Analysing community-level spending behaviour contributing to high carbon emissions using stochastic block models

Ognyan Simeonov,
PhD Student, MAC-MIGS CDT

Full paper on Scientific Reports
doi.org/10.1038/s41598-025-14364-7

Expenditure Datasets + Network Science = Policy

Research Overview

1

Identify spending/emissions patterns from micro-level data (transactions)

- interventions on individual level
- changes within companies

2

Observe how these patterns connect to macro-level data (LCFS, regions, LSOA)

- interventions on wider policy level
- changes within industries/regions

Data Overview

Partner with ekko – a sustainable banking FinTech company, focusing on tracking consumer emissions

www.ekko.earth

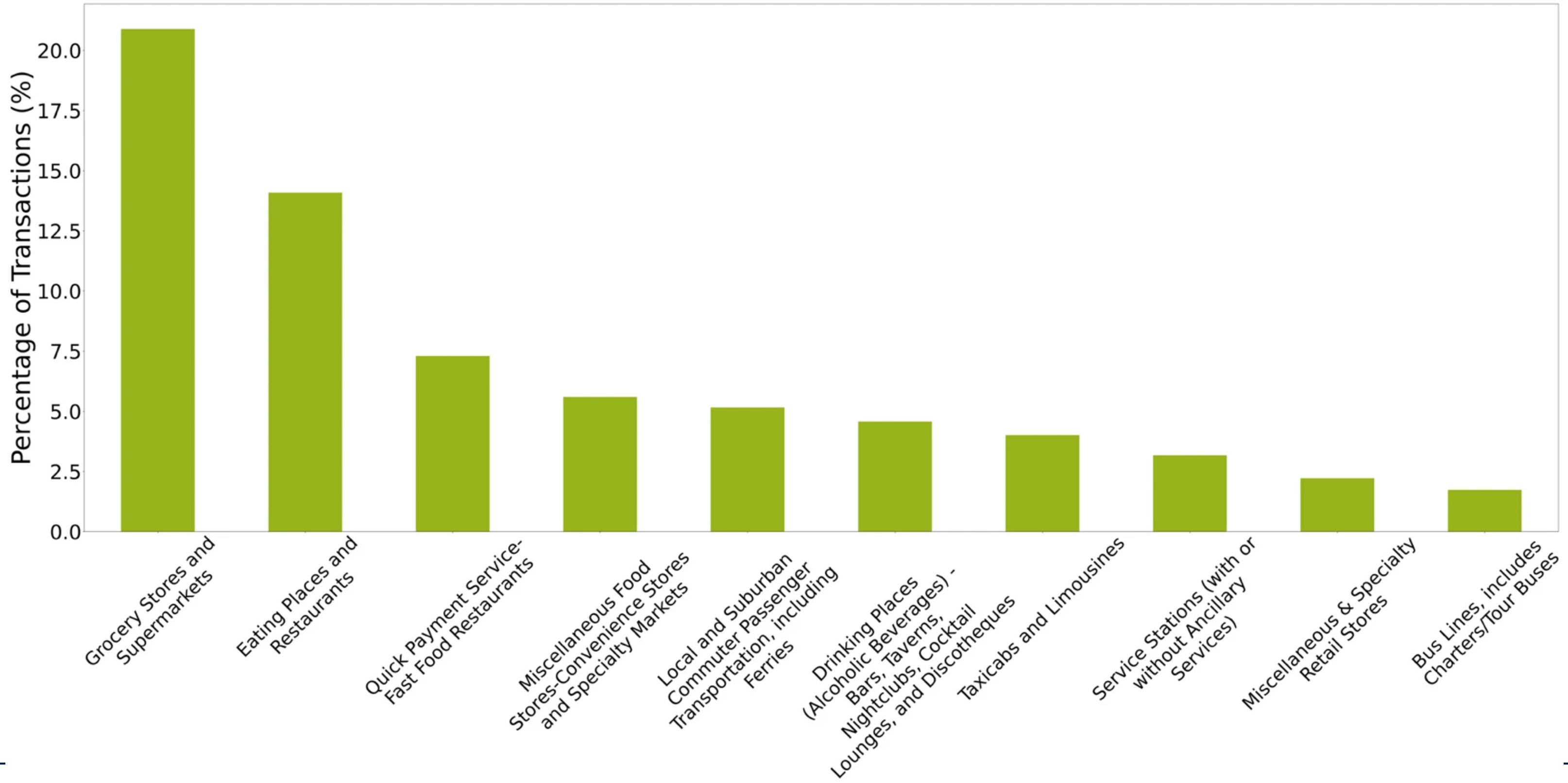
52,496 transactions spanning from 2021 to 2023 from 1,362 customers based in the UK

Data Overview

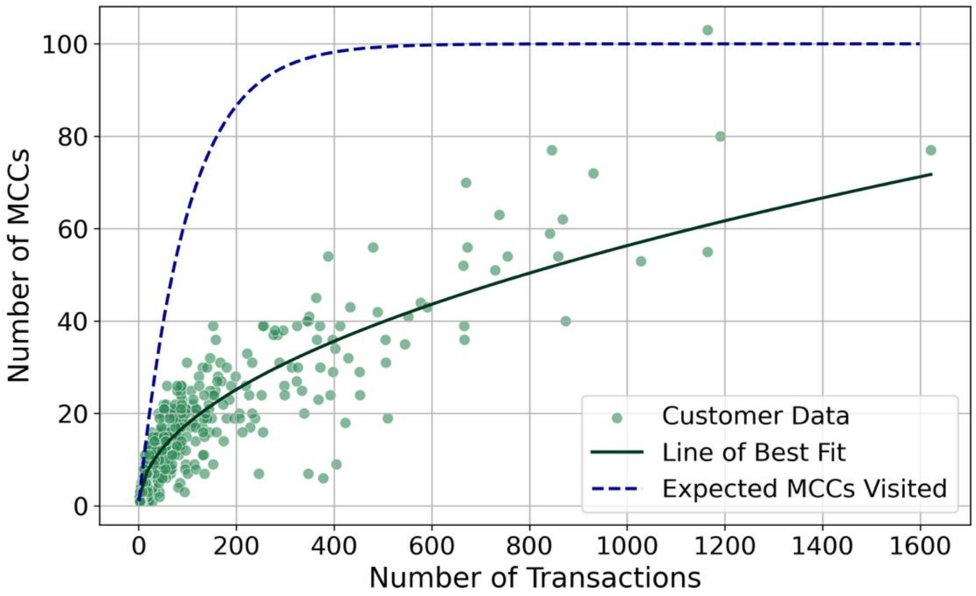
Transaction metrics

Transaction ID	Customer ID	Time of Transaction	Merchant Category Code (MCC)	Amount Spent (GBP)	CO2 Emissions (grams)
trans_0001	cus_num_01	2024-08-29T15:13:22.807Z	Grocery Stores and Supermarkets	17.6	6705.6

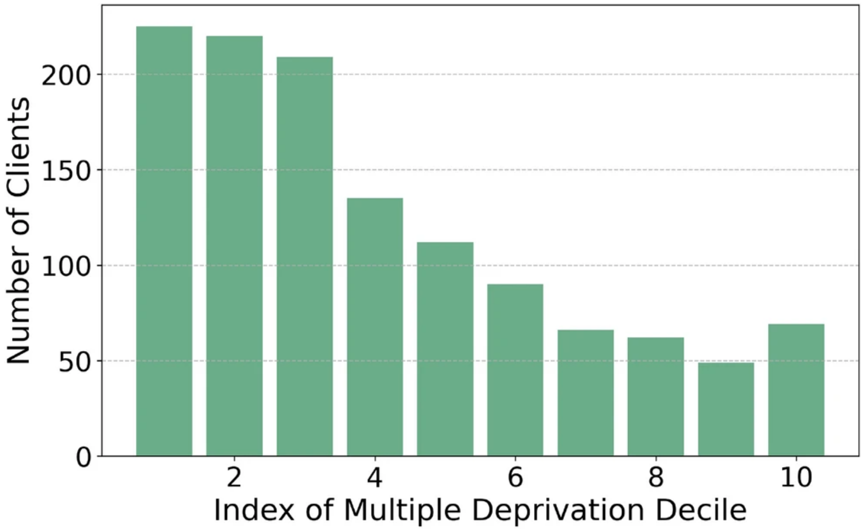
We focus on customers with more than 50 transactions across at least 10 different MCCs, which leaves with 272 customers.


Percentage of Transactions (%)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Grocery Stores and Supermarkets
Eating Places and Restaurants
Quick Payment Service-Fast Food Restaurants
Miscellaneous Food Stores-Convenience Stores and Specialty Markets
Local and Suburban Commuter Passenger Transportation, including Ferries
Drinking Places (Alcoholic Beverages) - Bars, Taverns, Nightclubs, Cocktail Lounges, and Discotheques
Taxicabs and Limousines
Service Stations (with or without Ancillary Services)
Miscellaneous & Specialty Retail Stores
Bus Lines, includes Charters/Tour Buses

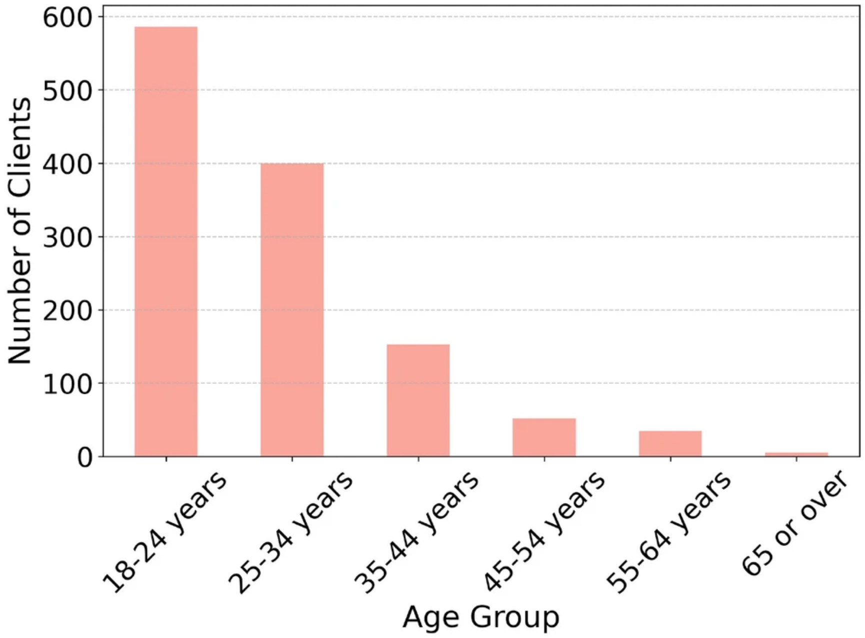
Expected vs observed MCC counts



$$\text{Number of MCC} = n\left(\frac{n^k - (n-1)^k}{n^k}\right)$$


Number of Clients
0
50
100
150
200
2
4
6
8
10
Index of Multiple Deprivation Decile

(a)


Number of Clients
0
100
200
300
400
500
600
18-24 years
25-34 years
35-44 years
45-54 years
55-64 years
65 or over
Age Group

(b)

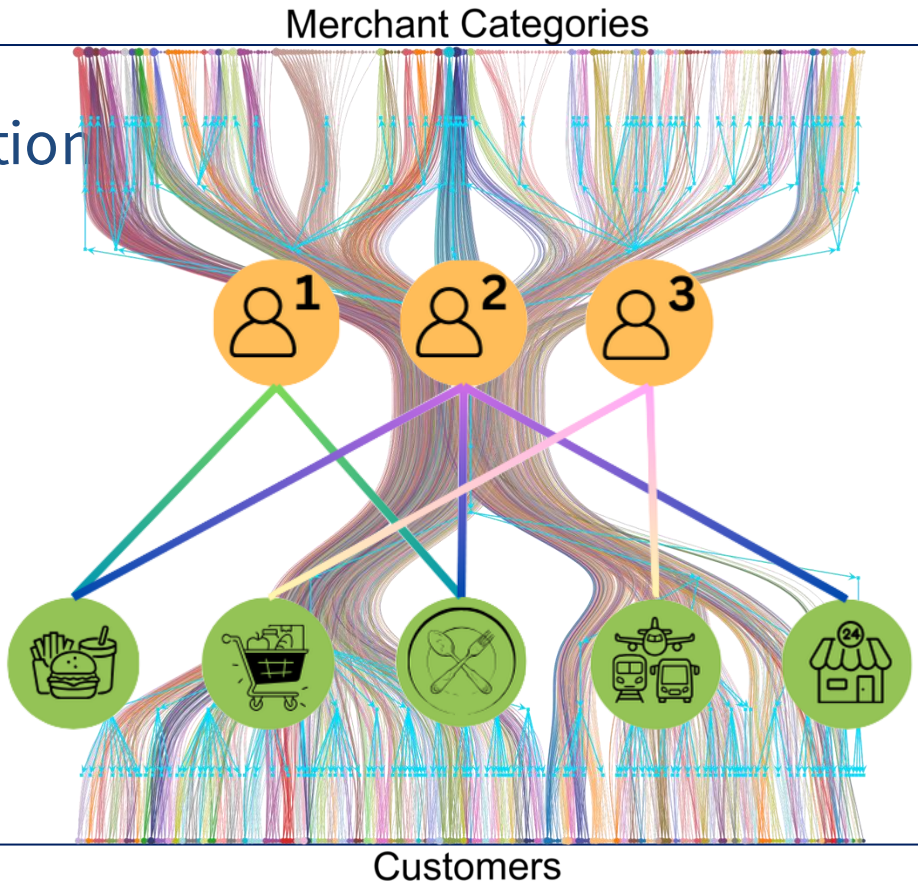
What similarities in customer spending patterns and clusters of consumers can we identified?

Why does it matter?

For financial institutions (banks):
- efficient method to design nudges for groups of customers
- personalised carbon tracking/ incentives/suggestions

For policy:
- identify groups of consumers to be targeted by similar policies

Network Creation



Method: Stochastic Block Modelling (SBM)

- Partition of N nodes into B building blocks, where each node is assigned to a block by $\mathbf{b_i} \in \{1, \ldots, B\}$

- Define the number of edges that connects nodes of two groups r and s by a matrix $\mathbf{e_{rs}}$

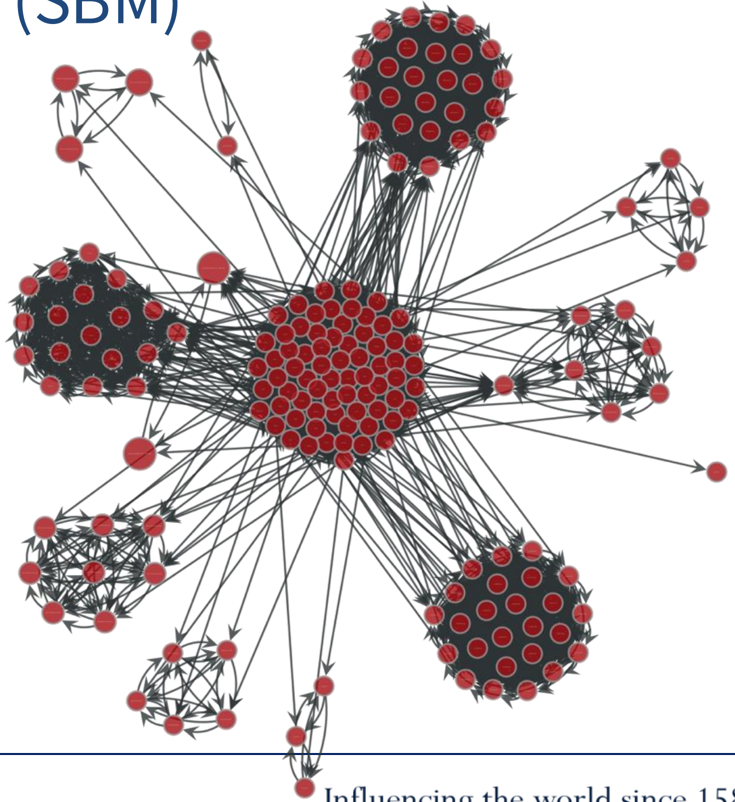
- Construct a generative model based on $\mathbf{b_i}$ and $\mathbf{e_{rs}}$

$$P(\boldsymbol{A}|\boldsymbol{b},\boldsymbol{e}),$$

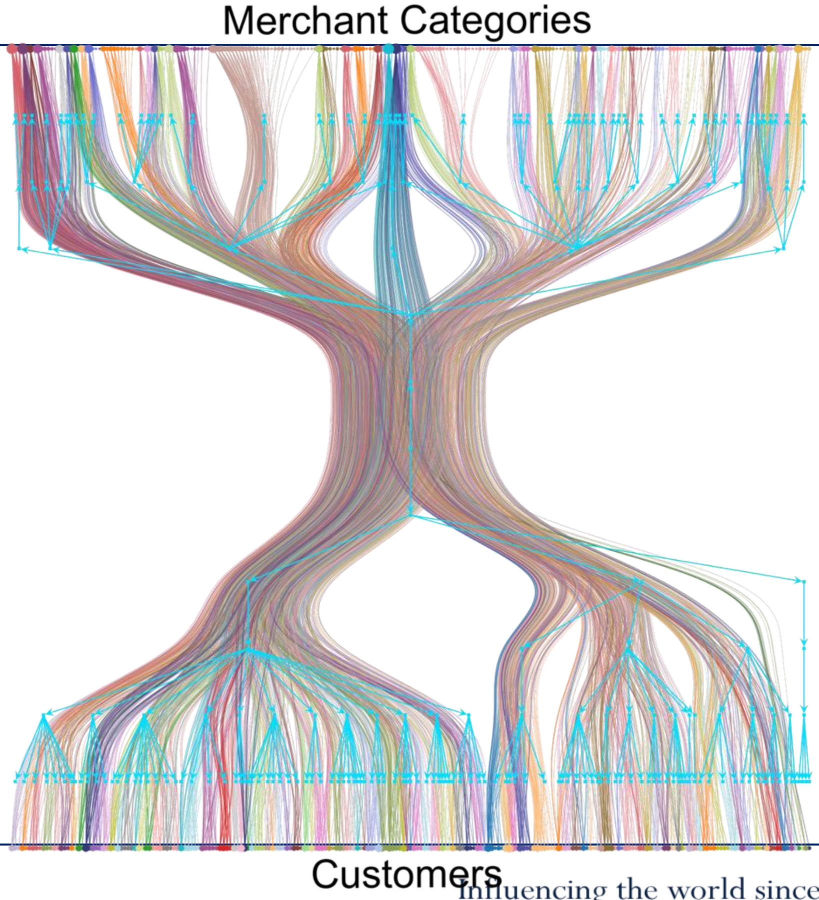
where A is the adjacency matrix of the graph

- We use Bayes formula to get the partitioning

$$P(\boldsymbol{b}|A) = \frac{P(A|\boldsymbol{b})P(\boldsymbol{b})}{P(A)}$$



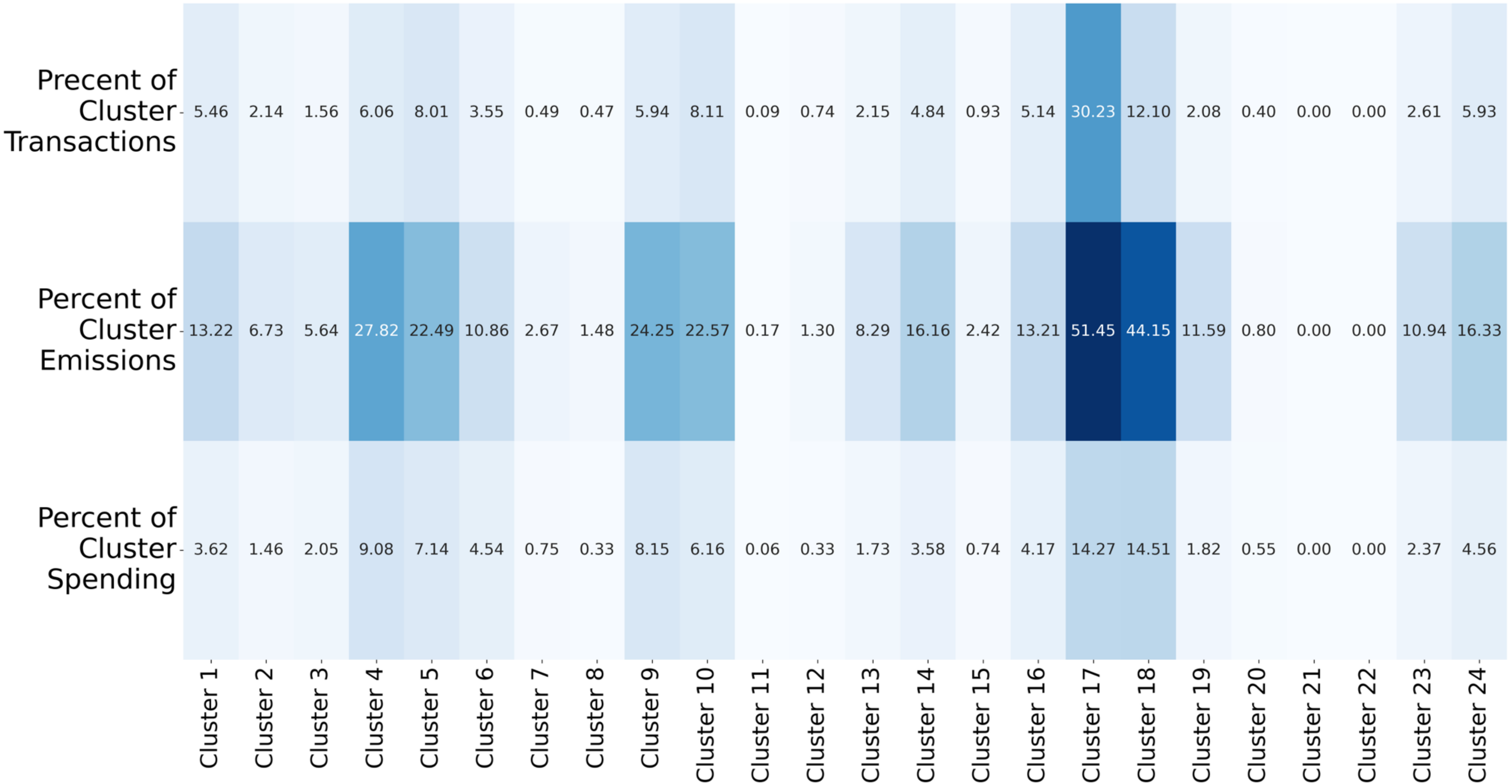
Bipartite Stochastic Block Model



	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12
Grocery Stores and Supermarkets	17.9%	18.8%	8.6%	27.4%	16.0%	10.8%	15.5%	11.6%	18.0%	10.3%	15.7%	5.5%
Service Stations (with or without Ancillary Services)	11.3%	0.8%	55.0%	8.1%	8.2%	26.7%	5.2%	4.1%	16.3%	1.1%	48.3%	35.1%
Taxicabs and Limousines	13.2%	6.7%	5.6%	27.8%	22.5%	10.9%	2.7%	1.5%	24.3%	22.6%	0.2%	1.3%
Eating Places and Restaurants	5.8%	16.3%	2.1%	9.5%	7.3%	7.1%	14.8%	14.8%	7.2%	6.7%	3.6%	2.4%
Passenger Rail (train)	1.7%	9.1%	0.0%	1.5%	3.8%	1.4%	2.9%	7.1%	4.8%	3.0%	0.1%	0.6%
Local and Suburban Commuter Passenger Transportation, including Ferries	5.3%	1.7%	0.6%	2.2%	3.9%	18.0%	6.4%	3.5%	1.5%	10.0%	0.0%	2.1%
Quick Payment Service-Fast Food Restaurants	9.4%	2.3%	2.3%	2.5%	3.7%	2.7%	3.8%	2.6%	5.5%	4.1%	4.0%	2.3%
Drug Stores and Pharmacies	1.7%	5.0%	1.5%	2.4%	3.3%	1.8%	9.3%	4.4%	1.4%	6.0%	0.4%	1.7%
Miscellaneous & Specialty Retail Stores	1.2%	1.2%	0.4%	0.5%	1.6%	1.0%	0.7%	5.1%	1.9%	1.7%	1.2%	19.4%
Miscellaneous Food Stores-Convenience Stores and Specialty Markets	2.0%	2.7%	12.1%	1.8%	4.5%	4.5%	2.0%	6.8%	1.4%	2.6%	1.2%	0.6%

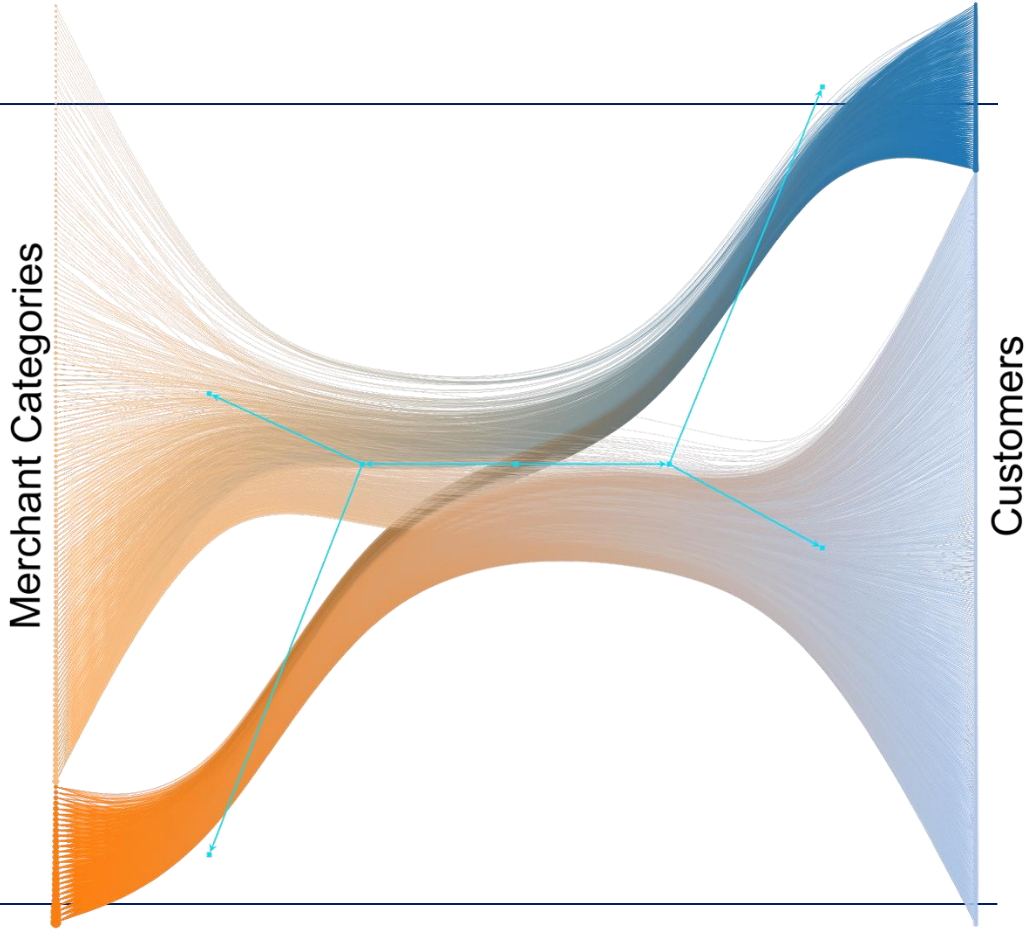
	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18	Cluster 19	Cluster 20	Cluster 21	Cluster 22	Cluster 23	Cluster 24
Grocery Stores and Supermarkets	13.3%	10.6%	16.1%	18.4%	6.9%	1.4%	21.2%	7.2%	5.3%	75.3%	2.0%	9.7%
Service Stations (with or without Ancillary Services)	17.2%	16.6%	15.5%	1.5%	7.4%	1.2%	16.4%	2.3%	4.2%	0.8%	0.0%	12.0%
Taxicabs and Limousines	8.3%	16.2%	2.4%	13.2%	51.4%	44.1%	11.6%	0.8%	0.0%	0.0%	10.9%	16.3%
Eating Places and Restaurants	6.4%	12.9%	11.1%	7.9%	11.7%	8.3%	21.1%	3.1%	0.0%	0.3%	13.2%	10.0%
Passenger Rail (train)	0.4%	0.0%	9.6%	9.6%	0.6%	5.9%	0.0%	2.9%	58.5%	0.0%	35.5%	1.9%
Local and Suburban Commuter Passenger Transportation, including Ferries	1.8%	0.0%	6.9%	8.4%	0.1%	3.0%	0.3%	77.0%	0.6%	0.0%	2.2%	0.7%
Quick Payment Service-Fast Food Restaurants	5.2%	0.9%	2.1%	3.1%	1.2%	1.7%	3.8%	0.7%	1.1%	0.6%	4.2%	4.1%
Drug Stores and Pharmacies	0.6%	0.0%	4.8%	6.5%	1.1%	1.2%	2.2%	1.2%	0.0%	4.4%	6.2%	3.1%
Miscellaneous & Specialty Retail Stores	7.4%	6.0%	0.4%	2.3%	1.8%	1.6%	5.8%	0.0%	0.0%	3.9%	0.2%	0.6%
Miscellaneous Food Stores-Convenience Stores and Specialty Markets	4.1%	2.0%	1.6%	2.9%	2.1%	3.9%	2.2%	0.3%	0.0%	0.3%	1.1%	2.0%

THE UNIVERSITY of EDINBURGH



SBM Modifications

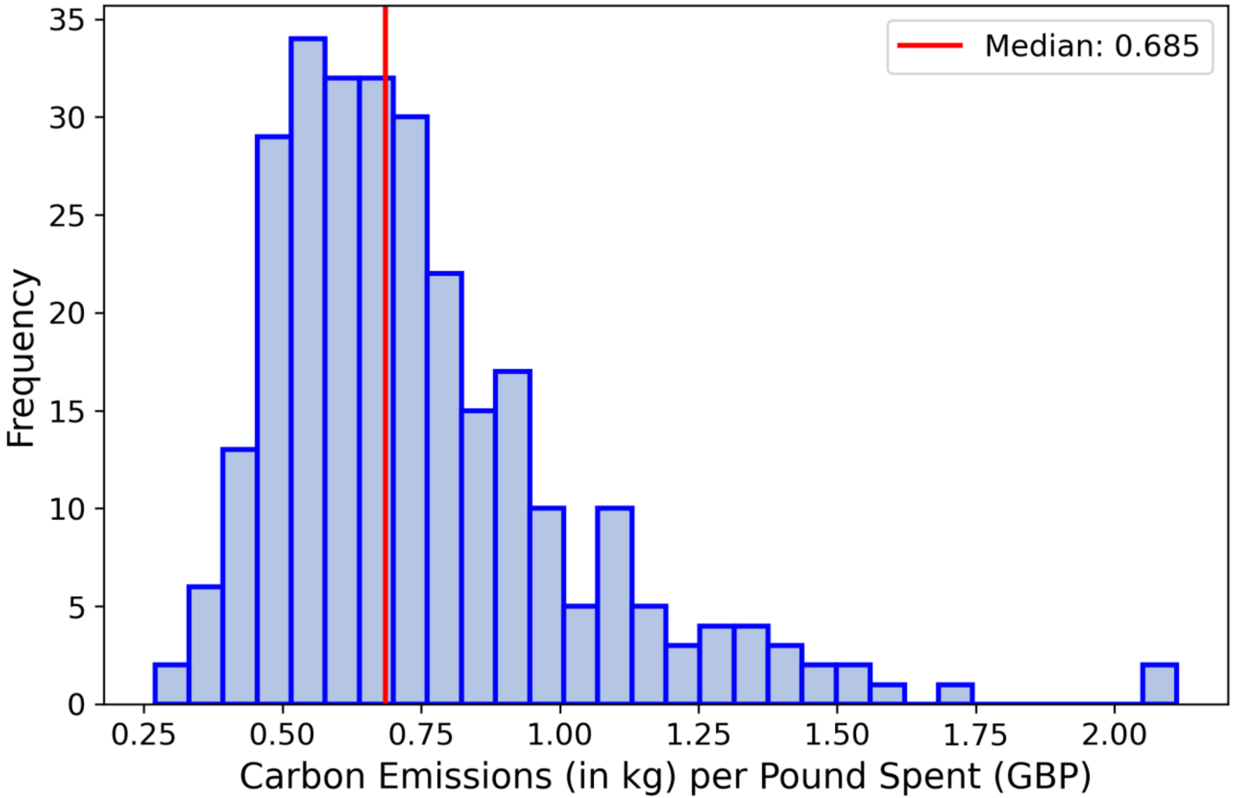
- using COICOP categories instead of MCCs
- introducing number of transactions in a customer-MCC connection as an edge weight
- finding relative customer spending per category and introducing these values as an edge weight.

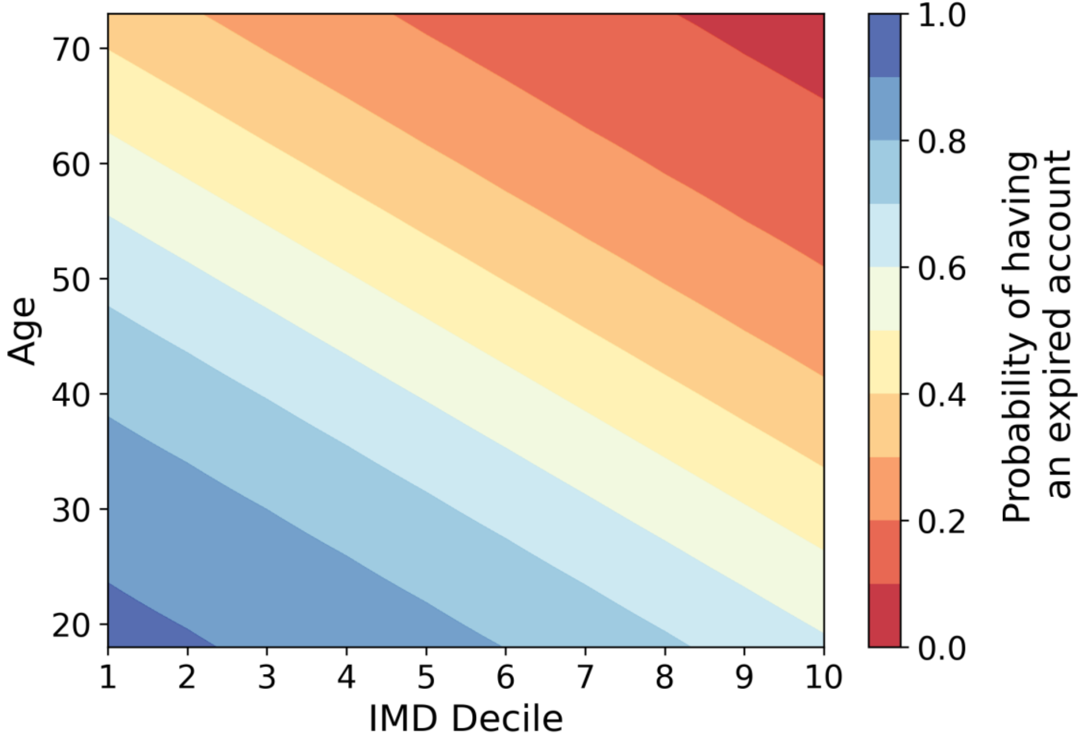


THE UNIVERSITY *of* EDINBURGH

	Cluster 1	Cluster 2	Total Percent of Population Spending	Median	MAD
Grocery Stores and Supermarkets	18.49%	19.63%	18.89%	18.57%	9.28%
Eating Places and Restaurants	17.59%	14.84%	16.63%	12.34%	6.85%
Quick Payment Service-Fast Food Restaurants	4.88%	6.22%	5.35%	5.25%	2.79%
Drinking Places (Alcoholic Beverages) - Bars, Taverns, Nightclubs, Cocktail Lounges, and Discotheques	4.64%	5.86%	5.07%	4.73%	3.45%
Service Stations (with or without Ancillary Services)	3.40%	5.50%	4.13%	4.45%	3.81%
Taxicabs and Limousines	2.78%	6.03%	3.91%	4.29%	3.53%
Miscellaneous Food Stores-Convenience Stores and Specialty Markets	3.32%	3.63%	3.42%	2.39%	1.93%
Miscellaneous & Specialty Retail Stores	2.48%	4.49%	3.18%	1.92%	1.55%
Men's and Women's Clothing Stores	2.34%	1.90%	2.19%	1.85%	1.33%
Department Stores	1.58%	2.38%	1.86%	1.75%	1.51%

0.175 0.150 0.125 0.100 0.075 0.050 0.025


Median: 0.685
Frequency
0
5
10
15
20
25
30
35
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Carbon Emissions (in kg) per Pound Spent (GBP)


70
60
50
40
30
20
Age
1
2
3
4
5
6
7
8
9
10
IMD Decile
1.0
0.8
0.6
0.4
0.2
0.0
Probability of having an expired account

Limitations?

- Data hard to access, NDAs
- Sample size
- Uneven distribution of customers across study area (UK)
- Incomplete spending profiles (for customers with multiple cards or cash users)
- Calculation of emissions per purchase
  * MCCs can be imprecise, e.g. grocery stores MCC
  *different prices across the UK (Edinburgh vs London)
- Limited policy applications (data hard to access)

Alternative Data Source

1

Identify spending/emissions patterns from micro-level data (transactions)

- interventions on individual level
- changes within companies

2

Observe how these patterns connect to macro-level data (LCFS, regions, LSOA)

- interventions on wider policy level
- changes within industries/regions

Living Cost and Food Survey

The LCFS is a continuous, nationally representative expenditure survey conducted by the Office for National Statistics

- Data hard to access, NDAs
- Sample size
- Uneven distribution of customers across study area (UK)
- Incomplete spending profiles (for customers with multiple cards or cash users)
- Calculation of emissions per purchase
  * MCCs can be imprecise, e.g. grocery stores MCC
  *different prices across the UK (Edinburgh vs London)
- Limited policy applications (data hard to access)

Conclusions:

- Expenditure data rocks!
  * hundreds of expenditure categories
  * thousands (millions) of consumers
  * hundreds of socio-economic classifications for consumers (including temporal and spatial aspects)
  * connects directly to human behaviour!
- Network science is a great tool for extracting information from these large transaction datasets
- Quantifying emissions from consumer spending is not as easy on small scales
- Understanding spending behaviour can help target consumers effectively

Questions?
(about this research or PhD life in general)