?]

METHODS

BallMapper

In depth descriptions of Ball Mapper's theory have been covered in various papers. As the aim of our paper is to provided a introductory foray into BallMapper and TDA in general, we will omit some of the more technical details below but provide reference to more in-depth explanations.

Mathematically, the concept of Ball Mapper supposes we are given a dataset X in K dimensions with N observations. Then for some point $x \in X$ and given $\epsilon > 0$, we create a ball, $b(x, \epsilon)$, centered around x with radius ϵ. Our aim to create set of balls, B, such that $B = \bigcup_{i=0}^{n} b(x, \epsilon)$ for all $x \in B$.

We call our dataset X a Point Cloud where each of the N rows represents a point in our cloud. Our dimension, K, are each column of our data and commonly our point cloud has dimensions $K > 2$. When we are selecting a point, $x_i \in X$, to draw a ball, we are randomly selecting a row of our point cloud and then taking he Euclidean Distance in K dimensions such that $i \neq j$ and $x_i, x_j \in X$ to every

---

[B]Table 2.1, BEA NIPA

other x_j in our point cloud. Then with our given epsilon, $\epsilon > 0$, we assign any points where the distance is less to epsilon to a ball with center x_i (Equation 3).

$$d(x_i, x_j) = \sqrt{(x_{i_1} - x_{j_1})^2 + (x_{i_2} - x_{j_2})^2 + ... + (x_{i_k} - x_{j_k})^2} < \epsilon \tag{3}$$

As an algorithm, Ball Mapper is as follows:

1. Select a random point x_i from point cloud X.
2. Given $\epsilon > 0$, construct a ball $b(x_i, \epsilon)$ with a center x_i by associating all other points where $d(x_i, x_j) < \epsilon$ and $i \neq j$. (Equation 3)
3. Place $b_i, \epsilon)$ in a set of balls B.
4. Repeat steps 1-3 until all $x \in X$ belong to some $b(x, \epsilon)$.
5. Draw an edge between $b_i, b_j \in B$ if they contain the same x, weighting the edge based on number distinct x values b_i and b_j both contain.

Analysis

As we saw above we need three parts for BallMapper: a Point Cloud, a variable to color by, and an epsilon. For this analysis, our point cloud was made up of 16 variables (dimensions) which can be see in Table 3. It is important to note two dimensions in our analysis were created: *Unemployment Change* and *Fed Rate Change*. [Add reasoning] All other variables where transformed but no new data created from them. Albeit, Unemployment is a lagging indicator and the Federal Funds Rate is a policy or reactive choice, we thought these two measures were important. Since Unemployment is a flow, is it useful to see if the hose of unemployed is expanding or shrinking. Similarly, it is informative to see changes in the Federal Funds Rate to see the policy choices made are working or not.

Data Series	Analysis Name
All	Year
Producer Price Index - All Commodities	PPI Change
Personal Income & Its Disposition	Personal Income
Personal Income & Its Disposition	Compensation
Personal Income & Its Disposition	Entrepreneurship
Personal Income & Its Disposition	Transfers
Gross Domestic Product	Consumption
Gross Domestic Product	Domestic Investment
Gross Domestic Product	Government Spending
Gross Domestic Product	Exports
Gross Domestic Product	Imports
Consumer Price Index	Inflation
Unemployment Rate	Unemployment
Unemployment Rate	Unemployment Change
Federal Funds Rate	Fed Rate
Federal Funds Rate	Fed Rate Change
New Privately-Owned Housing Units Started	Housing Change

Table 3: Variables inculde in BallMapper Point Cloud

To our knowledge there is no generally accepted way of choosing epsilon in the literature. Since BallMapper constructs maps by selecting a point and taking a radius distance of epsilon, this is should in theory, if given an epsilon small enough, produce a graph such that every point is its own ball. Conversely, this should lead to all of our points being included in one node. For our analysis here, this is indeed the case. So to find an appropriate epsilon we first start by finding these lower and upper bounds. What we look for in our lower bound is a graph is an epsilon small enough make every point in our point cloud a singleton node, not connected to any other. For our upper bound we do the converse and look for an epsilon so large where it creates a singular node that houses all of our points. Additionally, to reduce some of the computational time, we look for a lower bound where a connected component started to form, and an upper bound where we had only two connected nodes.

For our data we first found a lower and upper bound which created an interval of $[0.3 - 1.4]$. We then applied a function to generated around 100 graphs and reviewed them, narrowing our interval to $[0.4 - 0.75]$. During this narrowing process we look for interesting features such as connected components forming or dissolving, flares coming off of any components, or notable sizing or coloration

patterns. We finally decided on the value 0.511 because it presented a little of all the aforementioned features we look for: coloration and size patterning and connected components.

The standard BallMapper package in R produces a sufficient graph for exploratory data analysis. It's graphing function uses R base plot function and default coloring scheme. We however, felt that this graph output could be improved upon and recognized that we could use alternative, and more comprehensive packages to create our graph. That is, the output of the BallMapper algorithm could use an alternative package specifically for a network-like graph such as ours. Thus, translating the direct output from the BallMapper algorithm, we created the same graph but changed the visual representation and added another helpful piece of information that we were given but not used previously. Looking at Figure 1, we have the standard Ball Mapper graph output on the left (Figure 1a) and our new graph output on the right (Figure 1b). The most prominent changes is the spread of the network and coloration.

(a) Ball Mapper output using standard function (ColorIgraphPlot)   (b) Ball Maper output using new graph output (ggraph)
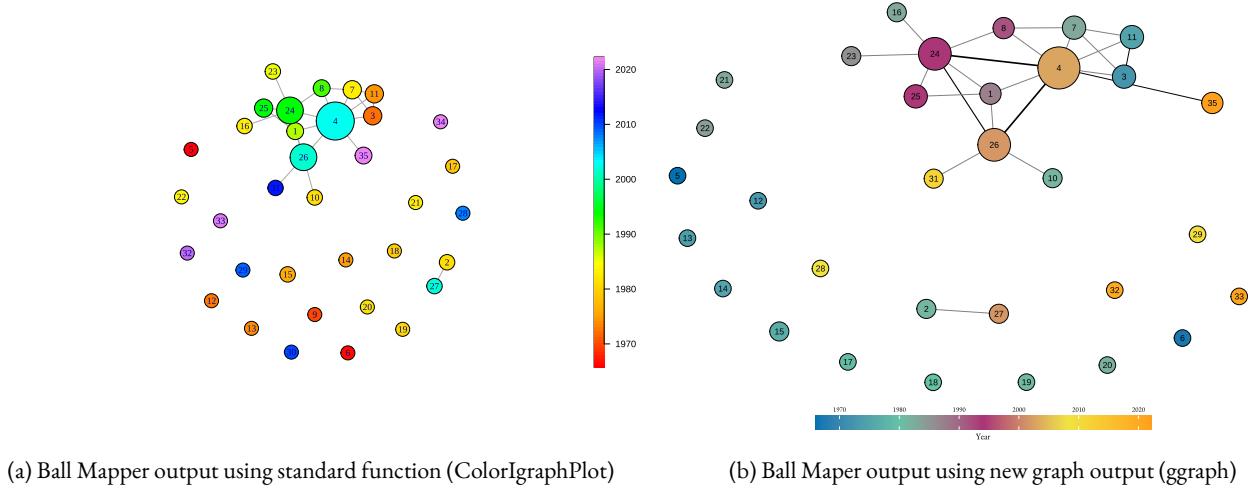
Figure 1: Ball Mapper Graph colored by Year with Epsilon set to 0.511.

Having the nodes spread out helps us see the structure of our graph more clearly as well as the new piece of information that was given to use but not see in Figure 1a, the edge strength. We see that in Figure 1b, highlighting the edge strength presents an insight about the relation between certain nodes. Further, out data consists of 65 rows, but one could imagine the advantage of a data set much larger that this one (e.g. looking at this data on a quarterly or monthly basis). Additionally, we changed the color palette for accessibility and readability. Looking at Figure 1a, any nodes colored in the royal blue (years approximately 2010-2015) mask the number identifier of the node.

For Figure 1b and Figure 2, below, we use Year as our coloring variable. We started with this variable because a large motivation for this paper came from a paper by Dłotko, Rudkin, and Qiu, *Topologically Mapping the Macroeconomy*. In this paper was a two pronged macroeconomic analysis with Ball Mapper. The first was a dataset of five macroeconomic variables from 16 countries from the 1870s-2017. Coloring by Year they looked at the Ball Mapper graph structure as well as the the effect and recovery of the Great Depression on these countries. The second prong was using Ball Mapper to look at relationship between private credit growth and GDP using Ball Mapper. This part of the paper they colored by various dimension within their point cloud as well as Year which was not included in the their point cloud.[12]

because Year is a metric that is not dependent on any economic variables and is a good marker for large macroeconomic events such as the dot com bubble of the 2000s. *Year* is not included in the construction of our Point Cloud. This is an important point because we show later that being able to create metrics to color by adds a depth the Ball Mapper's usefulness, even if the metrics themselves are not included in our Point Cloud.

CONCLUSION

Exploratory Data Analysis (EDA) Application

For our exploratory data analysis exploration we colored our graph by the *Year*, Figure 2. The coloration of each node represents the average of all the years within in a node. The size of our node is commensurate with the number of years it houses. Between two nodes, th edge thickness is determined the number of same data points in each endpoint.Looking at Figure 2, we first observe the overall structure of our graph. There is one large connected component, one small component, and the rest singleton nodes.[c] For the largest connected component (C1), we immediately are interested in the three large nodes (4, 24, and 26) because they are the largest in size for C1 and the same for their edge thickness (edge strength). Additionally, we also notice that each of the three large nodes have at least one flare (node connected only to themselves) as well as sharing others. What this structure tells us is, there are three distinct groups and each flare representing similar data to that specific group but perhaps is distinctly different to the other flares.
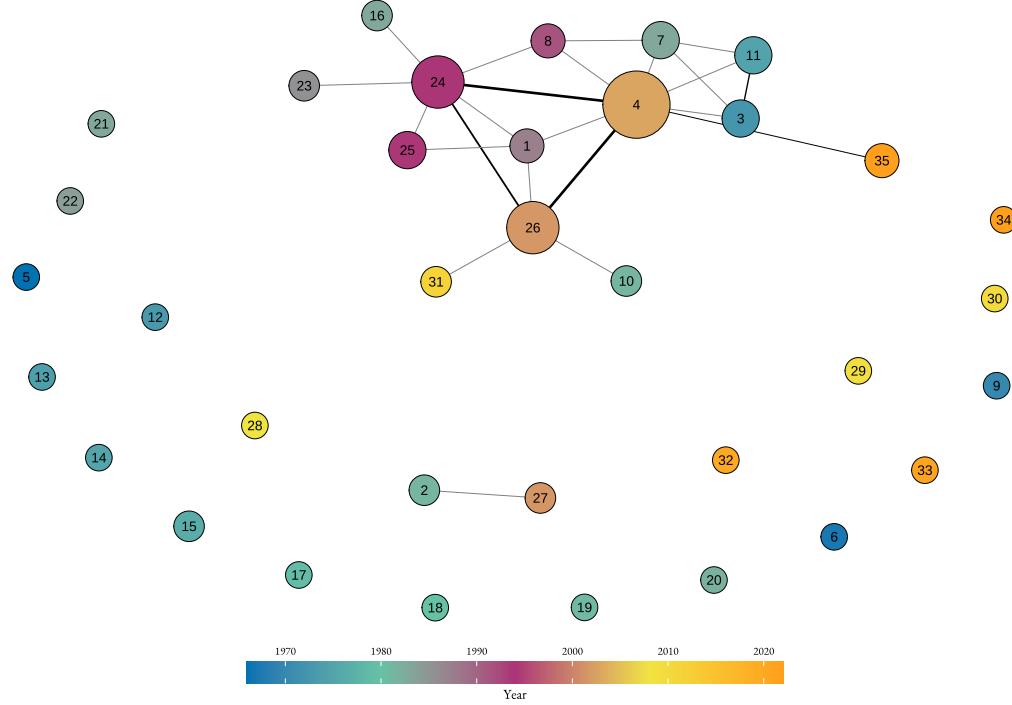

Figure 2: Ball Mapper Graph colored by Year with Epsilon set to 0.511.

The small component (C2), nodes 2 and 27, are of interest because it tells us us that these related nodes are somehow distinctly different from C1 and all other nodes. Seeing a small components like this sometimes can indicate outliers in data. In our case, this could be years where large economic events happened such as a recessions or extraordinary growth. However, we also can see this behavior in singleton nodes much like we have Figure 2. When we look at the coloration of the singleton nodes, we see indication of outlier events such as The Great Recession and the COVID-19 Pandemic.

To hark back to our discussion of C1, when we investigate nodes 4, 24, and 26 we indeed notice a pattern with the three generally: for the whole year the US was in an expansionary state. Node 4 shows us an "ideal" version of the economy. We have personal income and compensation slightly higher than inflation, sitting around 2%. Unemployment is a little elevated but historically not unusually high. For Nodes 24 and 26 we see two different sides of large economic shocks. Node 24 shows us years where the economy is starting to get "over it's skis". That is, it is starting growing faster than it can keep up with but it's not necessarily too late. Node 26 on the other hand consists of years following right after recessions or points where the economy was over heating. Most of the years in Node 26, follow what economist call a "Jobless Recovery".[13] Usually marked by growth returning to the economy but unemployment being "sticky" and not falling as business starts to take off. For the flares and smaller points that are connect each of our large nodes in C1, we find that some years are in both the smaller node and larger one, which creates our edge. However they also contain additional years which are related nodes 4, 24, or 26 but instead describe transition years to these large nodes.

Fisher Effect

---

[c]A Connected Component is a subset of nodes in our graph such that each point in the subset has a path to all other points in the subset.

Using the Fisher Equation we can color our topological graph with the approximated real interested rate (Equation 4).[14] Although we know that the Fisher Equation is an approximation and does not hold well when nominal interest rates (Fed Rate) are relatively high and time spans are short, Figure 3 shows us a stark picture of our economic history and important time periods.[15]

$$\text{Real Interest Rate} \approx \text{Fed Rate} - \text{Inflation} \tag{4}$$

Looking at nodes 19-22, we notice that these years consists of the years 1981-1984, the years during and following the "Volker Shock".[16] What is interesting about why these nodes are colored is that these nodes see to show the impact of Volker's decision to raise interest and the lag that took place when raising rates.[17]

On the lower end the of Fisher Effect, we notice nodes 14, 28, 31, 33, 34 (1975, 2008, 2011-2012, 2021, 2022 respectively) show a negative value for Equation 4. In these year rates were already extremely low, or the economy just came off a huge economic shock. In 1975 the fed aggressively lowered rates due to climbing inflation and high unemployment rates from the 1973-'74 oil embargo.[18] In 2021-'22 the COVID-19 Pandemic stimulus checks were going to cause inflation but the fed did not want to raise rates too aggressively.

We also observe that for our C1 component, node 24 does show an approximate value from our Fisher Equation to be above the stated target rate of 2%. This indicates that indeed the economy for these nodes are showing signs of a "hot" economy.
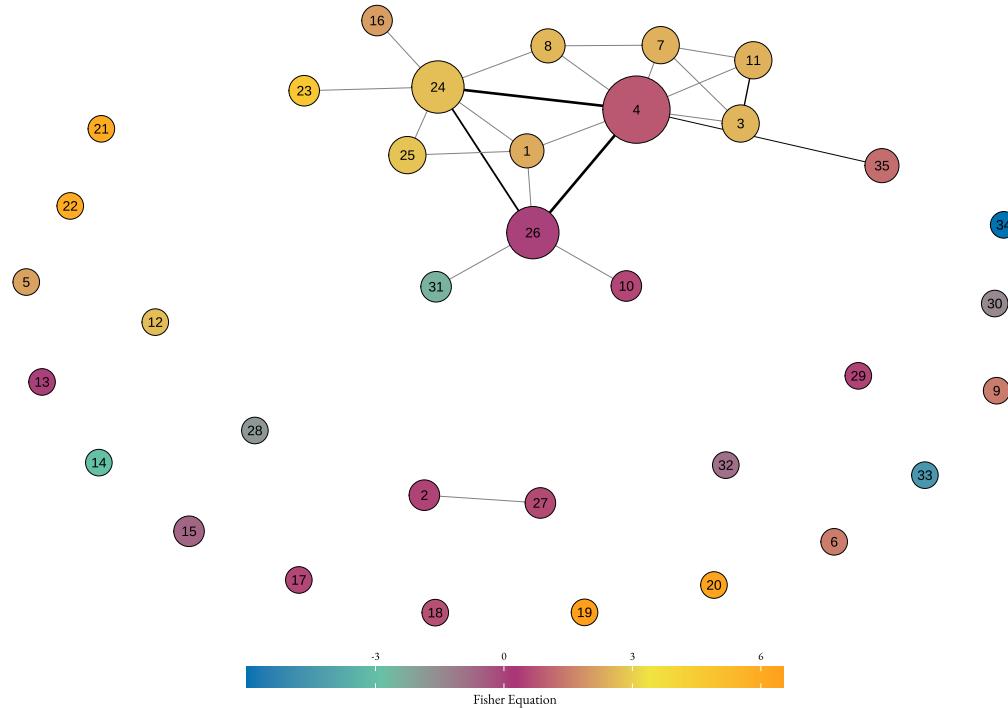

Figure 3: Ball Mapper Graph colored by Fisher Equation with Epsilon set to 0.511.
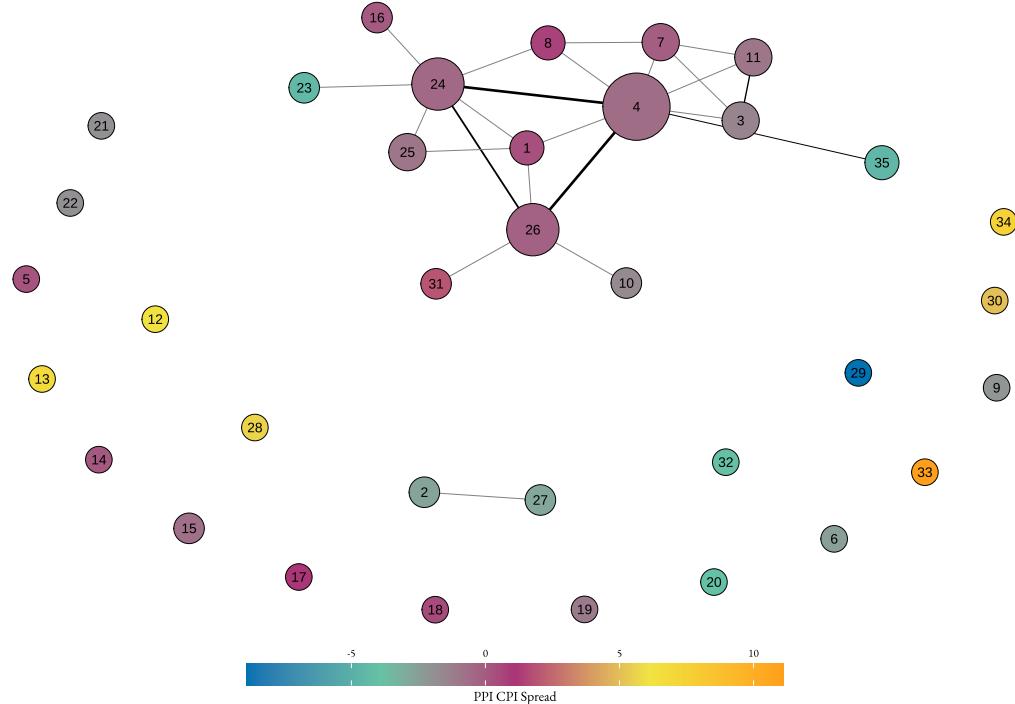
Figure 4: Ball Mapper Graph colored by the difference between PPI and CPI with Epsilon set to 0.511.

Robustness and Stabilization

- Shuffle Dataset –> Run BallMapper
    - Look at:
        *

        # nodes

        * coloring consistency
        * look at financne application and brexit vote

DISCUSSION

REFERENCES

SUPPLEMENTARY

```
node_4 <- final_data[V(BM_Year)$members[4] |> unlist(), ]
node_24 <- final_data[V(BM_Year)$members[24] |> unlist(), ]
node_26 <- final_data[V(BM_Year)$members[26] |> unlist(), ]

node_4 <- node_4 |>
```

```
  summarise(across(-Year, mean)) |>
  pivot_longer(everything(), names_to = "Measure",
              values_to = "Node 4 Avg. - The Sweet Spot")
node_24 <- node_24 |>
  summarise(across(-Year, mean)) |>
  pivot_longer(everything(), names_to = "Measure",
              values_to = "Node 24 Avg. - Overheating")
node_26 <- node_26 |>
  summarise(across(-Year, mean)) |>
  pivot_longer(everything(), names_to = "Measure",
              values_to = "Node 26 Avg. - Healthy Recovery")

all_nodes <- node_4 |>
  left_join(node_24, by = "Measure") |>
  left_join(node_26, by = "Measure") |>
  filter(Measure != "row_id") |>
  gt() |>
  opt_row_striping()
```

Year	PPI_Change	Personal_Income	Compensation	Entrepreneurship	Transfers	GDP	Consumption
1963	-0.23724804	3.731312	4.087260	0.8372622	4.5584131	4.4	4.1
1964	0.18457487	5.627684	5.729020	3.4204277	2.6284659	5.8	6.0
1994	1.28818409	2.407677	2.267551	3.9193375	1.8741264	4.0	3.9
1995	3.51438036	2.837265	1.912013	2.4681910	3.5996948	2.7	2.9
1996	2.31747076	3.020134	2.187463	9.3354433	2.1543086	3.8	3.5
1997	-0.06528694	3.626595	4.088269	4.8170322	0.4135061	4.4	3.8
1999	0.83364383	2.845324	4.163589	6.2261586	2.0363770	4.8	5.4
2003	5.20715602	1.289027	1.227378	0.8257153	2.4423212	2.8	3.2
2004	6.00601825	2.847558	2.966330	4.3995092	2.7135075	3.8	3.8
2005	7.05710812	1.848092	1.719347	-1.6373993	3.1929878	3.5	3.5
2006	4.56388374	3.751035	2.516709	3.8985395	3.0340329	2.8	2.9
2014	0.93380498	3.356754	2.989498	-0.2632255	3.1185974	2.5	2.8
2015	-7.52075668	4.438330	4.620931	-1.7580072	5.3847529	2.9	3.4
2016	-2.69651355	1.380372	1.453018	-1.1487527	2.0952864	1.8	2.5
2017	4.30689625	2.654343	2.386592	3.6106069	0.6842140	2.5	2.6
2018	4.25703791	2.653720	2.577156	2.1531535	1.7263802	3.0	2.7
2019	-1.04947557	2.856578	2.570816	2.1656075	3.7838706	2.6	2.1
2024	-0.45876015	2.540428	2.700120	1.1835512	3.8735863	2.8	2.9

Table 4: Component C1, Node 4 Members

Year	PPI_Change	Personal_Income	Compensation	Entrepreneurship	Transfers	GDP	Consumption
1986	-2.93487742	3.819274	4.355013	4.278550864	4.04803495	3.5	4.1
1987	2.60309480	2.102181	3.338292	7.542323766	0.09155415	3.5	3.4
1988	3.94171012	4.086906	3.942467	8.762714220	2.05634064	4.2	4.2
1989	4.83707704	2.985721	1.617455	0.002884136	4.30784088	3.7	2.9
1994	1.28818409	2.407677	2.267551	3.919337479	1.87412645	4.0	3.9
1995	3.51438036	2.837265	1.912013	2.468190951	3.59969483	2.7	2.9
1996	2.31747076	3.020134	2.187463	9.335443284	2.15430862	3.8	3.5
1997	-0.06528694	3.626595	4.088269	4.817032211	0.41350613	4.4	3.8
2004	6.00601825	2.847558	2.966330	4.399509244	2.71350749	3.8	3.8
2006	4.56388374	3.751035	2.516709	3.898539532	3.03403293	2.8	2.9

Table 5: Component C1, Node 24 Members

Year	PPI_Change	Personal_Income	Compensation	Entrepreneurship	Transfers	GDP	Consumption
1992	0.5633416	3.2409853	3.122708	9.2134780	8.298746	3.5	3.7
1993	1.4542117	1.3265049	1.056303	3.7969739	2.909281	2.7	3.5
1994	1.2881841	2.4076768	2.267551	3.9193375	1.874126	4.0	3.9
1995	3.5143804	2.8372652	1.912013	2.4681910	3.599695	2.7	2.9
1996	2.3174708	3.0201336	2.187463	9.3354433	2.154309	3.8	3.5
2003	5.2071560	1.2890269	1.227378	0.8257153	2.442321	2.8	3.2
2004	6.0060183	2.8475578	2.966330	4.3995092	2.713507	3.8	3.8
2012	0.5497585	2.4164836	2.011891	3.6123112	-1.868500	2.3	1.4
2013	0.6164234	-0.3764019	1.621224	2.4534348	1.107013	2.1	1.7
2014	0.9338050	3.3567538	2.989498	-0.2632255	3.118597	2.5	2.8

Table 6: Component C1, Node 26 Members

Old Paper

Discussion

This paper serves as a proof-of-concept of the usefulness of TDABM as a methodology for exploratory data analysis. By no means is this a totally comprehensive method to gain insight into data, but instead TDA BallMapper proves itself to be useful tool wehen employed in addition to other traditional analyses.

In our section we briefly went over two ways to interpret TDABM graphs. As seen in many other, longer papers written by Dlotko and colleagues, there is a lot of room to expound our TDABM graph here,though that is beyond the scope of this paper. In general, though, we can see that TDBM can give insight into our data that we might not see otherwise. Seeing that there is data that is not like the others gives an indication that there is something to investigate. Consider our comparison of the 08' - 09' financial crisis and the 2020 COVID Pandemic: We see learned that the years following the financial crisis were similar economically. Another way to view this is, clustering revels that the financial crisis was a singular problem that spurred the collapse of the financial system. Conversely, the COVID pandemic was a singular issue but exposed many different weaknesses in our current world.

For the future development of TDABM, and TDA in general, there are many possibilities of future research and one glaring downside. The downside to TDA is that it is a high barrier to entry to understand the methodologies inner workings. Those with a technical background will have an easier time, but nonetheless a higher barrier than methods such as linear regressions. Subject wise, TDA has an

advantage in some of the social sciences because it can be somewhat of a bridge between qualitative and quantitative analyses. TDABM has been used to analyze the Brexit Vote data, seeking to understand quantitative and qualitative motivations behind its outcome[3]. TDA also has the advantages of dimension reduction which could be useful to the life-sciences. Fields such as genetics, chemistry, and biology have many fields of data, and being able to consolidate it into an understandable form could benefit the body of knowledge greatly.

REFERENCES

1.   Madukpe VN, Ugoala BC, Zulkepli NFS. A comprehensive review of the Mapper algorithm, a topological data analysis technique, and its applications across various fields (2007–2025). arXiv preprint arXiv:250409042 [Internet]. 2025; Available from: https://arxiv.org/abs/2504.09042

2.   Chazal F, Michel B. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. Front Artif Intell. 2021;4:667963.

3.   Dłotko P. Ball mapper: A shape summary for topological data analysis. arXiv preprint arXiv:190107410 [Internet]. 2019; Available from: https://arxiv.org/abs/1901.07410

4.   Bureau of Economic Analysis. Interactive data application [Internet]. n.d. Available from: https://www.bea.gov/itable

5.   Federal Reserve Bank of St. Louis. Federal reserve economic data (FRED) [Internet]. n.d. Available from: https://fred.stlouisfed.org/

6.   Batch A, Chen J, Kampas W. Bea.r: Bureau of economic analysis API [Internet]. 2018. Available from: https://CRAN.R-project.org/package=bea.R

7.   Boysel S, Vaughan D. Fredr: An r client for the 'FRED' API [Internet]. 2021. Available from: https://CRAN.R-project.org/package=fredr

8.   Hughes-Cromwick E, Coronado J. The value of US government data to US business decisions. J Econ Perspect. 2019;33(1):131–46.

9.   Federal Reserve Bank of Dallas. Deflating nominal values to real values [Internet]. n.d. Available from: https://www.dallasfed.org/research/basics/nominal

10.   Hamilton JD, Chinn M. Use of logarithms in economics [Internet]. 2014. Available from: https://econbrowser.com/archives/2014/02/use-of-logarithms-in-economics

11.   Bureau of Labor Statistics. Producer price indexes: concepts [Internet]. 2025. Available from: https://www.bls.gov/opub/hom/ppi/concepts.htm

12.   Dłotko P, Rudkin S, Qiu W. Topologically mapping the macroeconomy. arXiv preprint arXiv:191110476 [Internet]. 2019; Available from: https://arxiv.org/abs/1911.10476

13.   Kolesnikova NA, Liu Y. Jobless recoveries: Causes and consequences. The Regional Economist [Internet]. 2011;18–9. Available from: https://www.stlouisfed.org/publications/regional-economist/april-2011/jobless-recoveries-causes-and-consequences

14.   McClung S. Fisher effect [Internet]. 2024. Available from: https://inomics.com/terms/fisher-effect-1545486

15.   Michaelides PG. The fisher equation. In: 21 equations that shaped the world economy: Understanding the theory behind the equations [Internet]. Springer; 2024. p. 113–20. Available from: https://link.springer.com/chapter/10.1007/978-3-031-76140-9_9

16.   Bryan M. The great inflation [Internet]. 2013. Available from: https://www.federalreservehistory.org/essays/great-inflation

17.   Federal Reserve Bank of St. Louis. What are "long and variable lags" in monetary policy? [Internet]. 2023. Available from: https://www.stlouisfed.org/on-the-economy/2023/oct/what-are-long-variable-lags-monetary-policy

18.   Corbett M. Oil shock of 1973–74 [Internet]. 2013. Available from: https://www.federalreservehistory.org/essays/oil-shock-of-1973-74

Data Source

- Bureau of Economic Analysis

  - Real Exports and Imports
  - New Foreign Direct Investment in the United States
  - Real GDP
  - Personal Income and Its Disposition

- Federal Reserve Bank

  - Selected Interests Rates

- Bureau of Labor Statistics

  - CPI
  - Employment Cost Index
  - PPI
  - Housing Starts
  - Unemployment Rate