A Motif-Based Network Analysis of Transaction Data from Ekko : To Understand Consumer Behavior and their Carbon Emissions

Raga Sreehitha Paritala

Master of Science
School of Informatics
University of Edinburgh
2024

Abstract

This study investigates the relationship between consumer transaction patterns and carbon emissions using advanced network analysis techniques. Analyzing a unique card transaction dataset from Ekko that tracks CO2 emissions for individual transactions, we construct and compare transaction networks for high and low carbon-emitting customers. Our research aims to uncover spending patterns related to carbon emissions, differentiate between customer groups, and identify strategies to promote sustainable spending habits. We employ a novel methodology combining network science with carbon footprint assessment, focusing on 3-node motifs in transaction networks.

These motifs, representing recurring transaction patterns, provide crucial insights into the microstructures of consumer behavior and their environmental impact. Our analysis reveals significant structural differences between high and low emission customer networks, with high emitters exhibiting larger, more complex networks and overrepresentation of multi-step transaction sequences. Key findings include a non-linear relationship between motif complexity and emissions, and an association between low-emission behaviors and more diverse transaction patterns. This research contributes to sustainable finance by providing a framework for understanding and influencing the environmental impact of consumer spending through detailed transaction pattern analysis.

Keywords: Consumer behavior, Carbon emissions, Network analysis, Transaction patterns, Network motifs, Sustainable finance

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 925065
Date when approval was obtained: 2024-06-24

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Raga Sreehitha Paritala*)

Acknowledgements

I would like to thank my supervisor, Valerio Restocchi, for his insightful suggestions, feedback, and guidance throughout this project. His guidance at various stages has been crucial in ensuring that the project remained on the right track.

I am also grateful to Ogy Simeonov for his mentorship, which provided constant encouragement and invaluable support throughout the project.A special thank you goes to Ekko for providing the unique dataset that is integral to the success of this research.

Additionally, I want to express my sincere appreciation to all my family and friends for their encouragement and support during the completion of this project and throughout the past academic year.

Table of Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Networks and Motifs 2

1.3 Research Aims, Objectives and Questions 3

1.3.1 Research Aims 3

1.3.2 Research Objectives 3

1.3.3 Research Questions 4

1.4 Significance of the study 5

1.5 Dissertation Outline 6

2 Literature Review **7**

2.1 Analyzing Card Transaction Data for Consumer Behavior 7

2.2 Network Analysis of Card Transactions 9

2.3 Network Motifs in Consumer Behavior Analysis 10

2.4 Motif Detection Algorithms 12

3 Methodology **14**

3.1 Data Collection and Preprocessing 14

3.1.1 Dataset Description 14

3.1.2 Data Filtering and Preparation 15

3.1.3 Customer Classification for Comparative Analysis 17

3.2 Network Construction 18

3.2.1 Graph Representation 18

3.2.2 Graph Construction process 18

3.2.3 Mathematical formulation 19

3.3 Motif Detection and Analysis 19

3.3.1 Motif Detection using PymFinder 19

3.4 Motif level Carbon Emission 21

4 Results **23**

4.1 Data Description 23

4.2 Network Analysis 23

4.2.1 Network Statistics 24

4.3 Motif Analysis 26

4.3.1 Motif Detection 26

4.3.2 Motif Frequency Distribution 27

4.3.3 Statistical Significance of Motifs: 28

4.3.4 Weighted Motif Analysis 30

4.3.5 Motif Level Carbon Emission Analysis 31

4.3.6 Motif Composition and Node Role Analysis 33

4.4 Key Findings 36

5 Discussions and Conclusions **38**

5.1 Discussions 38

5.2 Conclusion 39

Bibliography **41**

Chapter 1

Introduction

1.1 Motivation

Climate change stands as one of the most pressing challenges of our time, with far-reaching implications for ecosystems, economies, and human societies. The Intergovernmental Panel on Climate Change (IPCC) has warned of severe consequences if global temperatures rise beyond 1.5°C above pre-industrial levels, emphasizing the urgent need for rapid and significant reductions in greenhouse gas emissions across all sectors of the economy [14].

While much of the focus on climate action has been directed at industrial and governmental levels, there is growing recognition of the significant role that individual consumer behavior plays in driving environmental outcomes [12]. A groundbreaking study in the Journal of Industrial Ecology revealed that household consumption accounts for 60% of global greenhouse gas emissions and between 50% and 80% of total land, material, and water use [38]. This underscores the critical importance of understanding and influencing consumer behavior as part of broader climate mitigation strategies.

The financial sector, as a facilitator of consumer transactions, holds a unique position in influencing these behaviors [26]. Traditional banking systems have long been criticized for their lack of transparency and accountability regarding the environmental impact of the transactions they process [27]. However, the emergence of fintech companies focused on sustainability, such as Ekko, is revolutionizing this landscape. These innovative firms are not only providing financial services but are also actively tracking and reporting the environmental impact of each transaction, creating a rich dataset that offers unprecedented insights into the relationship between consumer spending and carbon emissions [28][31].

This research leverages the unique dataset provided by Ekko, to delve deep into the patterns of consumer spending and their associated carbon footprints. By applying advanced network science techniques to this data, we aim to uncover the intricate relationships between transaction sequences, merchant categories, and carbon emissions. This approach allows us to move beyond surface-level analysis and explore the complex dynamics that drive sustainable and unsustainable consumer behaviors [4].

As we move into the next section on Networks and Motifs in Consumer Behavior Analysis, we will delve deeper into these analytical tools and explore how they can be applied to extract meaningful patterns from complex transaction data.

1.2 Networks and Motifs

Network analysis has emerged as a powerful tool for uncovering patterns in complex systems of interactions, including consumer behavior. In the field of network science, networks are mathematical structures that consist of nodes (or vertices) connected by edges (or links). These networks can represent a wide variety of systems, including social interactions, biological systems, and financial transactions. When studying networks, one key area of interest is identifying patterns or structures within them, particularly motifs.

The concept of network motifs, first introduced by Milo et al. [25], has found applications in various fields, from biology to social network analysis. Motifs are recurring, small subgraphs that appear more frequently in a network than would be expected by random chance. These motifs can reveal underlying dynamics and significant patterns within the network that help explain the behavior of the system [24].

In this research, network science techniques will be applied to analyze transaction data with a focus on carbon emissions linked to consumer spending. By mapping transaction sequences involving Merchant Category Codes (MCCs) into networks, where nodes represent MCCs and edges represent transitions between them, it becomes possible to visualize and analyze the flow of consumer behavior [10]. This approach allows for the identification of network motifs that are associated with high or low carbon emissions, shedding light on spending patterns that have significant environmental impacts.

By combining the power of network analysis with Ekko's rich dataset that tracks CO2 emissions for every transaction, this study represents a novel approach to understanding and influencing consumer behavior for environmental sustainability. The identification of network motifs in transaction patterns offers a new lens through which to view the relationship between spending habits and carbon emissions, potentially revealing insights that traditional analysis methods might miss [8]. This research stands at the intersection of network science and consumer behavior, offering a unique perspective on one of the most critical challenges of our time—understanding the environmental impact of consumer spending patterns.

Moreover, this approach has the potential to uncover actionable insights that can inform targeted interventions and policy decisions [31]. By identifying specific transaction sequences and patterns associated with higher or lower carbon emissions, we can develop more nuanced strategies for promoting sustainable consumption [34]. This granular level of analysis could lead to the development of personalized recommendations for consumers, innovative product offerings from financial institutions, and more effective policy instruments to incentivize sustainable spending habits.

1.3 Research Aims, Objectives and Questions

1.3.1 Research Aims

This research aims to analyze real-world transaction data from Ekko, that tracks CO2 emissions for every transaction using network science techniques to uncover spending patterns related to carbon emissions, differentiate between high and low CO2 emission customers, and identify potential strategies to promote more sustainable spending habits. By focusing on the sequence of Merchant Category Codes (MCCs) in customer purchases and their associated carbon emissions, we seek to provide a novel perspective on the relationship between consumer behavior and environmental impact.

1.3.2 Research Objectives

- **RO1:** To construct and analyze transaction networks representing customer spending patterns and their associated carbon emissions. This involves creating directed networks where nodes represent unique MCCs and edges represent transitions between MCCs in a customer's transaction sequence, with edge weights reflecting transition frequencies and carbon emissions.

- **RO2:** To identify and characterize network motifs that differentiate high and low carbon-emitting customers. This includes applying specialized motif detection algorithms to uncover recurring subgraphs within the transaction networks and analyzing their prevalence and characteristics across different customer segments.

- **RO3:** To analyze the carbon emissions of transaction motifs between high and low carbon-emitting customer groups. This involves developing a robust methodology for calculating carbon emissions associated with each identified 3-node motif and conducting comparative analyses between customer groups.
- **RO4:** To identify specific Merchant Category Code (MCC) sequences that contribute significantly to variations in carbon emissions between customer groups. This includes analyzing high-impact sequences to uncover patterns and behaviors that drive carbon emission disparities.

1.3.3 Research Questions

- **RQ1:** What are the characteristic network structures and motifs that distinguish high carbon-emitting customers from low carbon-emitting customers? This question aims to identify specific patterns in transaction networks that are indicative of high or low carbon emissions.
- **RQ2:** How do the frequencies and compositions of specific 3-node motifs compare between high and low carbon-emitting customer groups? This question focuses on understanding which particular transaction patterns are more prevalent in high-emission versus low-emission customer behaviors.
- **RQ3:** How do carbon emissions of the transaction motifs differ between high and low carbon-emitting customers? This question seeks to uncover whether there are significant differences in the carbon intensity of spending patterns between these two groups.
- **RQ4:** What specific sequences of Merchant Category Codes (MCCs) are most strongly associated with high carbon emissions, and how do these differ from sequences typical of low-emission customers? This question seeks to identify concrete transaction patterns that contribute significantly to higher carbon footprints.

1.4 Significance of the study

The research brings new knowledge and insight to the field by providing a quantitative, data-driven approach to understanding the link between granular transaction patterns and carbon emissions. This fills a significant gap in the literature, moving beyond qualitative studies of behavioral drivers to provide actionable insights based on real-world transaction data. The use of network motifs to analyze carbon emission patterns in consumer transactions is a novel approach that can reveal complex patterns not easily discernible through traditional analysis methods [24][8].

Specifically, the analysis of 3-node motifs within the MCC sequence network offers a unique perspective on the microstructures of consumer spending behavior and their environmental impact [15]. By calculating carbon emissions for these motifs and comparing them between high and low carbon-emitting customer groups, provides a nuanced understanding of how specific spending patterns contribute to overall carbon footprints. This level of granularity in analysis is unprecedented in the field of sustainable finance and consumer behavior studies.

The findings from this study have the potential to significantly impact the wider field of sustainable finance and environmental economics. By demonstrating the effectiveness of network analysis in uncovering patterns of sustainable and unsustainable spending, we provide a scalable framework for financial institutions to promote environmentally conscious consumption. The insights gained from analyzing the carbon emissions associated with specific transaction motifs and MCC sequences can inform targeted interventions and incentives to encourage more sustainable consumer choices.

Moreover, this research contributes to the growing field of green fintech by providing a methodological framework that can be adopted and adapted by other financial institutions seeking to understand and influence the environmental impact of their customers' spending. It demonstrates the potential of leveraging transaction data for sustainability purposes, opening up new avenues for research and innovation in sustainable finance.

The practical implications of this research are substantial. Financial institutions can use these insights to develop more targeted and effective sustainability initiatives, such as personalized recommendations for reducing carbon footprint or incentive programs that reward sustainable spending patterns [20]. Policymakers can leverage this information to design more effective regulations and incentives to promote sustainable consumption [35]. For consumers, this research could lead to the development of more informative and actionable tools for understanding and reducing their individual carbon footprints.

In conclusion, this study represents a significant step forward in our understanding of the relationship between consumer spending patterns and carbon emissions. By applying sophisticated network analysis techniques to a unique dataset, we have uncovered insights that have the potential to drive meaningful change in consumer behavior and contribute to global efforts in mitigating climate change.

1.5 Dissertation Outline

The subsequent chapters of this dissertation are structured as follows: Chapter 2 provides a comprehensive literature review, encompassing card transaction data analysis, network analysis in financial contexts, and the application of network motifs to consumer behavior studies. Chapter 3 details the methodological framework, including data preprocessing, network construction, motif detection using pymfinder, and our novel approach to motif-level carbon emission calculation.

Then, the Chapter 4 presents the empirical results, focusing on network characteristics and motif analysis of high and low carbon-emitting customer groups. Chapter 5 discusses the implications of our findings, addresses study limitations, and concludes with key insights and future research directions.

Chapter 2

Literature Review

This chapter provides a comprehensive review of the existing literature relevant to our study on analyzing card transaction data for consumer behavior and its environmental impact. It begins by exploring the broader field of transaction data analysis, discussing its importance in understanding consumer patterns and economic trends in Section 2.1. Section 2.2 narrows the focus to network analysis of card transactions, examining how graph theory and network science have been applied to financial data. Section 2.3, delves into the concept of network motifs, their significance in revealing underlying structures in complex networks, and their potential applications in consumer behavior analysis. Finally, Section 2.4 discusses various motif detection algorithms, evaluating their strengths and limitations in the context of transaction network analysis. This literature review sets the foundation for our methodology and highlights the gaps in current research that our study aims to address.

2.1 Analyzing Card Transaction Data for Consumer Behavior

This is a field of study leverages detailed financial records to gain insights into purchasing patterns, preferences, and economic trends. This topic is of significant importance in the modern digital economy, as it allows businesses, financial institutions, and researchers to understand consumer habits, predict future behaviors, and tailor products and services accordingly.

The importance of this research area has grown exponentially with the increasing prevalence of electronic payments and the vast amounts of data generated by card transactions. It provides valuable information for marketing strategies, fraud detection, economic forecasting, and personalized financial services. Moreover, it offers a real-time window into consumer spending habits, which can be crucial for businesses adapting to rapidly changing market conditions.

Numerous studies have been conducted in this field, employing various methodologies and analytical techniques. For instance, leveraging large-scale transaction data has enabled researchers to assess the carbon footprints associated with different consumption patterns, advancing our understanding of sustainability implications [34]. This research has yielded significant results, such as detailed spending patterns, consumer profiles, and improved targeting strategies.

A significant contribution to this field comes from Kooti et al. [18], who conducted a large-scale study of online shopping behavior using email data. They analyzed 121 million purchases made by 20.1 million shoppers, extracted from email confirmations sent by merchants. Their study provided insights into demographic factors affecting online shopping, temporal patterns in purchasing behavior, and the influence of social networks on consumer choices.

The results of work done in this area have been promising and diverse. Krumme et al. [19] found that consumers tend to follow regular patterns in their shopping behavior, which can be predicted with a high degree of accuracy using transaction data. This insight has significant implications for inventory management and targeted marketing. Recent studies have continued to expand our understanding of consumer behavior through transaction data analysis. For example, Di Clemente et al. [4] developed a framework for extracting mobility and consumption patterns from credit card data. Their research showed how these patterns could be used to create detailed profiles of consumer behavior, potentially aiding in urban planning and business strategy development.

Furthermore, [23] showed that analyzing transaction data could detect fraudulent activities with high accuracy by using temporal motifs. Dong et al. [6] utilized machine learning techniques on credit card transaction data to predict consumer loyalty, demonstrating the potential of this approach in customer relationship management.

In summary, the analysis of card transaction data for consumer behavior is a powerful tool for understanding and predicting economic activities at both micro and macro levels. It offers invaluable insights for businesses, policymakers, and researchers alike. This topic naturally links to network analysis of consumer transactions, as both fields deal with extracting meaningful patterns from financial data. While card transaction analysis focuses on individual behaviors and broad trends, network analysis examines the interconnections between consumers, businesses, and financial institutions. Together, these approaches provide a comprehensive view of the complex ecosystem of modern financial transactions, enabling more sophisticated strategies for economic growth, risk management, and consumer services.

2.2 Network Analysis of Card Transactions

Network analysis is a powerful tool for understanding complex systems of interactions, including consumer spending patterns. In the context of card transaction data, network analysis involves representing purchases as nodes (typically Merchant Category Codes or MCCs) and the transitions between purchases as edges. This approach allows researchers to uncover structural patterns and dynamics within consumer behavior that may not be apparent through traditional statistical methods.

The importance of this topic lies in its ability to reveal hidden patterns and relationships within large-scale transaction data. By analyzing the structure and properties of transaction networks, researchers can identify common spending sequences, influential MCCs, and patterns that differentiate various consumer groups and their spending patterns. This information can be valuable for understanding consumer behavior, predicting future purchases, and developing targeted interventions to influence spending habits.

Previous work in this area has primarily focused on applications in marketing and fraud detection. For example, Raeder and Chawla [29] used network analysis to model customer behavior in online retail, identifying patterns that could predict future purchases. In the financial sector, [32] applied network analysis techniques to detect fraudulent transactions by identifying anomalous patterns in transaction networks. Recent work in this area has expanded beyond marketing and fraud detection to include applications in sustainability and financial technology. [11] applied network analysis to examine the structure of global trade networks and their impact on carbon emissions. In the fintech sector, [22] used network-based approaches to analyze peer-to-peer lending behaviors and credit risk assessment.

Results from these studies have further established the effectiveness of network analysis in uncovering meaningful patterns in transaction data. For instance, Gianluca et. al. [11] found that certain network structures in global trade were associated with higher carbon emissions, highlighting the potential of network analysis in sustainability research. Li et al. [22] showed that network-based credit assessment models outperformed traditional methods in predicting loan defaults.

While previous research has expanded the use of network analysis to understand consumer behavior and its broader implications, there remains a significant opportunity to apply these techniques specifically to the environmental impact of individual transactions. Existing studies have explored various applications, including marketing and fraud detection, but have not addressed how network analysis can reveal the environmental consequences of spending patterns. This project aims to bridge this gap by using network analysis to examine the relationship between card transaction patterns and their associated carbon emissions.

By leveraging detailed transaction data, where emissions are tracked by Merchant Category Codes (MCCs), this project seeks to uncover new insights into how individual spending behaviors contribute to overall carbon footprints. Understanding these relationships can lead to more targeted strategies for promoting sustainable consumer practices. To achieve this, the project will focus on analyzing network motifs within the transaction data, which are specific patterns or structures that recur in the network of consumer transactions. The next section will delve into the concept of network motifs, exploring how these recurring patterns in transaction sequences can provide deeper insights into consumer behavior and its environmental impact.

2.3 Network Motifs in Consumer Behavior Analysis

Network motifs are recurring, statistically significant subgraphs or patterns within a larger network. In the context of consumer behavior analysis, motifs represent common sequences or combinations of transactions that occur more frequently than would be expected by chance. This topic is crucial for understanding the underlying structure of consumer spending patterns and identifying key behavioral indicators.

The importance of studying network motifs in consumer transactions lies in their potential to reveal fundamental building blocks of consumer behavior. By identifying and analyzing these recurring patterns, researchers can gain insights into the decision-making processes that drive spending habits, predict future behaviors, and understand the factors that contribute to specific outcomes, such as high carbon emissions or customer churn.

Milo et al. [25] introduced the concept of network motifs as simple building blocks of complex networks. They demonstrated that certain subgraphs appear in real-world networks significantly more often than in randomized networks, suggesting that these motifs may carry important functional roles in the system they represent. This work laid the foundation for applying motif analysis to various domains, including consumer behavior.

In the context of banking and financial networks, Fontes et al. [8] applied heterogeneous network motifs to analyze real-world transaction data for fraud detection. They proposed a novel randomization process specifically designed for banking datasets, which operates directly on tabular data to overcome limitations of traditional network randomization methods in this context. Their work demonstrated that heterogeneous network motifs can provide more informative and interpretable results than traditional homogeneous motifs when applied to banking transaction data.

Fontes et al. [8] explored two types of graph representations for banking data: entity graphs and transaction graphs. In entity graphs, nodes represent entities such as clients, merchants, and cards, while edges represent transactions between these entities. In transaction graphs, nodes represent individual transactions, and edges connect transactions that share entities within a specific time window. This dual representation allows for a comprehensive analysis of both entity relationships and transaction patterns.

The authors found that certain motifs were significantly over represented in the real transaction networks compared to randomized networks. For example, in entity graphs, motifs involving multiple fraudulent transactions by the same client or card were prominent, reflecting the tendency of fraud to cluster around specific individuals or accounts. In transaction graphs, motifs representing repeated transactions between the same client and merchant within a short time frame were common, potentially indicating normal purchasing behavior or, in some cases, fraudulent activity.

Importantly, Fontes et al. [25] also identified anti-motifs - subgraphs that appear significantly less frequently than expected. These anti-motifs provided valuable insights into unusual or rare transaction patterns that might be indicative of fraud or other anomalous behavior.

The application of network motifs to consumer behavior analysis extends beyond fraud detection. In e-commerce and online retail, researchers have used motif analysis to improve product recommendation systems and predict customer churn [33][13]. For instance, Srivastava [33] applied network motifs to analyze user-item interaction patterns on online shopping platforms, leading to improved accuracy in product recommendations.

The results from these studies underscore the importance of network motifs as critical indicators of broader behavioral patterns. For example, in customer churn analysis, certain motifs, such as frequent transitions between specific categories, were linked to a higher likelihood of churn [13][39]. In e-commerce studies, specific motif structures were associated with higher customer engagement and increased likelihood of purchase [33].

By applying similar methodologies to transaction data with a focus on carbon emissions, this research aims to identify motifs within transaction networks that are associated with high carbon emissions, bridging the gap between consumer behavior and environmental sustainability. This approach has the potential to reveal specific transaction sequences or combinations that contribute disproportionately to an individual's carbon footprint, providing valuable insights for developing targeted interventions to promote more sustainable consumption patterns.

The study of network motifs in consumer behavior analysis connects closely with the previously discussed topics of Network analysis of Card transactions . It provides a more granular level of analysis that can complement broader network measures and help explain why certain spending patterns result in higher or lower carbon emissions. By identifying specific motifs associated with high-emission behaviors, researchers can develop more nuanced strategies for promoting sustainable consumption, tailored to the actual spending patterns of consumers.

2.4 Motif Detection Algorithms

Network motif discovery has been a significant area of research in complex network analysis. Various algorithms have been developed to identify these recurring, statistically significant subgraphs. In the context of analyzing small transaction networks, some of the notable algorithms include:

1. SUBDUE [5]: SUBDUE uses the minimum description length (MDL) principle to iteratively find motifs. For small networks, SUBDUE might be computationally feasible, but its greedy search strategy could potentially miss important motifs. In the context of transaction networks, where capturing all relevant patterns is crucial, this limitation could be significant even in smaller datasets.

2. GREW [21]: GREW is designed for finding exact isomorphic and vertex-disjoint motifs. While it can handle smaller networks efficiently, its approach of doubling subgraph size in each iteration might skip over important intermediate-sized

motifs. For small transaction networks, this could result in missing crucial patterns that exist between the size increments.

3. FANMOD [37]: FANMOD can enumerate or sample from networks and is capable of finding motifs with up to eight vertices. While it can handle small networks well, it doesn't natively support weighted networks. This is a significant limitation for transaction network analysis, where the weight (value) of transactions is a crucial factor, even in small networks.

4. NeMoFinder [3]: NeMoFinder can mine frequent motifs up to size 12, which is suitable for small networks. However, its focus on tree-based partitioning might not be optimal for transaction networks where cyclic patterns could be equally important, regardless of network size.

5. Kavosh [16]: Kavosh can mine overlapping embeddings and find motifs of size more than eight vertices. For small networks, its thorough enumeration strategy could be beneficial. However, the computational intensity of its approach might be unnecessary for very small networks and could potentially overcomplex the analysis.

6. Mfinder : Developed by Kashtan et al. [17], mfinder employs a sampling method for subgraph counting. This approach allows for efficient estimation of subgraph frequencies in large networks, providing a balance between accuracy and computational efficiency.

Among these algorithms, mfinder stands out for its efficient sampling approach and ability to handle large and small networks. However, for the specific needs of analyzing transaction networks, particularly those of smaller size, Pymfinder emerges as the most suitable choice. It offers enhanced features for comprehensive motif analysis in both binary and weighted networks [2].

The subsequent section further explores how these features can be effectively applied to transaction network analysis. It will provide an in-depth examination of the specific functionalities and methodologies employed by pymfinder, demonstrating its utility in gaining insights into the complexities of financial transactions and their associated carbon emissions.

Chapter 3

Methodology

This chapter outlines the detailed methodology employed in our study to analyze card transaction data and its associated carbon emissions using network science techniques. We begin in Section 3.1 by describing our data collection and preprocessing steps, including the unique features of the Ekko dataset and our criteria for filtering and preparing the data for analysis. Section 3.2 explains our approach to network construction, detailing how we transform transaction sequences into directed graphs. In Section 3.3, we provide a comprehensive explanation of our motif detection and analysis process, including our use of the pymfinder tool and our rationale for focusing on 3-node motifs. Finally, Section 3.4 introduces our novel methodology for calculating motif-level carbon emissions, which integrates structural insights from motif analysis with transaction-specific emission data. This chapter provides a thorough and transparent account of our research process, enabling replication and validation of our findings.

3.1 Data Collection and Preprocessing

3.1.1 Dataset Description

The foundation of our research is the Ekko dataset, a rich repository of consumer card transaction information that provides a window into the financial behaviors and environmental impacts of a diverse customer base. This dataset is a goldmine of information, containing several crucial elements that enable our in-depth analysis:

1. **Merchant Category Codes (MCCs):** These standardized four-digit numbers classify each business by the type of goods or services it provides. MCCs allow us to categorize transactions and identify patterns across different types of purchases.

2. **Transaction Timestamps:** Each transaction is marked with a precise timestamp, allowing us to reconstruct the chronological sequence of purchases for each customer. This temporal data is crucial for understanding the flow of transactions and identifying recurring patterns over time.

3. **Transaction Amounts:** The dataset includes the monetary value of each transaction in pence (p). This fine-grained measurement of transaction values (1/100th of a pound sterling) provides a highly detailed view of spending patterns, allowing for precise analysis of even small purchases.

4. **Carbon Emissions:** A unique aspect of the Ekko dataset is the inclusion of carbon emission estimates for each transaction, measured in grams (g) of CO2 equivalent. These estimates are calculated using a definitive multiplier based on the MCC of the transaction. Specifically:

    - Each MCC is associated with a predefined carbon emission factor.
    - The carbon emission for a transaction is calculated by multiplying the transaction amount (in pence) by the MCC-specific emission factor.
    - The result is a carbon emission value in grams, providing a tangible and relatable measure of environmental impact.
    - This method provides a standardized approach to estimating the environmental impact of diverse types of purchases, allowing for consistent comparison across different transaction categories.

This comprehensive dataset allows us to directly link financial activity with environmental impact, providing a rare opportunity to study the relationship between consumer behavior and carbon footprints at a granular level. The combination of precise transaction amounts in pence and gram-level carbon calculations enables us to analyze not just the frequency of certain types of purchases, but also their exact impact on a customer's overall carbon footprint.

3.1.2 Data Filtering and Preparation

To ensure the robustness and relevance of the analysis, a series of filtering criteria are applied to the raw dataset:

1. **Transaction Type Filter:** The analysis considers only debit transactions. This focus on debit transactions allows for a more consistent analysis of direct consumer

spending patterns and their associated carbon emissions, excluding credit-based or other types of financial transactions that might introduce additional complexities or biases into the analysis.

2. **Transaction Amount Threshold:** A minimum transaction amount threshold of 10p is implemented. This decision is based on prior research suggesting that transactions below this amount are often preauthorization charges or other minor transactions that contribute minimally to overall carbon emissions [34]. By focusing on transactions above this threshold, the analysis captures the most impactful consumer behaviors while filtering out noise from inconsequential transactions.

3. **Customer Selection Criteria:** To guarantee that the analysis is based on customers with sufficiently rich and diverse transaction histories, two key filters are applied:

   - A minimum of 3 unique MCCs per customer
   - A minimum of 10 total transactions per customer

   These criteria serve multiple purposes:

   - They ensure that each customer in the dataset has engaged in a diverse range of transaction types, allowing for the construction of meaningful network representations of spending patterns.
   - They help to filter out customers with very limited or sporadic transaction histories, which could skew the analysis or provide insufficient data for reliable pattern detection.
   - By setting these thresholds, a balance is struck between maintaining a robust sample size and ensuring that each customer record contains enough data to reveal meaningful transaction patterns.

4. **Data Cleaning:** A thorough data cleaning process is undertaken to ensure the integrity and completeness of the dataset. Transactions with missing values for key features used in the analysis, such as Merchant Category Codes (MCCs) or transaction amounts, are excluded from the dataset. This step is crucial for maintaining the reliability and consistency of the subsequent analysis.

This careful preprocessing stage lays the groundwork for all subsequent analyses, ensuring to build complex network models and derive the insights from a foundation of clean, relevant, and meaningful data.

3.1.3 Customer Classification for Comparative Analysis

A critical component of the methodology, aligned with the primary objective of this study, was the classification of customers based on their carbon emission intensity relative to spending. This classification serves as the foundation for the comparative analysis of high and low emitters from both network and motif perspectives.

To establish a meaningful basis for comparison, an emission-to-spend ratio for each customer was devised and, calculated as follows:

$$\text{emission\_to\_spend\_ratio} = \sum \text{transaction\_emissions} \setminus \sum \text{transaction\_amounts}$$

where ***transaction_emissions*** , represents carbon emissions associated with transaction for a given customer, and ***transaction_amount***, is the total monetary value of that transaction.

Following the computation of this ratio for all customers, a percentile-based approach was employed to categorize them into distinct groups.

1. **High emitters:** Customers falling within the top 25th percentile of the emission-to-spend ratio distribution.
2. **Low emitters:** Customers in the bottom 25th percentile of this distribution.

This percentile-based approach ensures a clear delineation between groups, allowing for robust comparative analysis. By focusing on these two distinct segments, this experiment aim to identify significant differences in network topologies and motif compositions that may contribute to varying levels of carbon emissions. This approach aligns with established methodologies in environmental impact studies. For instance, Yue D. et al., [7] employed a similar percentile-based classification to study carbon footprints in household consumption, finding it effective for identifying key differences between high and low emission groups. Furthermore, Wang et al. [36] utilized quartile-based segmentation in their analysis of carbon emissions in rising middle and rich classes in China, demonstrating the value of this approach in revealing distinct consumption patterns associated with varying emission levels. These studies support the methodology, suggesting that such classification can reveal meaningful insights into the relationship between consumer behavior and carbon emissions.

3.2 Network Construction

3.2.1 Graph Representation

The core of this methodology involves the transformation of each customer's transaction history into a sophisticated directed graph structure using the NetworkX library. For each customer, a directed graph $G = (V, E)$ is constructed, where:

- V (Vertices): The set of nodes in the graph, where each node $v \in V$ represents a unique Merchant Category Code (MCC).
- E (Edges): The set of directed edges in the graph, where each edge $e = (u, v) \in E$ represents a transition from one MCC u to another MCC v in the customer's chronological transaction sequence.

Each edge e = (u, v) in the graph is characterized by two critical attributes:

1. **Weight (w)**: A function $w : E \to N$ that assigns a positive integer to each edge, representing the frequency of transitions from MCC u to MCC v.
2. **Cumulative Carbon Emissions (c)**: A function $c : E \to R^+$ that assigns to each edge the total carbon emissions associated with all transitions from MCC u to MCC v.

3.2.2 Graph Construction process

The construction of these intricate transaction networks follows a systematic process that transforms raw transaction data into rich, informative graph structures:

1. **Node Creation:** For each unique MCC m encountered in a customer's transaction history, a corresponding node v_m is created in the graph. The set of all such nodes forms V.
2. **Edge Creation and Attribute Assignment:** The process iterates through the chronologically ordered transactions. For each pair of consecutive transactions (t_i, t_{i+1}) with MCCs m_i and m_{i+1} respectively:

- If an edge $e = (v_{m_i}, v_{m_{i+1}})$ exists, its weight $w(e)$ is incremented by 1.
- If it's a new edge, it is created with an initial weight $w(e) = 1$.
- The carbon emissions associated with transaction t_{i+1} are added to $c(e)$.

Graph Storage and Optimization: The constructed graphs are stored using the NetworkX library, optimizing the graph structure for efficient querying and analysis in subsequent stages of the study.

3.2.3 Mathematical formulation

Consider a customer's transaction sequence $T = [(mcc_1, e_1), (mcc_2, e_2), \ldots, (mcc_n, e_n)]$, where mcc_i represents a Merchant Category Code (MCC) and e_i the corresponding emission value:

- **Edge Weight:** For an edge (u, v),

$$w(u,v) = |\{(i, i+1) \mid mcc_i = u, mcc_{i+1} = v, 1 \leq i < n\}|$$

- **Edge Emissions:** For an edge (u, v),

$$c(u,v) = \sum_{(i,i+1)} e_{i+1} \text{ for all } (i, i+1) \text{ where } mcc_i = u \text{ and } mcc_{i+1} = v$$

- **Node Emissions:** For a node v,

$$n(v) = \text{average}(\{e_i \mid mcc_i = v, 1 \leq i \leq n\})$$

This formulation ensures that the graph G accurately represents the sequence and characteristics of a customer's transactions, capturing both the frequency of transitions between merchant categories and the associated carbon emissions.

3.3 Motif Detection and Analysis

3.3.1 Motif Detection using PymFinder

In this study, `pymfinder` was employed to analyze network motifs in the transaction data, chosen for its suitability in analyzing small, weighted networks. This process consists of the following key steps:

1. **Network Representation:** The transaction network was represented as a directed graph, where nodes correspond to Merchant Category Codes (MCCs) and edges represent transitions between MCCs in transaction sequences. The network was input to `pymfinder` as a list of edges, each represented as a tuple:

$$(\text{source\_MCC}, \text{target\_MCC}, \text{weight})$$

2. **Weight Normalization:** Weight normalization was a crucial step to ensure comparability across different transaction values. The following normalization formula was applied:

$$\text{normalized\_weight} = \frac{\text{weight} \times 100}{\text{total\_weight}}$$

   Where total_weight is the sum of all weights in the network. This normalization scales the weights to a percentage of the total weight, preserving relative importance while providing a standardized scale.

3. **Pymfinder Configuration:** In configuring `pymfinder` for the analysis, parameters were carefully selected to balance computational feasibility with analytical depth.

   - `motifsize = 3`; focusing the analysis on 3-node motifs. This choice is grounded in the common practice of network analysis, offering a balance between complexity and interpretability.
   - `weighted = True`; enabling the incorporation of the normalized weights in the motif analysis.
   - `nrandomizations = 100`; In the literature, it is common for authors to use either 100 or 1000 randomized networks; for this study, 100 was chosen due to the small to moderate size of the network [30].
   - `links = True`; to calculate statistics for links as well as nodes and thus to calculate the link participation and roles.
   - `weighted = True`; to incorporate the normalized weights in the analysis.
   - `networktype = "unipartite"`; specifying that the network is a unipartite network.

The results returned by Pymfinder provided a dataset with the details of motifs detected for further analysis. This included information on the overall motif structure of the network, frequencies of different motif types, statistical significance of motifs compared to random networks, and both node-specific and edge-specific motif roles. The subsequent analysis focused on three main areas: identifying the most frequent and statistically significant motifs, analyzing the distribution of weighted motifs to understand the impact of transaction values on network structure, and examining node-specific motif participation to identify key MCCs in the transaction patterns.

Through the application of Pymfinder with this methodology, meaningful patterns in the small, weighted transaction network were uncovered. This approach provided insights into both the structural aspects of spending patterns, offering a nuanced understanding of the underlying dynamics in the transaction data. The complexity of financial transactions was captured through the combination of topological analysis with weighted considerations, revealing aspects that might have been overlooked by purely structural or quantitative analyses alone.

3.4 Motif level Carbon Emission

This study introduces a novel methodology for analyzing carbon emissions associated with transaction patterns in consumer spending. It builds upon the motif analysis provided by pymfinder, integrating these structural insights with transaction-specific emission data to provide a nuanced understanding of how different spending patterns contribute to overall carbon footprints.

The foundation of the analysis lies in the application of pymfinder to detect motifs within the transaction network. Pymfinder provides a comprehensive breakdown of the network's structure, identifying recurring patterns of transactions (motifs), quantifying their frequencies, and detailing how individual nodes and links participate in these motifs. This structural analysis serves as the foundation upon which the emission attribution methodology is constructed.

Building upon the motif detection provided by pymfinder, a multi-step process was developed to attribute carbon emissions to specific transaction patterns:

Normalized Link Emission Calculation:

For each transition between merchant categories (represented as edges in the network), a normalized emission value is calculated:

$$\text{normalized\_emission}(u,v) = \frac{\text{emissions}(u,v)}{\text{weight}(u,v)}$$

where $\text{emissions}(u,v)$ is the total carbon emission associated with the edge, and

$\text{weight}(u,v)$ is the frequency of the transition.

Proportional Emission Attribution:

Recognizing that a single transaction can be part of multiple motifs, a proportional attribution scheme was developed. When a link participates in multiple motifs, its normalized emission is attributed to each motif based on the frequency of the link's appearance in that specific motif:

$$\text{attributed\_emission}(u,v,m) = \text{normalized\_emission}(u,v) \times \left(\frac{\text{participation\_count}(u,v,m)}{\text{weight}(u,v)}\right)$$

where $\text{participation\_count}(u,v,m)$ is the number of times the edge (u,v) appears in motif m.

Initial Node Emission:

This experiment incorporates a distinct treatment for the initial nodes of certain motifs. It is based on the understanding that in the transaction network, edge emissions represent the carbon footprint of the target MCC. Consequently, the emission of the initial node in a motif is not captured by any incoming edge:

$$\text{initial\_node\_emission}(v,m) = \text{avg\_emission}(v) \times \text{role\_count}(v,m)$$

where $\text{avg\_emission}(v)$ is the average emission of node v across all its occurrences, and $\text{role\_count}(v,m)$ is the number of times v plays an initial role in motif m.

Motif Emission Aggregation:

The final step in the emission attribution process involves aggregating the emissions for each motif, reflecting both the transitions between categories and the initiating transactions:

$$\text{motif\_emission}(m) = \sum(\text{attributed\_emission}(u,v,m)) + \sum(\text{initial\_node\_emission}(v,m))$$

summed over all edges (u,v) and initial nodes v in motif m.

Chapter 4

Results

This chapter presents the results of our analysis, following the methodology outlined in Chapter 3. We begin with an overview of our preprocessed dataset in Section 4.1. Section 4.2 details the findings from our network analysis, comparing the structural characteristics of transaction networks between high and low carbon-emitting customers. Section 4.3 presents the results of our motif analysis, including frequency distributions, statistical significance, weighted patterns, and associated carbon emissions. Throughout, we highlight key differences in transaction behaviors and network structures between customer groups, providing a foundation for our subsequent interpretations.

4.1 Data Description

Following the pre-processing steps outlined in Section 3.1.2, our final dataset consisted of 332 customers. As described in Section 3.1.3, we classified these customers into high and low emission groups based on their emission-to-spend ratio. This classification resulted in two balanced groups of 83 customers each, representing the extremes of our emission-to-spend ratio distribution. The subsequent analyses in this chapter are based on these two customer groups, allowing for direct comparison of transaction patterns and network structures associated with high and low emission behaviors.

4.2 Network Analysis

Following the network construction methodology described in Section 3.2, we generated transaction networks for both high and low emission customers. These networks represent the patterns of transactions between different Merchant Category Codes

(MCCs), with nodes symbolizing MCCs and edges representing transitions between them.

4.2.1 Network Statistics

To characterize these networks, we computed several key network statistics, as presented in Table 4.1 below.

Table 4.1: Comparison of key network metrics between high and low carbon-emitting customer transaction networks. Values represent means with standard deviations in parentheses. Significance was determined using t-tests with $p < 0.05$

Metric	**High_Emission Customers**	**Low_Emission Customers**	**T-Statistic**	**Significance**
Average Degree	Mean: 4.60, Std Dev: 1.32	Mean: 4.01, Std Dev: 1.27	2.96	Significant $p < 0.05$
Number of Nodes	Mean: 12.22, Std Dev: 8.05	Mean: 9.61, Std Dev: 4.63	2.55	Significant $p < 0.05$
Number of Edges	Mean: 31.69, Std Dev: 30.42	Mean: 20.46, Std Dev: 15.06	3.01	Significant $p < 0.05$
Density	Mean: 0.2959, Std Dev: 0.1382	Mean: 0.2822, Std Dev: 0.1421	-0.28	Not Significant

Statistical analysis of the network metrics reveals significant structural differences between the transaction networks of high and low carbon-emitting customers. High emitters exhibit larger and more complex networks, characterized by a greater number of nodes and edges. This suggests that high-emission customers interact with a more diverse range of merchant categories and engage in more frequent inter-category transactions.

The higher average degree observed in high-emitter networks indicates increased connectivity between MCCs, implying more varied transaction sequences. This heightened interconnectedness could be indicative of more complex spending patterns among high-emission customers, potentially contributing to their elevated carbon footprint.

Interestingly, despite these differences in size and connectivity, network density does not significantly differ between the two groups. This similarity in density suggests that the fundamental structure of transactional relationships remains consistent across emission levels, with differences primarily manifesting in the scale and complexity of these relationships.

The consistently larger standard deviations observed in the high-emission group across all metrics point to greater variability in network structures. This heterogeneity implies that high-emission behavior may not conform to a single, uniform pattern, but rather encompasses a spectrum of complex spending habits that collectively result in elevated carbon emissions.

Figure 4.1 illustrates representative transaction networks for a high and a low carbon-emitting customer, providing a visual representation of the structural differences between these groups.

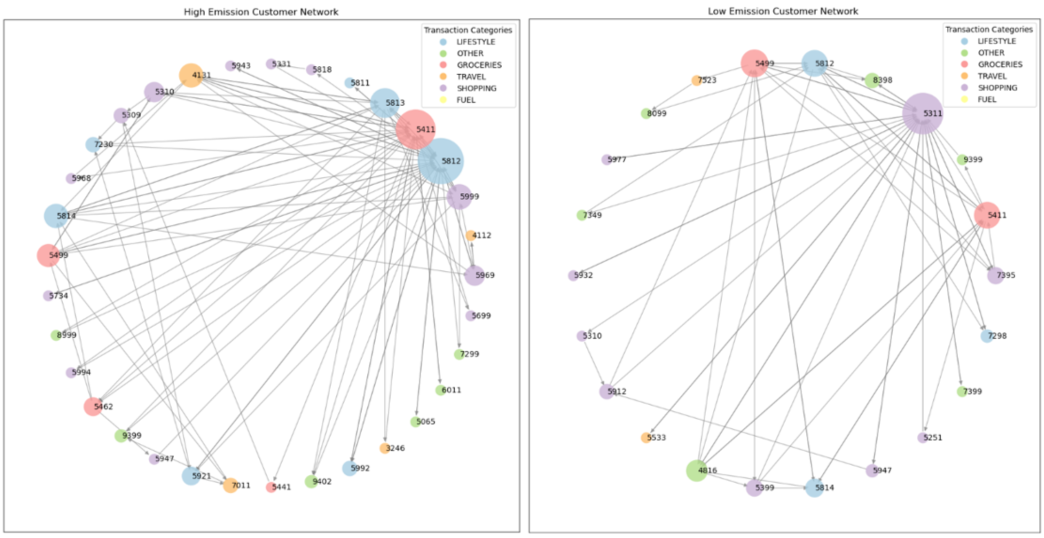


Figure 4.1: Representative transaction networks for (a) a high carbon-emitting customer and (b) a low carbon-emitting customer. Nodes represent Merchant Category Codes (MCCs), and edges represent transitions between MCCs. Node size indicates transaction frequency.

While these high-level network statistics provide valuable insights into the overall structure of transaction networks for high and low emission customers, they don't capture the specific patterns of transactions that may be driving these differences. The similar network densities, combined with differences in network size and connectivity, hint at more detailed structural differences that these broad metrics don't fully capture.

To gain a deeper understanding of the underlying transaction behaviors and their relationship to carbon emissions, we turn to motif analysis as described in Section 3.3. In the subsequent sections, we present our findings on motif distributions, their associated carbon emissions, and the distinctions observed between high and low emission groups.

4.3 Motif Analysis

Following the motif detection and analysis methodology described in Section 3.3, we examined the prevalence and characteristics of 3-node motifs in our transaction networks. This section presents the results of our motif analysis, including frequency distributions, statistical significance, weighted patterns, and associated carbon emissions.

4.3.1 Motif Detection

As per the Motif results from Pymfinder, 13 possible 3-node motifs in directed networks are detected, as illustrated in Figure 4.2. These motifs, labeled according to Alon et al.'s "motif dictionary" [1], represent all potential transaction patterns involving three Merchant Category Codes (MCCs) in our transaction networks.

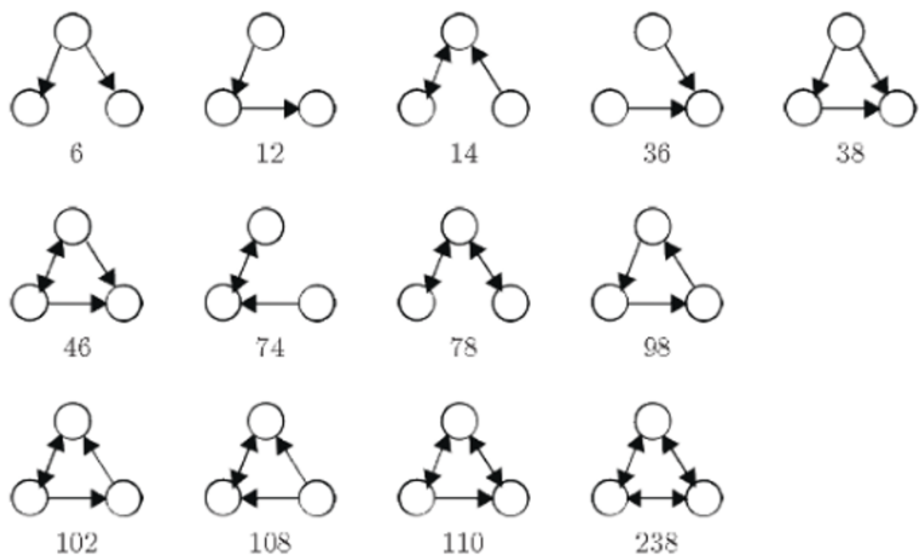


Figure 4.2: The 13 possible 3-node motifs in directed networks, as identified by Alon et al. [1][9] Each motif represents a distinct pattern of transactions between three Merchant Category Codes (MCCs).

Each motif in Figure 4.2 represents a distinct pattern of transactions between three

MCCs. For instance:

- Motif 6 represents a simple chain of transactions across three categories.
- Motif 38 shows a feedback loop where the last MCC in a chain also transacts with the first.
- Motif 238 a fully connected triad represents a scenario where transactions occur between all pairs of three MCCs in both directions.

4.3.2 Motif Frequency Distribution

Analysing the frequency of Motifs from the Figure 4.3 frequency distribution of the 13 possible 3-node motifs among high and low emission customers, the key observations are listed below:

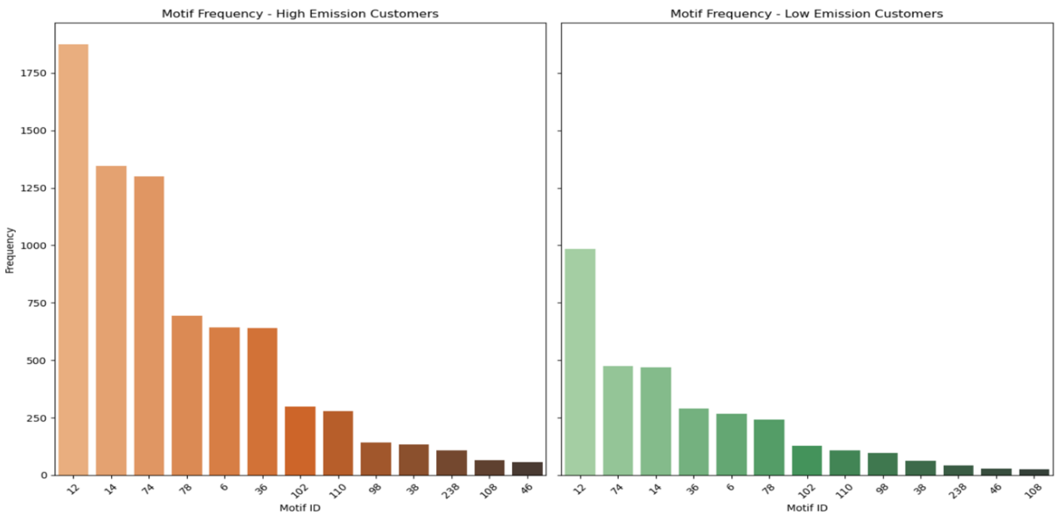


Figure 4.3: Frequency distribution of 3-node motifs among (a) high emission customers and (b) low emission customers. Motif IDs correspond to those in Figure 4.2.

- **Prevalence of Simple Motifs:** Motifs 12, 14, and 74 emerge as the most frequent in both high and low emission networks. However, their substantially higher absolute frequencies in high-emission networks suggest these patterns may be key contributors to increased emissions.
- **Complexity Gradient:** High emission customers exhibit a broader distribution of motif types, with notably higher frequencies of more complex motifs (e.g., 238, 110, 102). This suggests that intricate, multi-step transaction sequences are characteristic of high-emission behavior.

- **Structural Differences:** Low emission customers show a relative preference for simpler motifs (e.g., 6, 36), implying that more linear transaction patterns may be associated with lower emissions.

4.3.3 Statistical Significance of Motifs:

Z-Score Analysis: To assess the statistical significance of these observed frequencies, we analyse the Z-scores computed by Pymfinder for each motif, comparing their occurrence in the actual networks to randomized networks. Figure 4.4 presents the average Z-scores for each motif across high and low emission customers.

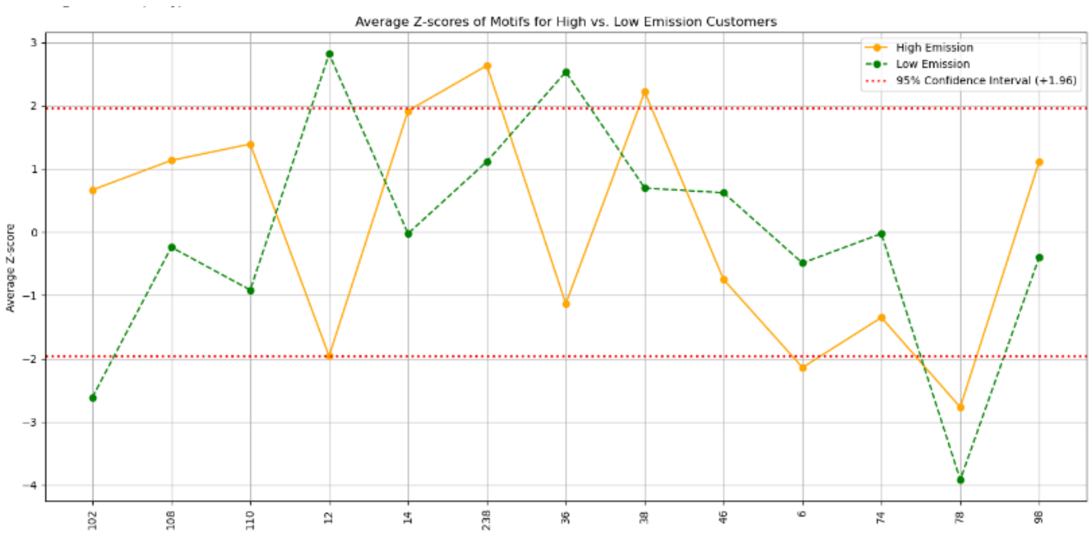


Figure 4.4: Bar chart comparing the average Z-scores of motifs for high versus low emission customers. Z-scores indicate the statistical significance of motif occurrences compared to randomized networks.

- **Significant Over representation:** Motifs 38 and 238 show significant relative over representation in high-emission networks, indicating these complex transaction patterns are characteristic of high-emission behavior.

- **Low Emission Patterns:** Motifs 12 and 36 are significantly over represented in low-emission networks, suggesting simpler, more linear transaction sequences may be associated with lower emissions.

- **Underrepresented Motifs:** Motif 78 is significantly underrepresented in both groups, but more so in high emitters, implying this pattern might be associated with more sustainable transaction behaviors.

Ratio Analysis: To provide a more intuitive measure of motif significance, we computed the ratio metric (r_i) for each motif, representing how many times more often a motif appears in the original network compared to randomized networks. Figure 4.5 presents this comparison across high and low emission customers.

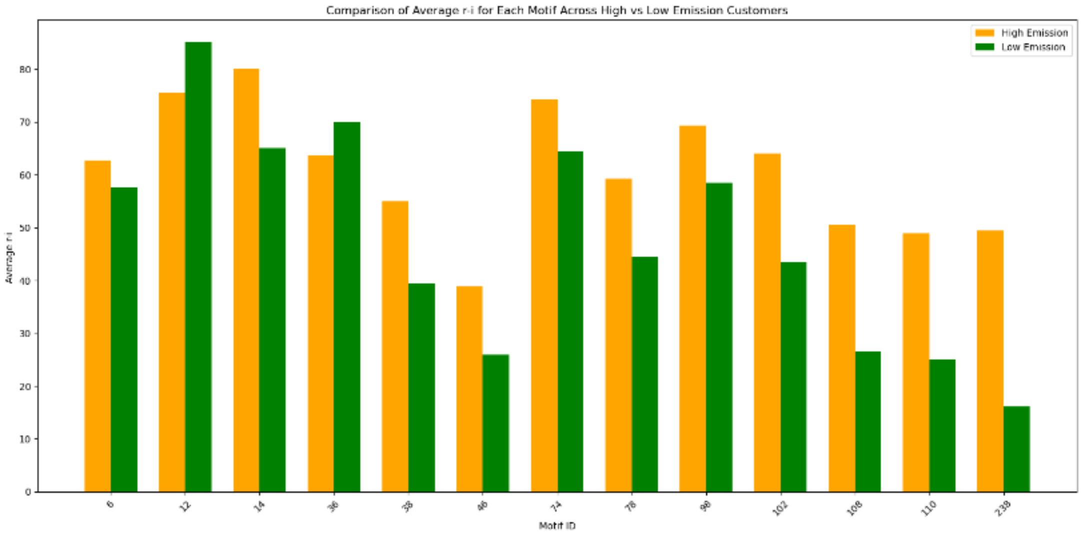


Figure 4.5: Comparison of average r_i values (ratio of motif count to random expectation) for each motif across high versus low emission customers. Higher r_i values indicate stronger overrepresentation of a motif compared to random networks.

- **Complex Motif Significance:** Motif 238 a fully connected triad shows a striking difference in r_i values, appearing over 100 times more often in high-emission networks than expected by chance, compared to only about 20 times more often in low-emission networks.
- **Feedback Loops:** Motifs involving feedback mechanisms (e.g., 14, 38) show higher r_i values for high-emission customers, suggesting that reciprocal or cyclical transaction patterns play a crucial role in high-emission behavior.
- **Simple Chain Importance:** Motif 6 (simple chain) shows a higher r_i value for high-emission customers, indicating that even basic transaction sequences are more pronounced in high-emission behavior relative to random expectation.

The comprehensive analysis of motif frequencies, Z-scores, and ratios has revealed significant structural differences in the transaction networks of high and low emission customers. High-emission networks are characterized by more complex, interconnected transaction patterns, with a higher prevalence of feedback loops and fully connected triads. Conversely, low-emission networks exhibit simpler, more linear transaction sequences. These findings suggest that the structure of transaction patterns, not just their volume or value, plays a crucial role in determining carbon emissions associated with consumer spending.

While the motif frequency analysis provides valuable insights into the general structure of transaction networks, a deeper understanding of the specific components and characteristics of these networks is necessary. As outlined in Section [3.3.1] of our methodology, we will now proceed to analyse motifs further with weights and emissions.

4.3.4 Weighted Motif Analysis

Following our examination of motif frequencies, we conducted a comparative analysis of the average weight mean versus motif occurrence for both high and low emission customer groups. This analysis provides insights into the relative importance and frequency of different transaction patterns. Figure 4.6 illustrates the relationship between average weight mean and motif occurrence for high and low emission customers.

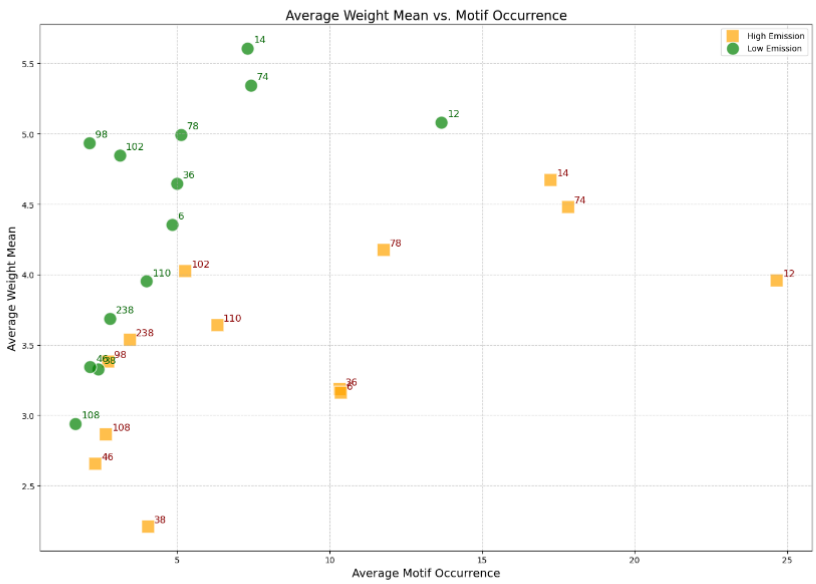


Figure 4.6: Scatter plot of average weight mean versus motif occurrence for high and low emission customers. Each point represents a unique motif, with its position indicating its frequency (x-axis) and average weight (y-axis) in the respective customer group.

Key observations from this analysis include:

1. **Frequency and Importance:** Consistent with our earlier frequency analysis,

motifs 12, 14, and 74 show high occurrences, particularly for high emitters. However, this weighted analysis reveals that their importance (weight mean) varies between groups, with low emitters often attributing higher importance to these common motifs despite lower occurrences.

2. **Complexity Gradient:** Reinforcing our previous findings, high emitters demonstrate a broader distribution of motif occurrences, including more complex motifs (e.g., 238, 110). Interestingly, while these complex motifs occur more frequently in high emitters, they often have higher weight means in low emitters, suggesting a nuanced role in transaction patterns.

3. **Simple vs. Complex Motifs:** Earlier, we noted low emitters' preference for simpler motifs. This analysis adds depth to that observation, showing that while simpler motifs (e.g., 6, 36) have lower occurrences in low emitters, they often have higher weight means, indicating their structural importance in low-emission transaction patterns.

4. **Feedback Loops:** Motifs with feedback mechanisms (e.g., 14, 38) continue to show significance. Motif 14, in particular, demonstrates high importance for both groups, aligning with its previously observed over representation. Motif 38, while less prominent in occurrences, maintains moderate weight means, suggesting a subtle but persistent role.

5. **Divergent Behaviors:** Some motifs, notably 12 and 6, show differences between groups in both occurrence and weight mean. This extends our understanding of how specific transaction patterns differentiate high and low emitters.

This weighted analysis provides a more nuanced view of our earlier findings, revealing that the importance of motifs isn't solely determined by their frequency. It highlights how certain transaction patterns, while less frequent, may play crucial roles in defining low-emission behaviors.

4.3.5 Motif Level Carbon Emission Analysis

Building on our weighted motif analysis and earlier frequency observations, we now examine the carbon emissions associated with each motif type, calculated as detailed in section [3.4]. Figure 4.7 presents the Average Carbon Emissions per Motif for High and Low Emission Customers

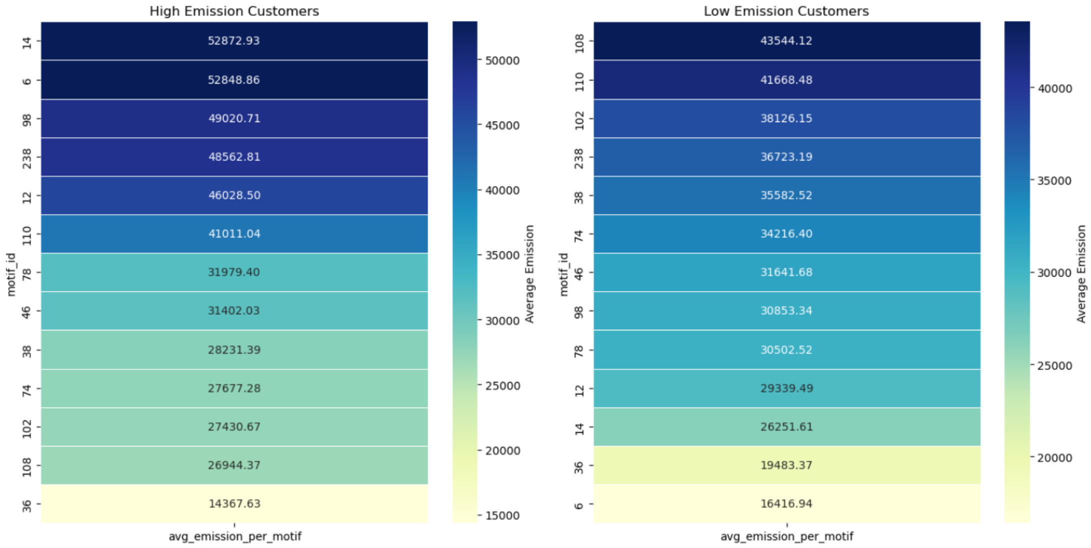


Figure 4.7: Bar chart illustrating the average carbon emissions per motif for high and low emission customers. Motif IDs correspond to those in Figure 4.2.

The emission analysis further refines our understanding:

1. **Complex Motif Emissions:** Aligning with our earlier observations on the complexity gradient, the most complex motifs (e.g., 238, 110) indeed show the high emissions, especially for high emitters. This confirms the environmental impact of intricate transaction patterns.

2. **Frequency-Emission Relationship:** Highly frequent motifs (12, 14, 74) identified in our initial analysis generally show varying emission levels. However, their emissions aren't always proportional to their frequency, revealing a complex relationship between transaction patterns and environmental impact.

3. **Feedback Loop Impact:** Motifs with feedback mechanisms, particularly 14, demonstrate high emissions across both groups, reinforcing their significant role identified in earlier analyses. Motif 38, while less prominent in emissions, still contributes to the overall pattern of feedback loops' importance.

4. **Simple Motif Variability:** Our initial observation of low emitters preferring simpler motifs is nuanced by the emission data. Some simple motifs show lower emissions for high emitters, suggesting that the relationship between motif simplicity and emissions is more complex than initially thought.

5. **Z-score and Emission Correlation:** Motifs previously identified as significantly over represented (e.g., 238) align with high emission levels, especially in

high emitters. This correlation strengthens our understanding of how structural significance relates to environmental impact.

This emission analysis complements our previous findings by quantifying the environmental impact of different transaction patterns. It confirms the significance of complex and feedback-loop motifs in higher emissions while revealing nuances in how simpler transaction patterns contribute to overall carbon footprints.

The combined insights from our frequency, weighted, and emission analyses provide a comprehensive view of how transaction network structures relate to carbon emissions.

In the next section, we will explore the specific Merchant Category Code (MCC) sequences that compose these high-impact motifs, further refining our understanding of the relationship between transaction patterns and carbon emissions.

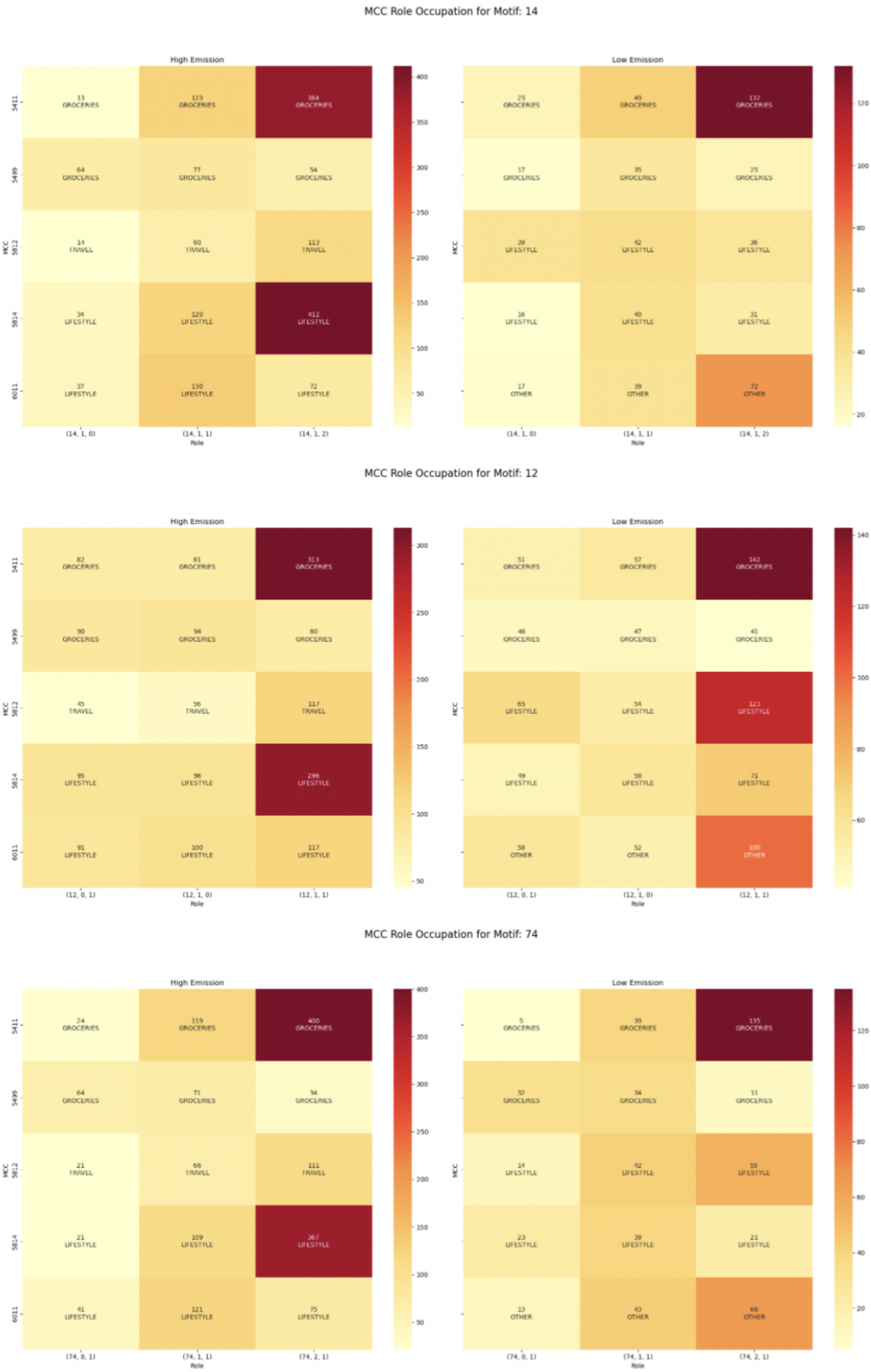
4.3.6 Motif Composition and Node Role Analysis

Building upon our earlier frequency and Z-score analyses, we examined the composition and node roles of the most significant motifs for high and low emission customers. Motifs 12, 14, and 74 were most frequent for both high and low emission customers. Z-score analysis revealed motifs 238 and 38 as most significant for high emitters, while motifs 12 and 36 were most significant for low emitters. Table 4.2 summarizes the composition of top motifs identified based on frequency and Z-score significance. Figures 4.8 illustrate the MCC distribution across different node roles for motifs 14, 12, and 74 in both high and low emission customers.

Node roles in pymfinder are represented using a notation that captures the position of a node within a motif. The notation follows the format (motif ID, number of outgoing links, number of incoming links). For example, a node role labeled as (14, 1, 2) indicates that the node is part of motif 14, has 1 outgoing link, and 2 incoming links within that motif. This notation allows for a detailed understanding of how each node contributes to the structure of various motifs in the network.

Interpretation

1. **Motif Composition Patterns:** The most frequent motifs (12, 14, 74) are common to both high and low emitters, suggesting fundamental transaction patterns across all consumers. However, the composition of these motifs differs significantly between the two groups. High emitters show strong bidirectional flows between Grocery Stores and Eating Places across all top motifs. In contrast, low emit-



externalParty.merchantCategoryCode	externalParty.transactionCategory	MCC_List
5411	GROCERIES	Grocery Stores and Supermarkets
5499	GROCERIES	Miscellaneous Food Stores-Convenience Stores and Specialty Markets
5541	TRAVEL	Service Stations (with or without Ancillary Services)
5812	LIFESTYLE	Eating Places and Restaurants
5814	LIFESTYLE	Quick Payment Service-Fast Food Restaurants
6011	OTHER	Financial Institutions--Automated Cash Disbursements

Figure 4.8: MCC distribution across different node roles for motifs 14, 12, and 74 in both high and low emission customers. Each bar represents the proportion of MCCs from different categories occupying specific roles within the motif.

Table 4.2: Composition of Top Motifs for High and Low Emission Customers

Motif ID	Emission Group	Top Transitions	Dominant MCCs
12	High	High: Grocery → Misc. Food (38), Grocery → Fast Food (30)	High: Eating Places (487), Grocery (476), Fast Food (308)
	Low	Low: Fast Food → Grocery (28), Grocery → Eating Places (25)	Low: Grocery (250), Eating Places (242), Financial (210)
14	High	High: Grocery ↔ Eating Places (73, 71)	High: Eating Places (566), Grocery (520), Fast Food (239)
	Low	Low: Grocery → Eating (32), Financial → Grocery (31)	Low: Grocery (206), Financial (128), Eating Places (117)
74	High	High: Eating Places ↔ Grocery (69 each)	High: Grocery (543), Eating Places (497), Fast Food (237)
	Low	Low: Financial → Grocery (29), Eating → Grocery (29)	Low: Grocery (179), Financial (122), Eating Places (111)
238	High	HighEating Places ↔ Grocery (37 each)	Grocery (70), Eating Places (63), Fast Food (34)
38	High	Grocery → Misc. Food (7), Local Transport → Misc. Food (5)	Misc. Food (36), Grocery (32), Fast Food (27)
36	Low	Misc. Food → Grocery (19), Misc. Retail → Grocery (15)	Eating Places (169), Grocery (169), Fast Food (105)

ters exhibit more diverse transitions, with significant involvement of Financial Institutions, particularly in motifs 14 and 74.

2. **Category Dominance:** While both groups have Grocery Stores and Eating Places as dominant MCCs, high emitters show a higher frequency of Fast Food establishments. Low emitters, notably, have a strong presence of Financial Institutions across their top motifs, suggesting a potential link between financial management and lower emissions.

3. **Complex vs. Structured Patterns:** Motifs 238 and 38, significantly overrepresented in high emitters, represent more complex, interconnected transaction patterns. Low emitters' significant motifs (12 and 36) show more structured, single-input patterns, potentially indicating more planned purchasing behavior.

4. **Node Role Distribution:** In feed-forward loops (motif 14), high emitters show a concentration of GROCERIES and LIFESTYLE categories across all roles. Low emitters display a more balanced distribution, with the OTHER category (primarily Financial Institutions) playing a significant role, especially in the (14, 1, 2) position.

5. **Feedback Mechanisms:** The fully connected motif (74) in high emitters reveals

intense cycling between GROCERIES and LIFESTYLE categories across all positions. Low emitters incorporate more OTHER category transactions in these cycles, particularly in intermediary positions, suggesting financial transactions play a key role in their feedback patterns.

6. **Role Specialization:** High emitters demonstrate role specialization, with GROCERIES and LIFESTYLE dominating specific positions, especially endpoints. Low emitters show less specialization, maintaining a more balanced category distribution across different roles, including starting positions.

These findings align with and extend our previous results, highlighting that while basic transaction patterns are common across all consumers, the specific composition, complexity, and role distribution within these patterns significantly influence emission levels. The concentration and cycling of food-related categories in high emitters versus the more balanced, financially-integrated patterns in low emitters provide crucial insights into how transaction behaviors relate to carbon footprints.

The over representation of complex, fully connected motifs in high emitters versus more structured, single-input patterns in low emitters suggests that transaction complexity itself may be a factor in higher emissions. The prominent role of financial transactions in low emitter patterns is a notable difference from high emitters. This observation suggests a potential relationship between the frequency of financial transactions and lower emission patterns. However, the nature of this relationship and its implications for sustainable spending habits require further investigation. It's possible that this pattern reflects differences in overall financial management or transaction strategies between the two groups, rather than a direct causal link to lower emissions.

These insights offer a nuanced understanding of the relationship between consumer transaction patterns and carbon emissions, providing a foundation for developing targeted interventions to promote more sustainable consumer behaviors.

4.4 Key Findings

This study's comprehensive analysis of transaction networks reveals significant structural and behavioral differences between high and low carbon-emitting consumers, providing valuable insights into the relationship between spending patterns and environmental impact. The key findings from each analytical phase build upon one another to form a cohesive understanding of this complex relationship:

1. **Network Structure:** High-emission customers exhibit larger, more complex transaction networks with greater connectivity between merchant categories, suggesting a link between diverse, frequent inter-category transactions and elevated carbon footprints.

2. **Motif Analysis:** While certain basic patterns are common across all consumers, high emitters show a broader distribution of motif types, with significantly higher occurrences of complex motifs. Z-score analysis confirms the overrepresentation of intricate, multi-step transaction sequences in high-emission behavior.

3. **Weighted Patterns:** Some simpler motifs show higher weight means in low-emission networks, indicating their importance in sustainable transaction patterns despite lower frequency.

4. **Emission Quantification:** Complex motifs (e.g., 238, 110) indeed show the highest emission levels, but the relationship between motif complexity and emissions is not strictly linear.

5. **Transaction Behaviors:** High emitters demonstrate strong bidirectional flows between grocery and lifestyle categories, while low emitters exhibit more diverse transitions, notably incorporating financial institutions across their significant motifs.

This study reveals that the relationship between consumer transaction patterns and carbon emissions is more nuanced than previously understood. The structure and composition of spending behaviors, rather than just volume or value, play a crucial role in determining environmental impact. High carbon emissions are associated with complex, repetitive transaction patterns, particularly those cycling between grocery and lifestyle purchases. In contrast, lower emissions are linked to more diverse, structured patterns with a stronger presence of financial transactions.

These findings provide a novel perspective on consumer behavior and environmental impact, suggesting that promoting sustainable consumption requires addressing not only what consumers purchase but also how they structure their overall spending patterns. This insight opens new avenues for developing targeted interventions and strategies to encourage more environmentally friendly consumer behaviors, potentially revolutionizing approaches to sustainability in personal finance and consumption.

Chapter 5

Discussions and Conclusions

5.1 Discussions

Our study reveals significant structural and behavioral differences in transaction networks between high and low carbon-emitting consumers, providing novel insights into the relationship between spending patterns and environmental impact.

Network Structure and Complexity: High-emission customers exhibit larger, more complex transaction networks with greater connectivity between merchant categories. This aligns with Di Clemente et al.'s [4] work on urban lifestyles but extends it by directly linking complex spending patterns to higher carbon emissions. Our findings challenge simplistic volume-based models of consumption and emissions, highlighting the importance of transaction network structure in contributing to elevated carbon footprints [34].

Motif Analysis and Emission Patterns: The application of network motifs to carbon emission analysis bridges structural network analysis and environmental impact studies. Our findings support Milo et al.'s [25] principle that complex structures often indicate significant functional roles in networks. The overrepresentation of complex motifs in high-emission networks provides a new tool for understanding and potentially predicting high-emission behaviors.

Weighted Patterns and Sustainable Transactions: Our weighted motif analysis challenges the assumption that complexity always correlates with higher emissions. This aligns with recent research suggesting that sustainable behaviors can be embedded in simple, repetitive patterns [20]. It implies that promoting sustainable consumption may involve encouraging certain types of transaction patterns rather than simply reducing overall consumption.

Non-Linear Relationship Between Motif Complexity and Emissions: The observed non-linear relationship between motif complexity and emissions aligns with recent work in complex systems theory applied to environmental studies such as Ma et al., [24]. These findings challenges researchers to develop more sophisticated models of consumer behavior and its environmental impact.

Transaction Behaviors and Category Interactions: Our findings on distinct transaction behaviors between high and low emitters build upon previous work on consumer segmentation and lifestyle analysis by Dong et al. [6]. The observation that low emitters incorporate more financial transactions in their significant motifs is a novel finding that warrants further investigation, potentially linking financial management practices to lower emissions.

In conclusion, our study provides several novel insights that both support and extend existing literature on consumer behavior and environmental impact. By applying network analysis and motif detection to transaction data with a focus on carbon emissions, we've uncovered complex patterns that challenge simplistic models of consumption and sustainability. These findings suggest the need for more nuanced, structure-aware approaches to promoting sustainable consumer behavior, opening new avenues for research and intervention in the fight against climate change.

5.2 Conclusion

This study has provided novel insights into the relationship between consumer transaction patterns and carbon emissions through the application of network analysis and motif detection techniques. Our findings reveal a strong association between network structure complexity and higher carbon emissions, with specific network motifs characteristic of high and low emission behaviors. Notably, the relationship between motif complexity and carbon emissions is non-linear, challenging existing models of consumer behavior and environmental impact.

Key discoveries include the association of low-emission behaviors with more diverse transaction patterns and a higher frequency of financial transactions, suggesting a potential link between financial management practices and sustainable consumption. These findings address our initial research questions, identifying characteristic network structures, motif distributions, and specific MCC sequences associated with high and low carbon emissions.

However, it is important to acknowledge the limitations of this study. The reliance on a single dataset from one financial institution may limit the generalizability of our findings. The carbon emission calculations, based on predefined multipliers for each Merchant Category Code (MCC), may not capture the full complexity of product-specific emissions. Additionally, our focus on 3-node motifs might overlook larger-scale patterns, and we have not accounted for potential seasonal variations in transaction patterns.

Future research directions could address these limitations and extend our findings. Incorporating data from multiple financial institutions would improve the generalizability of the results. Developing more nuanced carbon emission calculations that consider product-specific factors could provide a more accurate picture of environmental impact. Longitudinal studies could account for seasonal and long-term changes in transaction patterns, while expanding the analysis to include larger motif sizes might uncover additional significant patterns.

This study has demonstrated the value of applying complex network analysis to consumer behavior and its environmental impact. If replicated with additional resources, future studies could benefit from more detailed data on specific products purchased within each MCC to refine emission calculations. Incorporating qualitative research methods to understand the motivations behind different transaction patterns could provide deeper insights. Furthermore, developing and testing interventions based on these findings to promote more sustainable transaction behaviors would be a logical next step.

In conclusion, this research has opened new avenues for understanding and potentially influencing consumer behavior towards more sustainable practices. By revealing the complex relationship between transaction patterns and carbon emissions, it provides a foundation for developing more targeted and effective strategies to promote environmental sustainability in consumer spending. As society grapples with the challenges of climate change, insights derived from such analyses will be crucial in shaping policies, business practices, and individual behaviors that can contribute to a more sustainable future.

Bibliography

[1] Uri Alon. Uri alon's systems biology lab at weizmann institute of science. `http://www.weizmann.ac.il/mcb/UriAlon/`. Accessed: Aug. 22, 2024.

[2] Bernat Bramon Mora, Alyssa Cirtwill, and Daniel Stouffer. pymfinder: a tool for the motif analysis of binary and quantitative complex networks, 2018.

[3] Jin Chen, Wynne Hsu, Mong Lee, and See-Kiong Ng. Nemofinder: Dissecting genome-wide protein-protein interactions with meso-scale network motifs. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006:106–115, 08 2006.

[4] R. Di Clemente, M. Luengo-Oroz, M. Travizano, et al. Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications*, 9(3330), 2018.

[5] Diane J. Cook and Lawrence B. Holder. Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1:231–255, 1993.

[6] X. Dong, Y. Suhara, B. Bozkaya, V. Singh, B. Lepri, and A. Pentland. Social bridges in urban purchase behavior. *ACM Transactions on Intelligent Systems and Technology*, 9:1–29, 2017.

[7] Yue Dou, Jun Zhao, Xiucheng Dong, and Kangyin Dong. Quantifying the impacts of energy inequality on carbon emissions in china: A household-level analysis. *Energy Economics*, 102:105502, 2021.

[8] X. Fontes, D. Aparício, M. I. Silva, B. Malveiro, J. T. Ascensão, and P. Bizarro. Finding nemo: Fishing in banking networks using network motifs, 2021.

[9] Reid Ginoza and Andrew Mugler. Network motifs come in sets: Correlations in the randomization process. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 82:011921, 07 2010.

[10] D. Goldenberg. Social network analysis: From graph theory to applications with python, 2021.

[11] G. Guidi, R. Mastrandrea, A. Facchini, T. Squartini, and C. Kennedy. Tracing two decades of carbon emissions using a network approach. *Scientific Reports*, 14(1):7251, 2024.

[12] S. Hampton and L. Whitmarsh. Choices for climate action: A review of the multiple roles individuals play. *One Earth*, 6(9):1157–1172, Sep. 2023.

[13] K. Inafuku, T. Fushimi, and T. Satoh. Extraction method of typical purchase patterns based on motif analysis of directed graphs. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services (iiWAS '16)*, pages 86–95, 2016.

[14] Intergovernmental Panel on Climate Change (IPCC). Summary for policymakers of ipcc special report on global warming of 1.5°c approved by governments. `https://www.ipcc.ch/2018/10/08/summary-for-policymakers-of-ipcc-special-report-on-global-warming-of-1-5c-` 2018. Accessed: 2024-08-14.

[15] J. Jiang, Y. Hu, X. Li, W. Ouyang, Z. Wang, F. Fu, and B. Cui. Analyzing online transaction networks with network motifs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, pages 3098–3106, 2022.

[16] Zahra Kashani, Hayedeh Ahrabian, Elahe Elahi, Abbas Nowzari, Elnaz Saberi Ansari, Sahar Asadi, Shahin Mohammadi, and Falk Schreiber. Kavosh: A new algorithm for finding network motifs. *BMC bioinformatics*, 10:318, 10 2009.

[17] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20 11:1746–58, 2004.

[18] F. Kooti, K. Lerman, L. M. Aiello, M. Grbovic, N. Djuric, and V. Radosavljevic. Portrait of an online shopper: Understanding and predicting consumer behavior.

In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 205–214, 2016.

[19] C. Krumme, A. Llorente, M. Cebrian, A. S. Pentland, and E. Moro. The predictability of consumer visitation patterns. *Scientific Reports*, 3:1645, 2013.

[20] S. Kumar, D. Sharma, S. Rao, et al. Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 2022.

[21] M. Kuramochi and G. Karypis. Grew - a scalable frequent subgraph discovery algorithm. In *Fourth IEEE International Conference on Data Mining (ICDM'04)*, pages 439–442, 2004.

[22] Y. Li, A. Hao, X. Zhang, and X. Xiong. Network topology and systemic risk in peer-to-peer lending market. *Physica A: Statistical Mechanics and its Applications*, 508:118–130, 2018.

[23] P. Liu, R. Acharyya, R. E. Tillman, S. Kimura, N. Masuda, and A. E. Sarıyüce. Temporal motifs for financial networks: A study on mercari, jpmc, and venmo platforms, 2023.

[24] N. Ma, H. Li, Y. Wang, S. Feng, J. Shi, and K. Wang. Interaction pattern features and driving forces of intersectoral co2 emissions in china: A network motif analysis. *Resources, Conservation and Recycling*, 149:391–412, 2019.

[25] M. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, Oct. 2002.

[26] D. Moran, R. Wood, E. Hertwich, K. Mattson, J. F. D. Rodriguez, K. Schanes, and J. Barrett. Quantifying the potential for consumer-oriented policy to reduce european and foreign carbon emissions. *Climate Policy*, 20(sup1):S28–S38, 2018.

[27] S. C. Obiora, O. Bamisile, E. Opoku-Mensah, and A. N. K. Frimpong. Impact of banking and financial systems on environmental sustainability: An overarching study of developing, emerging, and developed economies. *Sustainability*, 12(8074), 2020.

[28] T. Puschmann, C. H. Hoffmann, and V. Khmarskyi. How green fintech can alleviate the impact of climate change—the case of switzerland. *Sustainability*, 12(10691), 2020.

[29] T. Raeder and N. Chawla. Market basket analysis with networks. *Social Network Analysis and Mining*, 1:97–113, 2011.

[30] Pedro Ribeiro, Fernando Silva, and Marcus Kaiser. Strategies for network motifs discovery. In *2009 Fifth IEEE International Conference on e-Science*, pages 80–87, 2009.

[31] C. Rodriguez-Sanchez. The role of social marketing in achieving the planet sustainable development goals (sdgs). *International Review on Public and Nonprofit Marketing*, 20:559–571, 2023.

[32] D. Savage, Q. Wang, P. Chou, X. Zhang, and X. Yu. Detection of money laundering groups using supervised learning in networks, 2016.

[33] A. Srivastava. Motif analysis in the amazon product co-purchasing network, 2010.

[34] A. Trendl, A. Owen, L. Vomfell, L. Kilian, J. Gathergood, N. Stewart, and D. Leake. Estimating carbon footprints from large scale financial transaction data. *Journal of Industrial Ecology*, 27:56–70, 2023.

[35] United Nations. Report of the intergovernmental committee of experts on sustainable development financing. `https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=2404`, 2014. Accessed: 2024-08-14.

[36] Liyuan Wei, Cai Li, Ju Wang, Xiao Wang, Zhen Wang, Can Cui, Sha Peng, Yong Liu, Shuxia Yu, Ling Wang, and Zhihua Shi. Rising middle and rich classes drove china's carbon emissions. *Resources, Conservation and Recycling*, 159:104839, 2020.

[37] Sebastian Wernicke and Florian Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics (Oxford, England)*, 22(9):1152–1153, 2006.

[38] R. Wood, D. Moran, K. Stadler, D. Ivanova, K. Steen-Olsen, A. Tisserant, and E. G. Hertwich. Prioritizing consumption-based carbon policy based on the evaluation of mitigation potential using input-output methods. *Journal of Industrial Ecology*, 22:540–552, 2018.

[39] María Óskarsdóttir, Cristián Bravo, Wouter Verbeke, Carlos Sarraute, Bart Baesens, and Jan Vanthienen. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85:204–220, 2017.