MANCHESTER 1824
The University of Manchester

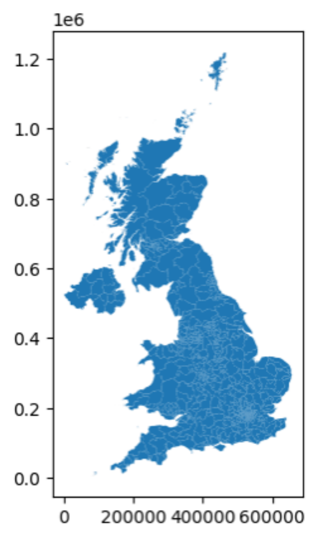
12th June 2025

The Shape of Britain

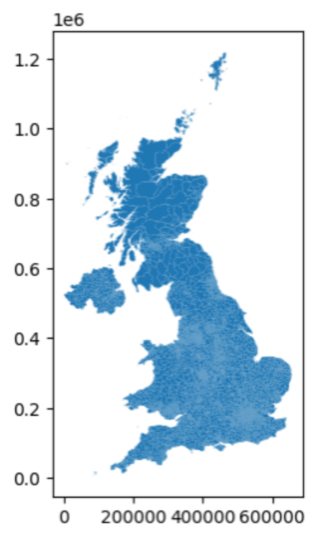
Topologically Mapping the Census

Dr Simon Rudkin
Senior Lecturer in Data Science

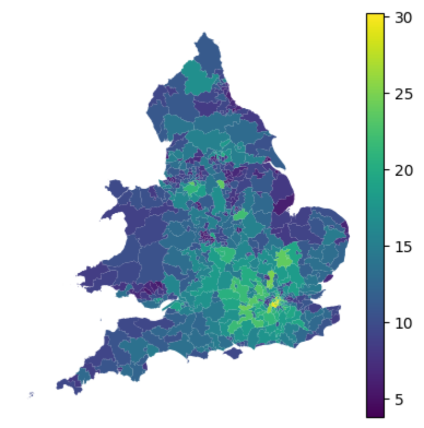
Shape of Britain*



- Geographic shape of Britain is understood
- The shapefiles provided by ONS include Scotland and Northern Ireland
- Examples studied will only be England and Wales



Shape of Britain* 2

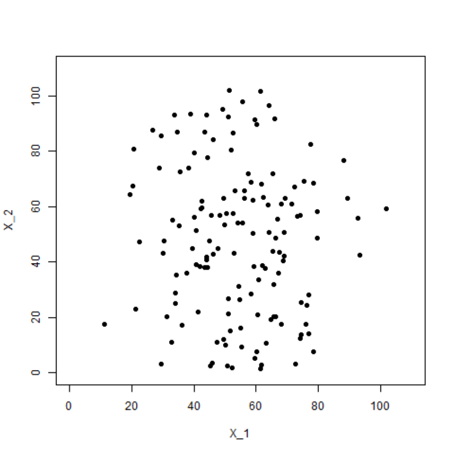


- Maps can be coloured to understand geographic spread
- Example here is proportion of residents in NSSEC 1
- Yellows are the higher values - concentrate around London
- Lighter colouration around Manchester is also seen

MANCHESTER 1824
The University of Manchester

What is the Shape of Data?

The Datasaurus (Matejka and Fitzmaurice, 2017)



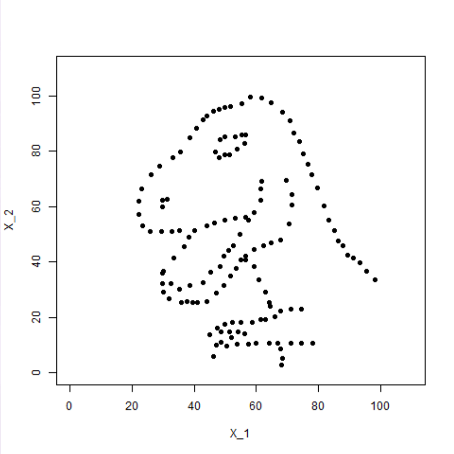
Imagine a scatterplot for the following data:

Variable	Mean	Std Dev	Correlation
X_1	54.26	16.77	-0.064
X_2	47.83	26.94	

Most will think of the Gaussian cloud, but no information in the table says that the data is a Gaussian cloud

MANCHESTER 1824
The University of Manchester

The Datasaurus (Matejka and Fitzmaurice, 2017)



Imagine a scatterplot for the following data:

Variable	Mean	Std Dev	Correlation
X_1	54.26	16.77	-0.064
X_2	47.83	26.94	

Most will think of the Gaussian cloud, but no information in the table says that the data is a Gaussian cloud

MANCHESTER 1824
The University of Manchester

Topological Data Analysis (TDA)

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 46, Number 2, April 2009, Pages 255–308
S 0273-0979(09)01249-X
Article electronically published on January 29, 2009

TOPOLOGY AND DATA

GUNNAR CARLSSON

1. INTRODUCTION

An important feature of modern science and engineering is that data of various kinds is being produced at an unprecedented rate. This is so in part because of new experimental methods, and in part because of the increase in the availability of high powered computing technology. It is also clear that the *nature* of the data we are obtaining is significantly different. For example, it is now often the case that we are given data in the form of very long vectors, where all but a few of the coordinates turn out to be irrelevant to the questions of interest, and further that we don't necessarily know which coordinates are the interesting ones. A related fact is that the data is often very high-dimensional, which severely restricts our ability to visualize it. The data obtained is also often much noisier than in the

Data has Shape and Shape has Meaning

- Seminal work Carlsson (2009)
- Mapper algorithms after Singh et al. (2007) - Stability?
- Applications in dynamic systems, visualization and modelling

What is meant by Mapping Data?

Point Clouds

Point Cloud - Set of data points plotted with 1 axis per dimension

- Simplest point cloud is scatter plot
- We are trained to understand scatterplots
- See Cleveland and McGill (1984a,b) for reviews of the "many faces of a scatterplot"

Key Motivating Arguments for Ball Mapper

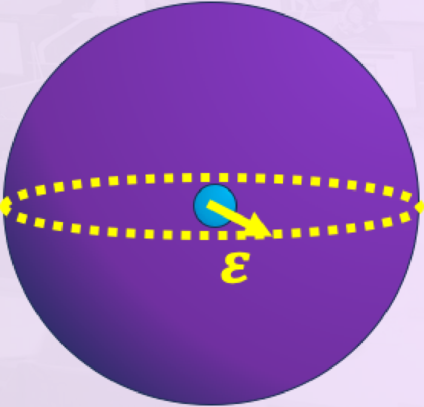
Visualising data is an essential phase of the modelling process (Anscombe, 1973)

Humans cannot see in multiple dimensions - dimension reduction

Mapping helps rationalise space - look to map our data

Summary statistics (1st and 2nd moments) are insufficient (Matejka and Fitzmaurice, 2017)

Topological Data Analysis Ball Mapper (Dłotko, 2019)



- TDABM "covers" data with "balls" of fixed radius ϵ (single parameter)
- TDABM has advantage of stability over mapper (Dłotko, 2019; Carriere and Oudot, 2018)
- Benefits of visualisation to understand Brexit (Rudkin et al., 2024)
- Capturing trajectories in regional development time series (Rudkin and Webber, 2023)
- Culture and migration (Tubadji and Rudkin, 2025)
- Finance (Qiu et al., 2020), Dłotko et al. (2024)
- Points in ball are "similar" in all dimensions

Core Elements of TDABM

Axis Variables - Must be continuous with sufficiently many observations. Would be plotted using a scatterplot

Outcome Variable - Must be available for each data point

Colouration Function - How should outcome values for each point be combined within the ball? (default is geometric mean)

TDABM Setup

- Dataset X - N observations on K variables, x_1, x_2, ..., x_k, ..., x_K
- Point p_i, $i \in [1, N]$ has co-ordinates x_{1i}, x_{2i}, ..., x_{ki}, ..., x_{Ki}
- Cover $B(X)$ - p_i is assigned to at least one ball $B_b(l_{kb}, \epsilon)$.
- l_k is co-ordinate of point at centre of ball on axis k
- l iteratively selected randomly from uncovered points
- ϵ - ball radius - only choice parameter
- V shown as discs to represent balls - E connect pairs where $B(p_1, \epsilon) \cap B(p_2, \epsilon) \neq \emptyset$
- Produce abstract two-dimensional visualization of $B(X, \epsilon)$, $G(V, E)$

MANCHESTER 1824
The University of Manchester

An Economic Topology of Brexit

REGIONAL STUDIES
https://doi.org/10.1080/00343404.2023.2204123

Routledge Taylor & Francis Group | Regional Studies Association

OPEN ACCESS | Check for updates

An economic topology of the Brexit vote

Simon Rudkin[a*], Lucy Barros[a], Paweł Dłotko[b] and Wanling Qiu[c**]

ABSTRACT
A desire to understand the UK voting to leave the European Union continually attracts attention. We generate a multidimensional map of the economic geography of Brexit voting at the regional level, visualising hitherto unidentified insights into the regional manifestation of leave voting. While we find broad patterns consistent with national heterogeneities and the geographies of discontent, we also demonstrate support for Brexit locates in a far more homogenous set of regions than support for remaining in the European Union. Our conclusions apply at the constituency and local authority levels and are robust to inclusion of additional cultural and economic regional characteristics.

KEYWORDS
topological data analysis; geographies of discontent; Brexit; local demographics; interaction effects; voting behaviour

JEL C1, D72, N44
HISTORY Received 25 September 2021; in revised form 4 April 2023

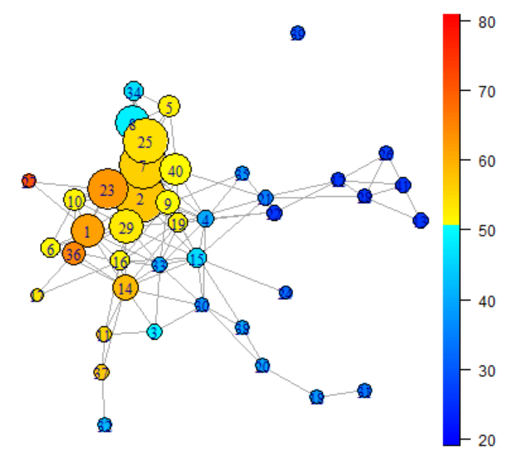
1. INTRODUCTION

Evaluation of the decision of the UK to leave the European Union (EU), 'Brexit',[1] has been rooted in geographies of discontent (Dijkstra et al., 2020; McCann, 2020) and the notion that regions were being 'left behind' (Heath & Goodwin, 2017). Behind these concepts sit mixtures of demographic characteristics such as education, home ownership, social status and age which have been variously associated with Leave voting. This paper takes the multidimensional dataset high-

What follows is based upon observed proportions of key demographic characteristics within parliamentary constituencies. Motivation for studying regional aggregation lies in the fact that voting behaviour cannot be divorced from local context; the characteristics of individuals interact with the broader features of their neighbours and neighbourhood (e.g., Johnston et al., 2004; Abreu & Öner, 2020). Constituency-level aggregation allows linkage directly to parliamentary election results that comment on voter party-political allegiance. Direct parallels are

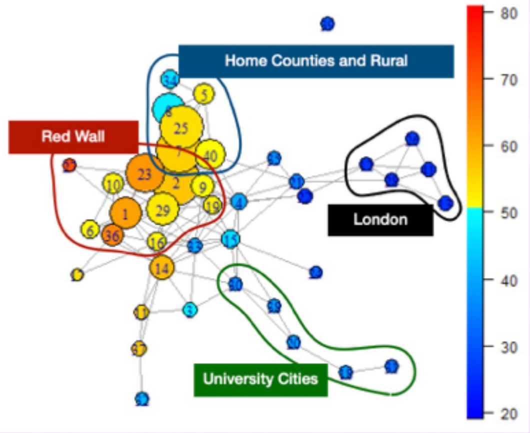
- Joint work with Dr Pawel Dlotko (Discouri Centre in Topological Data Analysis), Dr Wanling Rudkin (University of Exeter) and Dr Lucy Barros (Cardiff University)
- Also appears as a blog on "UK in a Changing Europe" in 2019

An Economic Topology of Brexit 2



- Leave percentage by constituency
- Concentration of Leave versus Remain

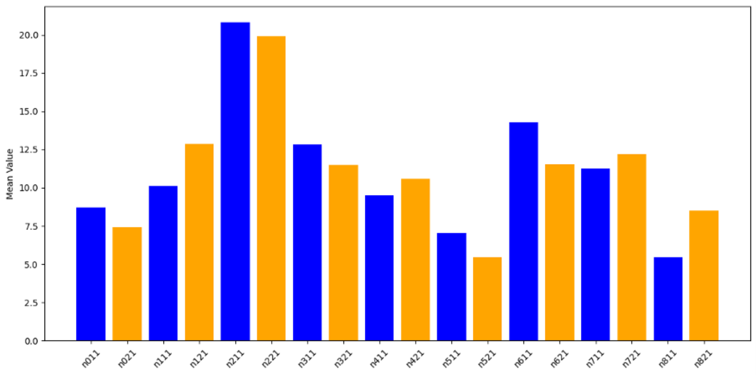
An Economic Topology of Brexit 2



- Leave percentage by constituency
- Concentration of Leave versus Remain
- Annotation from Otway and Rudkin (2024)
- London constituencies in right arm
- University cities group together
- Groups of constituencies form in space
- All balls are fully explorable

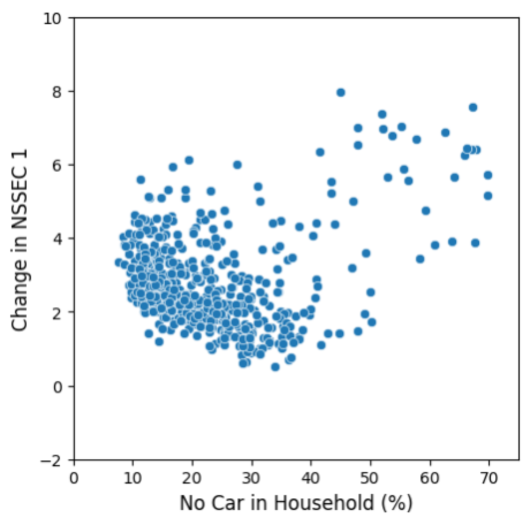
MANCHESTER 1824
The University of Manchester

Changes in NSSEC



Some levels have been falling (e.g 0), others rising (e.g. 1)

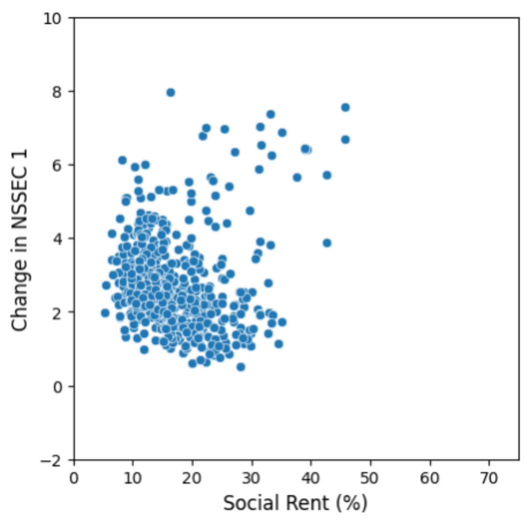
Changes in NSSEC 2



- There is a negative correlation between no car and having more NSSEC 1 residents than 2011
- Group of points to the top right which go against relationship

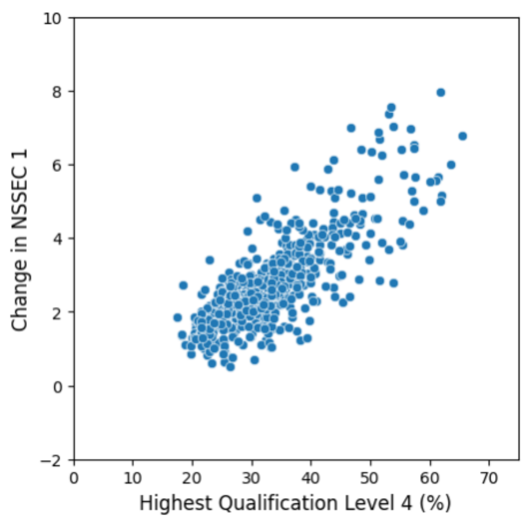
The University of Manchester

Changes in NSSEC 2



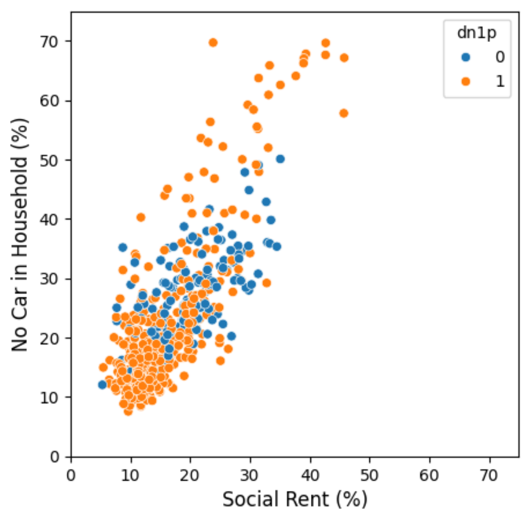
- There is a negative correlation between no car and having more NSSEC 1 residents than 2011
- Group of points to the top right which go against relationship
- Similar story for social rent

Changes in NSSEC 2



- There is a negative correlation between no car and having more NSSEC 1 residents than 2011
- Group of points to the top right which go against relationship
- Similar story for social rent
- Proportion with Level 4 qualifications is proportional

Changes in NSSEC 2



- There is a negative correlation between no car and having more NSSEC 1 residents than 2011
- Group of points to the top right which go against relationship
- Similar story for social rent
- Proportion with Level 4 qualifications is proportional
- Colouring shows where the rises in NSSEC 1 have been greatest

The University of Manchester

Changes in NSSEC 2

- There is a negative correlation between no car and having more NSSEC 1 residents than 2011
- Group of points to the top right which go against relationship
- Similar story for social rent
- Proportion with Level 4 qualifications is proportional
- Colouring shows where the rises in NSSEC 1 have been greatest

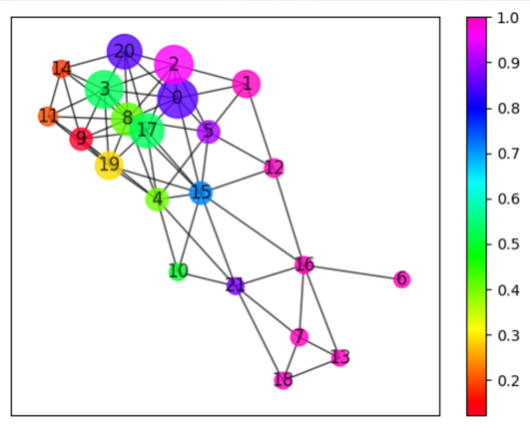
Changes in NSSEC 3

Variable	sf	Mean	s.d.	Min	Max
No Car in Household	car021	22.85	11.55	7.563	69.72
Full Time Employment	efull21	33.97	4.35	21.60	51.70
Health Very Good	hvg21	40.55	4.65	29.38	58.63
High Qualifications	q421	33.30	8.895	17.50	65.49
Social Rental	soc21	16.87	6.763	5.333	47.78
Dummy for Increase in NSSEC 1	dn1p	0.728	0.446	0	1

Variables used in basic example. Dummy is 1 if percentage in NSSEC1 has risen by 2 percentage points

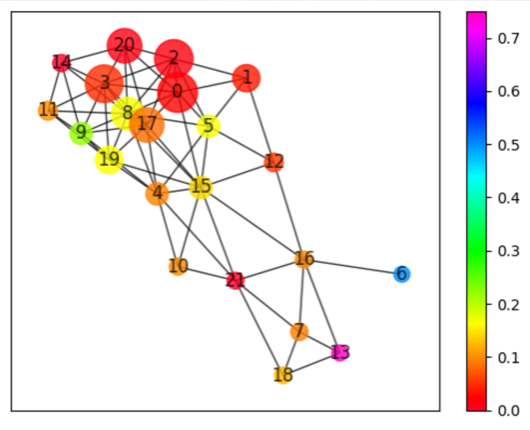
MANCHESTER 1824
The University of Manchester

Changes in NSSEC 2



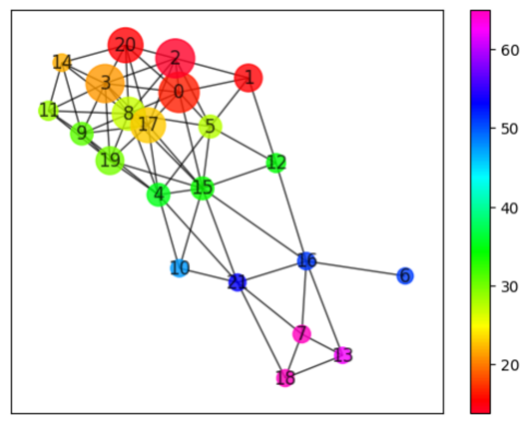
- Proportion with changes in NSSEC 1 above 2 percentage points

Changes in NSSEC 2



- Proportion with changes in NSSEC 1 above 2 percentage points
- Proportion of NSSEC Other increasing

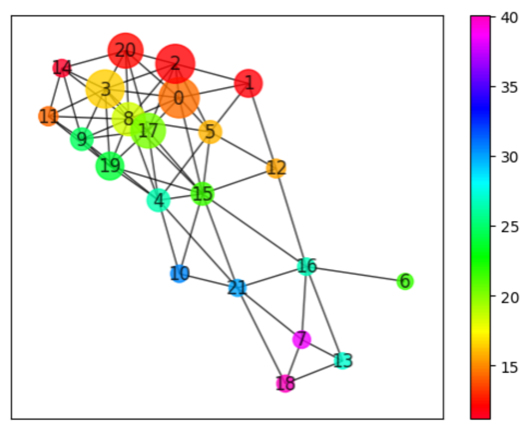
Changes in NSSEC 2



- Proportion with changes in NSSEC 1 above 2 percentage points
- Proportion of NSSEC Other increasing
- Households without cars highest to the lower right

MANCHESTER 1824
The University of Manchester

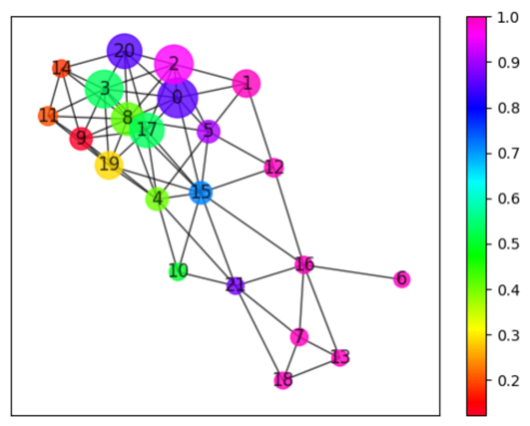
Changes in NSSEC 2



- Proportion with changes in NSSEC 1 above 2 percentage points
- Proportion of NSSEC Other increasing
- Households without cars highest to the lower right
- Similar colouration for the social rent but see balls to the lower middle with high social rent but lower proportions without a car

MANCHESTER 1824
The University of Manchester

Changes in NSSEC 3

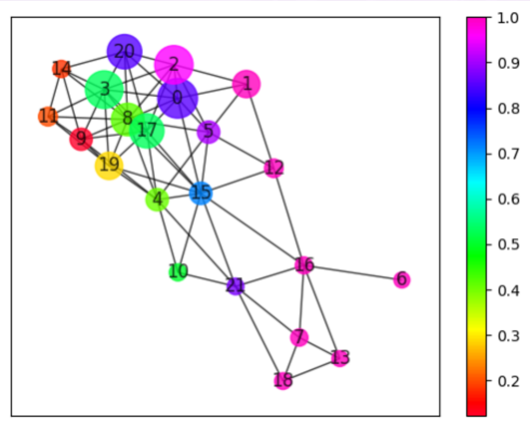


- Ball 6 is Battersea, Chelsea and Fulham, Streatham and Tooting
- Ball 13 is Cities of London and Westminster, Hampstead and Kilburn, Kensington, Westminster North
- Ball 1 contains 79 constituencies including Altrincham and Sale West, Beckenham, Woking, Wycombe and Gower

MANCHESTER 1824
The University of Manchester

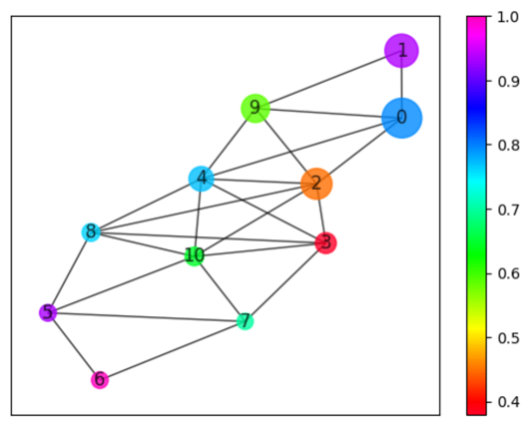
Are the inferences from TDABM Consistent?

Changes in NSSEC 4



- Proportion with changes in NSSEC 1 above 2 percentage points

Changes in NSSEC 4

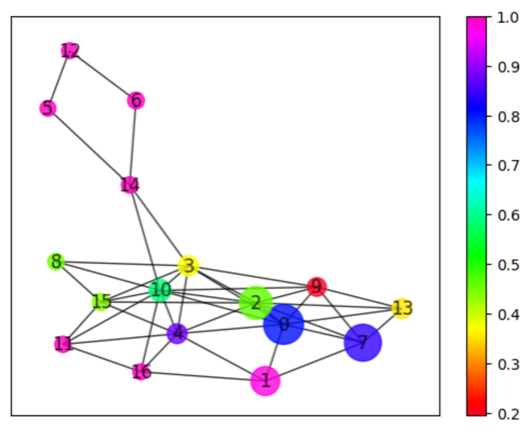


- Proportion with changes in NSSEC 1 above 2 percentage points
- Changing the axes to have part-time work, private ownership, highest qualifications and good health

MANCHESTER 1824

The University of Manchester

Changes in NSSEC 4



- Proportion with changes in NSSEC 1 above 2 percentage points
- Changing the axes to have part-time work, private ownership, highest qualifications and good health
- Expanded dataset with all previously used together with students and private rental
- Consistent inference is reached

MANCHESTER 1824
The University of Manchester

Further Work

Resilience - Does regional similarity assist in the preparedness of areas?

Politics - How does the socio-demographic mapping of data align with election results (Otway and Rudkin, 2024)

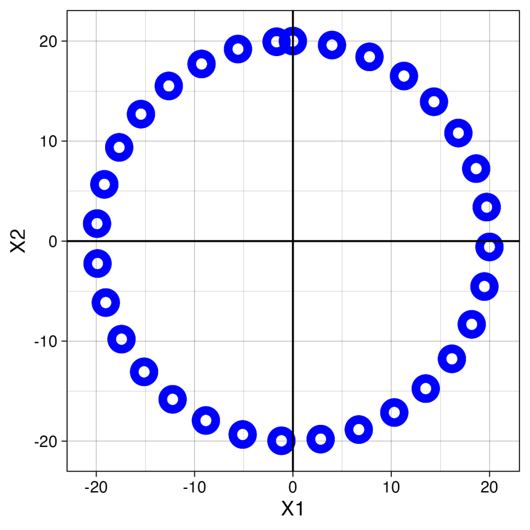
Unpaid Care - Patterns of unpaid care provision on the socio-demographic space reveal clear concentrations

Education Inequality at the Ward Level - Coursework on this years DATA70302

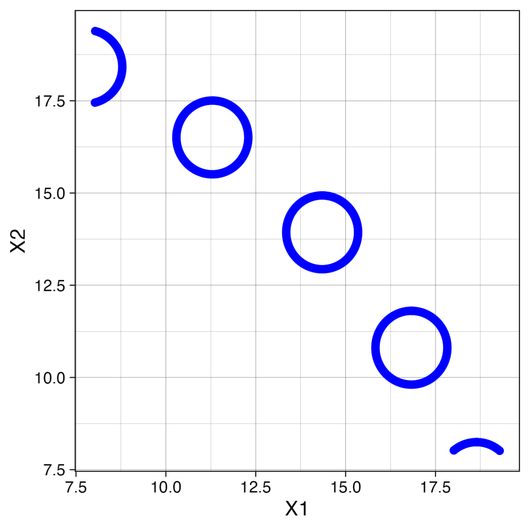
Summary

- Socio-economic phenomenon are the conflation of many factors
- Topological Data Analysis Ball Mapper offers new chance to see data
- Inference from the BM plots and summaries of the balls
- Each ball represents data - can take more characteristics from that data to discuss
- Interested in whole picture - BM shows the picture and enables understanding social phenomena and policy decisions
- Mapping Census data reveals patterns and directs further work

Optimum Radius? Example Datasets

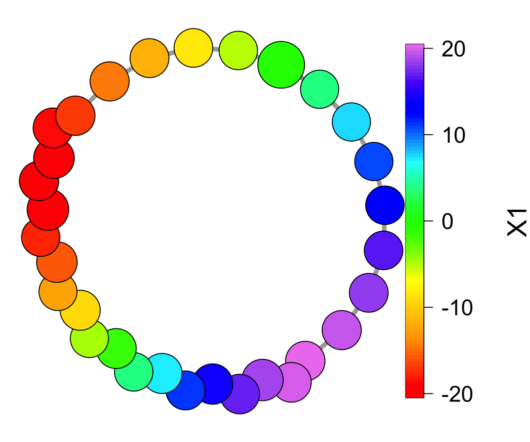


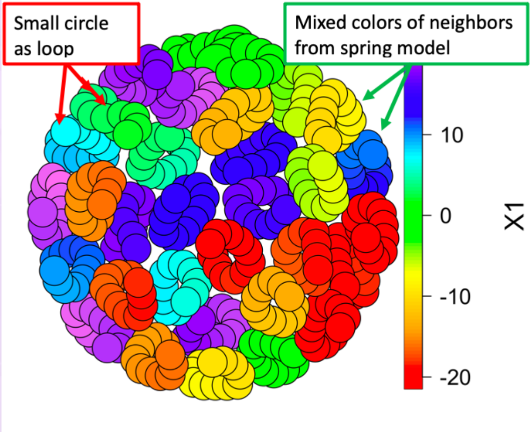
(a) Full Dataset



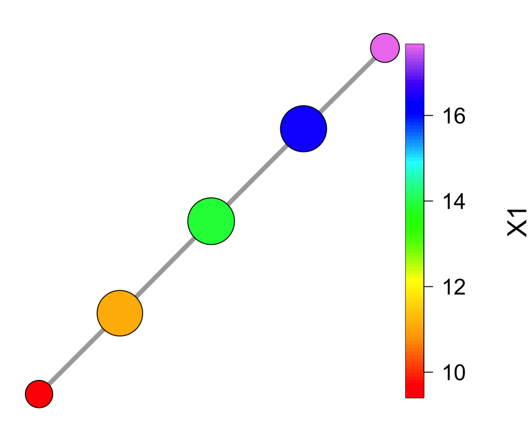
(b) Zoomed Section

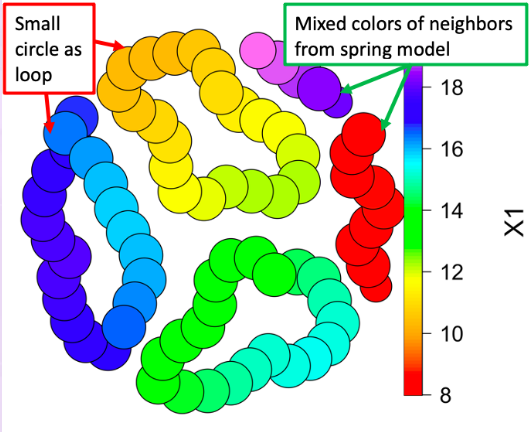
Optimum Radius ? (Full Dataset)





Optimum Radius ? (Zoomed Section)





MANCHESTER 1824
The University of Manchester

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1):17–21.

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308.

Carriere, M. and Oudot, S. (2018). Structure and stability of the one-dimensional mapper. *Foundations of Computational Mathematics*, 18(6):1333–1396.

Cleveland, W. S. and McGill, R. (1984a). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554.

Cleveland, W. S. and McGill, R. (1984b). The many faces of a scatterplot. *Journal of the American Statistical Association*, 79(388):807–822.

Dłotko, P. (2019). Ball mapper: A shape summary for topological data analysis. *arXiv preprint arXiv:1901.07410*.

Dłotko, P., Qiu, W., and Rudkin, S. (2024). Financial ratios and stock returns reappraised through a topological data analysis lens. *The European Journal of Finance*, pages 1–25.

Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 1290–1294.

Otway, A. and Rudkin, S. (2024). The shape of left-behindedness: The 2019 uk general election and britain's changing electoral geography. *Available at SSRN 4877434*.

Qiu, W., Rudkin, S., and Dłotko, P. (2020). Refining understanding of corporate failure through a topological data analysis mapping of altman's z-score model. *Expert Systems with Applications*, page 113475.

Rudkin, S., Barros, L., Dłotko, P., and Qiu, W. (2024). An economic topology of the brexit vote. *Regional Studies*, 58(3):601–618.

Rudkin, S. and Webber, D. J. (2023). Regional growth paths and regional resilience. *Available at SSRN 4333276*.

Singh, G., Mémoli, F., and Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3d object recognition. *SPBG*, 91:100.

Tubadji, A. and Rudkin, S. (2025). Cultural gravity and redistribution of growth through migration: Cohesion lessons from spatial econometrics and topological data analysis. *Papers in Regional Science*, 104(1):100064.