Spectral Clustering of Macroeconomic Indicators: Insights from Inflation and Unemployment

Vred Rudnick

September 26, 2025

1 Motivation

Understanding the behavior of macroeconomic variables is crucial for central banks to make informed decisions on monetary policy. In this work, we use spectral clustering in an attempt to find group countries based on their similarity in terms of price stability and labor markets. In the following sections, we provide background on spectral clustering, discuss the methodology of applying a spectral clustering algorithm to a cross-country data sample, and interpret the results.

2 Theory and Methodology

2.1 Spectral Clustering

Spectral clustering is a graph-based clustering technique. Its key strength is the ability to detect non-convex clusters, which distinguishes it from other popular approaches such as K-means. Suppose that we are given a data sample $x_1, ..., x_n \in \mathbb{R}^d$, where $n \in \mathbb{N}$ defines the sample size and $d \in \mathbb{N}$ the number of features. Our goal is to identify a sensible partition of the sample into k subsets for a given $k < n$. To use spectral clustering for this purpose, we view our sample as an undirected, weighted graph, in which the data points $x_1, ..., x_n$ define the vertices, and the matrix

$$W = (w_{ij})_{i,j=1}^n \in \mathbb{R}^{n \times n}$$

denotes the similarity matrix. Here, w_{ij} is non-negative and denotes the similarity between the vertices x_i and x_j. Note that W is symmetric because our graph is undirected. The output of our spectral clustering algorithm is derived from the normalized graph Laplacian matrix

$$L' = I - D^{-1/2} W D^{-1/2}$$

where D denotes the diagonal degree matrix $D = \text{diag}[d_1, ..., d_n]$ with $d_i := \sum_j w_{ij}$. In the literature, the matrix L' is often referred to as the *normalized symmetric Laplacian* to distinguish it from the *unnormalized Laplacian* $L = D - W$. A simple proof shows that

$$L' = D^{-1/2} W D^{-1/2}$$

holds. Additionally, $u \in \mathbb{R}^n$ is an eigenvector of L' to the eigenvalue λ if and only if it solves the following generalized eigenproblem (see von Luxburg (2007) for proof)

$$Lu = \lambda Du \tag{2.1}$$

Spectral clustering assigns each data point to one of the k clusters based on the eigenvectors to the k smallest eigenvectors of L'. In particular, the algorithm works as follows:

ALGORITHM 2.1 (Normalized Spectral Clustering for $k \geq 3$ (Shi and Malik 2000))**.** *Input: data sample $V = x_1, ...x_n$, the undirected graph $G = (V, E)$*

(1) Compute the similarity matrix $W \in \mathbb{R}^{n\times n}$ based on a similarity function, the degree matrix D and the unnormalized Laplacian matrix $L = D - W$.

(2) Solve the generalized eigenproblem (2.1) for the eigenvectors $v_1, ..., v_k$ to the smallest eigenvalues. Arrange them in the matrix $T \in \mathbb{R}^{n\times k}$.

(3) Group the rows $\psi_i \in \mathbb{R}^k$ of T using the k-means algorithm and assign the i-th vertex according to the cluster membership of ψ_i.

Output: a partition $A_1, ..., A_k$ of the vertex set V.

This technique is designed to find a partition that, while being as balanced as possible, preserves most of the similarity inside the clusters while cross-cluster similarity is low. Mathematically, this is achieved by finding the partition that minimizes the normalized cut of the data graph. This minimization, under certain relaxations, is equivalent to finding the eigenvectors to the k smallest eigenvalues of L' (for detail, see von Luxburg (2007, Section 5)).

When implementing spectral clustering, two key challenges consistently present themselves. The first challenge is to determine a similarity function from which to derive W. For our purposes, we use a Gaussian similarity kernel. Since $d \leq 2$ in all of our examples, we introduce the Gaussian kernel for one-dimensional data

$$K : \begin{cases} \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R} \\ (x, y) \longmapsto \exp\left(-\frac{\|x-y\|^2}{2h^2}\right) \end{cases} \tag{2.2}$$

where $h > 0$ denotes the kernel width. We select h according to Scott's rule for bandwidth selection

$$h = 1.06\sigma n^{-1/5} \tag{2.3}$$

In the two-dimensional case, the Gaussian kernel function is given by

$$K : \begin{cases} \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x, y) \longmapsto \exp\left(-\frac{1}{2}(x - y)^T H (x - y)\right) \end{cases} \tag{2.4}$$

Here, $H \in \mathbb{R}^{2\times 2}$ denotes the width matrix. A two-dimensional version of Scott's rule gives us the diagonal matrix

$$H = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix} \tag{2.5}$$

where h_1 and h_2 are given by $\sqrt{h_i} = n^{-1/6}\sigma_i$, where σ_1 and σ_2 are the standard deviations of the data's features. Note that Scott's rule has originally been developed for kernel density estimation (KDE). However, we still employ the method because it has proven to work reliably for our purposes. For more detail on bandwidth selection for KDE, see Silverman (1986). The second challenge is choosing the number of clusters. Absent prior knowledge about the underlying distribution, there exist a number of approaches, some of which are listed in von Luxburg (2007, Section 8.3). In our examples here, we use an eigengap heuristic, which suggests that clustering works more reliably if $\lambda_k - \lambda_{k-1}$, i.e. the gap between the k-th smallest eigenvalue λ_k and the next smallest eigenvalue λ_{k-1}, is large (see Rudnick (2025, Section 3.7) for detail). This approach is grounded in Davis-Kahan theory (see von Luxburg (2007, Section 7)).

It can be shown that, under certain assumptions, spectral clustering recognizes underlying data patterns reliably, and "converges" to the actual structure of the data. For details on the proof and a definition of convergence in this context, see von Luxburg, Belkin, and Bousquet (2008) and Rudnick (2025).

2.2 Data

Our sample consists of a number $n = 70$ of two-dimensional feature vectors, where each vector represents one of the countries. The data was obtained from the World Bank (2025). To ensure we select a diverse sample of countries, we chose them from across the Human Development Index (HDI) ranking (see United Nations Development Programme (2025, p. 274-278)). The features are the country's CPI YoY headline inflation rate and the country's unemployment rate (ILO-modeled) for a particular calendar year. It is important to note that central banks rely on a variety of indicators to inform monetary policy decisions, such as core inflation, loan cost for private lenders, GDP growth projections, and others (ECB 2025; Federal Reserve 2025). We choose the two metrics at hand mainly because they are well-known and widely available, but we may include additional metrics in this work, to better reflect the full scope of indicators that drive central bank action. Below, we provide full list of the analyzed countries and a scatterplot of the data for the year 2023(**??**).

List of the included countries: Malawi, Mali, Malta, Netherlands, Gambia, United Kingdom, United States, United Arab Emirates, Afghanistan, Austria, Australia, Burkina Faso, Burundi, Benin, Denmark, Ethiopia, Djibouti, Finland, France, Germany, Ireland, Japan, Lesotho, Luxembourg, Liberia, Niger, Norway, Pakistan, Slovenia, Singapore, Spain, Guinea, Guinea-Bissau, Hong Kong, Iceland, Switzerland, Sweden, Israel, South Korea, Mozambique, Chad, Tanzania, Sierra Leone, Madagascar, Central African Republic, Haiti, Senegal, Belgium, Canada, South Sudan, Czech Republic, Nigeria, Mauritania, Togo, Italy, Cyprus, Papua New

Guinea, Rwanda, Uganda, Ivory Coast, Greece, Poland, Estonia, Saudi Arabia, Bahrain, Lithuania, Solomon Islands, Cameroon, Zambia, Cambodia

2.3 Methodology

For our two-dimensional data, we employ Scott's rule in the sense of (2.5) with $n = 70$ while computing σ_1 and σ_2 empirically from the sample. We then use the Gaussian similarity kernel defined in (2.4) to calculate the similarity matrix W, and subsequently the normalized Laplacian matrix. For the year 2023, our eigengap heuristic suggests choosing $k = 9$, since $\lambda_9 - \lambda_8$ is the largest eigengap by far.

3 Results

We observe in Figure that a significant chunk of the countries belongs to a single cluster, with an inflation rate below 15 percent, and an ILO unemployment estimate below 10 percent.

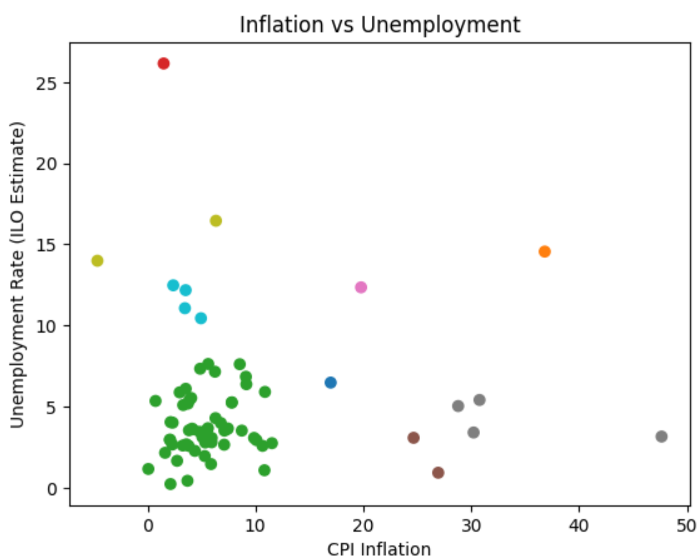

Figure 3.1: Scatterplot of the clustering result for $k = 9$ and the year 2023

This dense region in the plot contains a diverse group of countries with low to moderate inflation and unemployment, ranging from countries with low inflation and unemployment, such as Bahrain, low inflation and moderate unemployment, e.g. Burkina Faso, to those with moderate inflation and low to modest unemployment (e.g. Germany, Chad).

4 Conclusion and Outlook

We conclude from the analysis that neither region nor HDI seem to determine the clusters we observe. Therefore, it is necessary to conduct analysis of additional indicators to find out what the determining factor of the clusters is. Such analysis will be provided in the future, as this work is updated regularly. Additionally, the behavior of our two-dimensional metric over recent decades is intriguing, as we see notable shifts throughout the change of global economic environments.

References

ECB. 2025. *Monetary policy statement (with Q&A), 11 September 2025.* https://www.ecb.europa.eu/press/press_conference/monetary-policy-statement/2025/html/ecb.is250911~a13675b834.en.html. Accessed: 2025-09-25.

Federal Reserve, Open Market Committee. 2025. *Summary of Economic Projections, September 17, 2025.* https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20250917.pdf. Accessed: 2025-09-25.

Rudnick, Vred. 2025. "Mathematical Theory of Spectral Clustering." Master's thesis, Ruhr University Bochum, Germany.

Shi, Jianbo, and Jitendra Malik. 2000. "Normalized Cuts and Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8): 888–905. https://doi.org/10.1109/34.868688.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis.* Vol. 26. Monographs on Statistics and Applied Probability. Printed in 1998. Boca Ranton, FL, USA: Chapman & Hall. ISBN: 978-0-412-24620-3.

United Nations Development Programme. 2025. *Human Devlopment Report 2025: A Matter of Choice: People and Possibilities in the Age of AI.* Report. English version. Accessed on July 18, 2025. https://hdr.undp.org/content/human-development-report-2025.

von Luxburg, Ulrike. 2007. "A Tutorial on Spectral Clustering." *Statistics and Computing* 17 (4): 395–416. https://link.springer.com/article/10.1007/s11222-007-9033-z.

von Luxburg, Ulrike, Mikhail Belkin, and Olivier Bousquet. 2008. "Consistency of Spectral Clustering." *Annals of Statistics* 36 (2): 555–586. https://projecteuclid.org/journals/annals-of-statistics/volume-36/issue-2/Consistency-of-spectral-clustering/10.1214/009053607000000640.full.

World Bank. 2025. *Agriculture, Forestry, and Fishing, Value Added (% of GDP).* World Bank Development Indicators. Data for the year 2023. Downloaded on July 16, 2025. https://databank.worldbank.org/source/world-development-indicators#.