Semantic Regimes of Multiplication: An Exploration Using Entropy and UMAP

CRL-J

March 29, 2025

Abstract

This paper explores the structure of multiplication tables by applying Shannon entropy and UMAP for dimensionality reduction. We show that clusters of rows exhibit distinct, interpretable semantic regimes based on arithmetic properties such as the number of distinct prime divisors. This unsupervised analysis reveals deep insights into the multiplicative behavior of integers, providing a topological lens through which we can view arithmetic operations. We also use HDBSCAN to identify natural groupings of multiplication rows, associating these groups with prime, composite, and transitional regimes.

1 Introduction

The multiplication table, a foundational concept in number theory, offers a vast array of behaviors when analyzed through various mathematical lenses. Traditional approaches focus on the arithmetic structure of these tables, such as the frequency of products or divisibility. However, this work introduces a topological perspective, where we analyze the entropy and geometric structure of multiplication tables using unsupervised learning techniques such as UMAP (Uniform Manifold Approximation and Projection) and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). These techniques allow us to discover underlying patterns and semantic groupings of rows in a multiplication table based on their product distributions and divisibility properties.

2 Methodology

2.1 Data Representation

We represent an $n \times n$ multiplication table as a matrix M, where the entry at position (i, j) is the product $i \times j$. Each row of this table corresponds to the products of a particular integer i with every integer from 1 to n.

2.2 Shannon Entropy

Shannon entropy is a measure of uncertainty or disorder in a probability distribution. In this case, we use entropy to quantify the multiplicative complexity within each row of the multiplication table. A higher entropy indicates that the row has a broader range of products with more variation, while a lower entropy suggests the row has repeated or predictable values.

The Shannon entropy H_i for row i of the multiplication table is calculated as:

$$H_i = -\sum_{k=1}^{m} p_k \log_2 p_k$$

where: - p_k is the probability of the k-th product in the row, - m is the number of distinct products in that row.

This entropy measurement gives us a sense of the unpredictability in the row. For example:

- Low entropy rows correspond to numbers with repetitive patterns (such as rows dominated by multiples of prime numbers).
- High entropy rows correspond to numbers with more varied product structures, indicative of higher multiplicative complexity.

In the context of multiplication tables, entropy helps identify rows with more complex or erratic multiplicative patterns. We visualize these entropy values as a color gradient in the UMAP plots to capture how entropy varies across different regions of the table.

2.3 UMAP for Dimensionality Reduction

UMAP is employed to reduce the dimensionality of the multiplication table. Each row of the table is treated as a point in high-dimensional space, where each dimension corresponds to one of the multiplication results for that row. UMAP projects this high-dimensional data into a 2D space, maintaining the local and global structures of the data. The resulting 2D projection provides a way to visualize how similar or dissimilar the rows are based on their multiplication properties.

2.4 HDBSCAN Clustering

HDBSCAN is applied to the UMAP projection to identify clusters in the multiplication table. These clusters correspond to regimes of similar multiplicative behavior, such as rows with similar prime divisor structures or product distributions. The cluster labels are mapped back to the UMAP coordinates for visualization.

3 Results

3.1 UMAP Visualization

The UMAP projection reveals a smooth, curved manifold where rows of the multiplication table with similar behavior cluster together. These clusters are visualized in Figure 1, showing the distinct regions of arithmetic behavior.

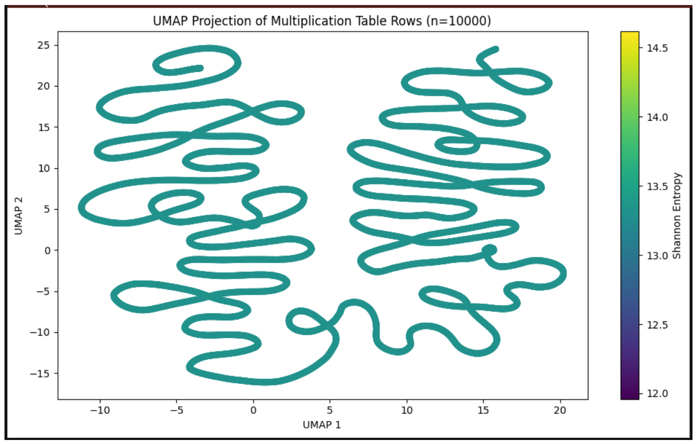

Figure 1: UMAP projection of multiplication table rows (n=10,000), colored by Shannon entropy. Rows with similar multiplicative behavior are grouped together in the 2D space. Higher entropy indicates more variation in products.

3.2 Cluster Interpretation

Upon examining the clusters, we identified five distinct semantic regimes, each exhibiting different arithmetic behaviors. These behaviors are characterized by a combination of the number of distinct prime divisors and Shannon entropy, which indicates the complexity of the multiplicative patterns. Here's how these clusters can be interpreted:

- **Prime-Sparse Base**: Rows in this cluster correspond to numbers with fewer distinct prime divisors (lower $\omega(n)$) and low entropy. These rows exhibit predictable, simple multiplicative patterns, often dominated by prime numbers or products of small primes. The low entropy suggests that the products in these rows are relatively easy to predict.

- **Transitional Wedge**: This cluster contains rows that transition from simple multiplicative patterns to more complex ones. The numbers in this group have a moderate number of prime divisors, and their entropy values are slightly higher, reflecting increased complexity. These rows represent a blend of prime and composite behaviors, showing the point at which multiplicative structures become less predictable.

- **Composite Crest**: Rows in this cluster have a higher number of prime divisors, indicating a greater degree of multiplicative complexity. These rows also exhibit higher entropy, suggesting that their product structures are more irregular and less predictable. The increased entropy reflects a shift toward numbers with many divisors, exhibiting more complex multiplicative behavior.
- **Multiplicative Mesh**: This cluster corresponds to the densest region in the multiplication table, consisting of rows with high prime divisor counts and significantly higher entropy. These rows represent highly composite numbers and exhibit substantial variation in their products. The high entropy suggests that these numbers are involved in complex multiplicative interactions, leading to a broader range of products.
- **Entropy Plateau**: Rows in this cluster are located at the highest indices of the multiplication table, dominated by highly composite numbers. These rows exhibit the highest entropy values, indicating that their products are the most unpredictable. The number of divisors is also high, but the dominant feature of this group is the saturated entropy, reflecting a level of multiplicative complexity that is difficult to predict or reduce further.

These clusters represent different semantic regimes of multiplication, where entropy is a key factor in determining the complexity and predictability of each row's products. Lower entropy indicates simpler, more predictable multiplication patterns, while higher entropy reflects rows with complex, less predictable multiplicative behavior. The relationship between the number of prime divisors and entropy further reveals how the structure of numbers influences the degree of complexity in their multiplication patterns.

3.3 Prime Divisors Distribution

Figure 2 shows the distribution of distinct prime divisors ($\omega(n)$) within each cluster. The results demonstrate how prime divisor complexity increases smoothly across clusters, with each cluster showing a different range of divisor counts.

3.4 Entropy vs. Prime Divisors

To explore the relationship between entropy and prime divisors, we plotted these two metrics for each cluster. As shown in Figure 3, we observe that entropy generally increases with the number of prime divisors, reflecting a shift from simpler, predictable rows to more complex, composite rows. This suggests that rows with more prime divisors tend to exhibit higher multiplicative variation.

3.5 HDBSCAN Clustering in UMAP Space

The results of HDBSCAN clustering in UMAP space are visualized in Figure 4. We observe that the clustering naturally separates the rows into distinct groups

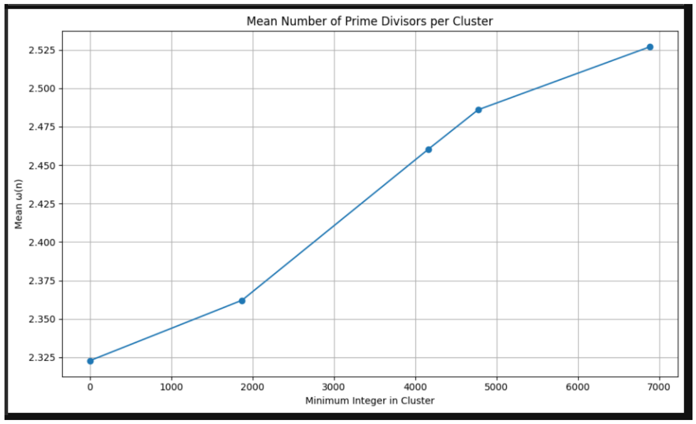

Figure 2: Mean number of prime divisors ($\omega(n)$) per cluster. As the index increases, the multiplicative complexity (as measured by the number of prime divisors) rises.

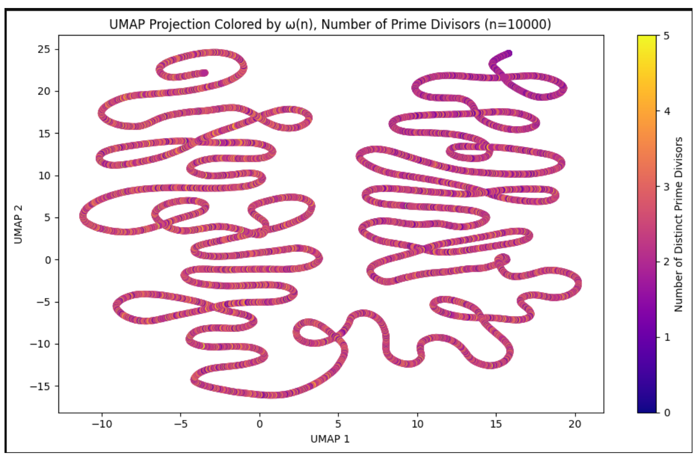

Figure 3: Comparison of entropy and number of prime divisors ($\omega(n)$) across clusters. Higher entropy is generally associated with rows having more prime divisors.

based on their multiplicative behavior, with some clusters representing rows with higher divisibility (and hence higher entropy), while others correspond to simpler rows with fewer divisors.

3.6 Semantic Regimes of Multiplication

Finally, we identified and labeled five semantic regimes of multiplication. These regimes describe the different types of arithmetic behavior present in the multi-

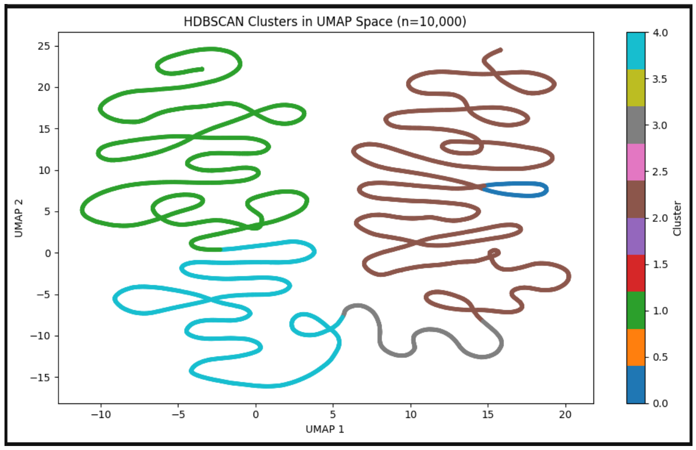

Figure 4: HDBSCAN clustering of UMAP-projected multiplication table rows (n=10,000). Each color represents a distinct cluster, with rows exhibiting different multiplicative properties.

plication table, as observed through their entropy and prime divisor characteristics. These regimes are mapped onto the UMAP projection in Figure 5.

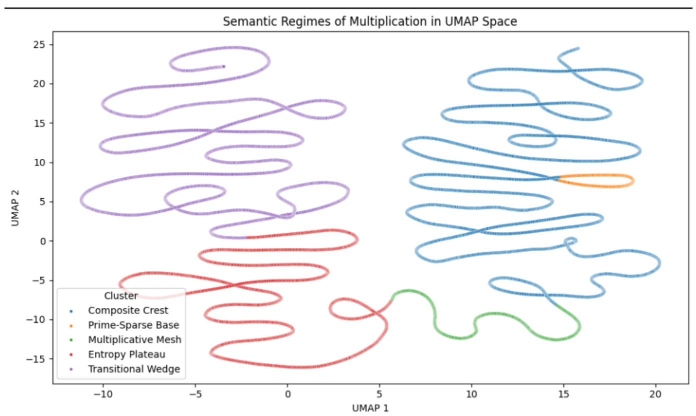

Figure 5: Semantic regimes of multiplication in UMAP space, based on HDBSCAN clustering. The distinct clusters are color-coded to reflect the different arithmetic behaviors, ranging from prime-heavy to highly composite rows.

4 Conclusion

In this study, we employed entropy and UMAP to reveal the hidden topological structure of multiplication tables. By applying HDBSCAN, we identified five semantic regimes that correspond to different types of multiplicative behavior, ranging from prime-heavy rows to highly composite ones. This work demonstrates how topological data analysis can uncover rich patterns in arithmetic, opening the door for further exploration of number-theoretic properties through data science.