An Economic Topology of the Brexit vote

Paweł Dłotko*[1], Simon Rudkin †[2], and Wanling Qiu‡[3]

[1]Mathematics Department, Swansea University, United Kingdom
[2]Economics Department, Swansea University, United Kingdom
[3]School of Management, University of Liverpool, United Kingdom

September 8, 2019

Abstract

A quest to understand the decision of the UK to leave the European Union, Brexit, in the referendum of June 2016 has occupied academics, the media and politicians alike. As the debate about what the future relationship will look like rages, the referendum is given renewed importance as an indicator of the likely success, or otherwise, of any forward plans. Topological data analysis offers an ability to faithfully extract maximal information from complex multi-dimensional datasets of the type that have been gathered on Brexit voting. Within the complexity it is shown that support for Leave drew from a far more similar demographic than Remain. Obtaining votes from this concise set was more straightforward for Leave campaigners than was Remain's task of mobilising a diverse group to oppose Brexit. Broad patterns are consistent with extant empirical work, but the strength of TDA Ball Mapper means that evidence is offered to enrich the narrative on immobility, and being "left-behind" by EU membership, that could not be found before. A detailed understanding emerges which comments robustly on why Britain voted as it did. A start point for the policy development that must follow is given.

Keywords: Topological Data Analysis, Voting Behaviour, Brexit, Local Demographics, Interaction Effects.

1 Introduction

June 23rd 2016 saw the United Kingdom vote by a margin of 52% to 48% to leave the European Union. Britain's decision to exit the EU has become known as "Brexit". Brexit's consequences are still yet to be fully understood three years on from that pivotal vote. As politicians continue to shape the form that withdrawal, if and when it happens, will take there is natural uncertainty about what will become of the UK economy. Many theorems for the result have been posited, including notions of the "left behind" signalling to the political elite and a growth of "Euroscepticism". Both are hypothesised to be fuelled by austerity and the particular challenges emerging from the Global Financial Crisis. These theories have many endogeneities and overlaps, creating a plethora of statistical challenges that evidence remains limited. However, what is more clear is that there is value in understanding why the vote went the way that it did, which local characteristics were most associated with Brexit? How did those characteristics vary across regions? How did multiple circumstances come together to produce such a seemingly irrational fundamental change in the UK position?[1]

*Full Address: Mathematics Department, College of Science, Swansea University, Bay Campus, Swansea, SA1 8EN, United Kingdom. Email:p.t.dlotko@swansea.ac.uk.

†**Corresponding Author**. Full Address: Economics Department, School of Management, Swansea University, Bay Campus, Swansea, SA1 8EN, United Kingdom. Tel: +44 (0)1792 606325 Email:s.t.rudkin@swansea.ac.uk

‡Full Address: Accounting and Finance Subject Group, School of Management, University of Liverpool, 20 Chatham Street, Liverpool, L69 7ZH, United Kingdom. Email:wanling.qiu@liverpool.ac.uk

[1]Before proceeding it is important to define a few key terms as they will be used in this paper. In all that follows region is used to define the eleven subdivisions of Great Britain used within the analysis. Constituencies are the areas defined at the time of the 2016 referendum such that each constituency has one member of the Houses of Parliament elected to represent

This paper contributes an answer to these questions from a data driven perspective, unlocking information content from voting pattern, and demographic, data at the UK Parliamentary Constituency level. Focusing in this way speaks to the geographical discussion begun in Harris and Charlton (2016), whilst recognising all of the elements of the demographic explorations in Becker et al. (2017) and others. Whilst the decision to hold the referendum was born of a Conservative Prime Minister the political will to do so was held by the United Kingdom Independence Party, Liberal Democrats and Green Party. The former was a single issue party whose successes in past elections had brought the question of EU membership up the agenda, whilst that latter two are keen advocates of stronger membership who saw the referendum as a chance to cement the UK will to integrate more (Sampson, 2017). Constituencies have the big advantage over other aggregations that they can be linked directly to parliamentary election results to comment on voter allegiance. Contributions are thus made on the link between electoral outcomes and the referendum result, as well as those with local socio-economic conditions. It is shown that in all dimensions the Brexit voting constituencies are concentrated within a small part of the data cloud, while Remain was highly spread. This only emerges from the multidimensionality facilitated by the novel approach adopted herein. Encouraging turnout and tailoring messages to appeal to potential Leave voters becomes easier and cements the result.

In the immediate aftermath of the Brexit result much was made of the demographic make up of those who had voted Leave. Initial summaries pointed to effects from age and education. Specifically those who were older and had lower levels of education were more likely to support Leave (Clarke and Whittaker, 2016; Becker et al., 2017; Manley et al., 2017; Arnorsson and Zoega, 2018; Sampson, 2017). To this mix many further studies have added additional importance to occupation (Harris and Charlton, 2016), income (Hobolt, 2016), and the related notion of income inequality (Bell and Machin, 2016; Darvas, 2016). More recent work has abstracted a quadratic effect from age, noting that those who are old enough to remember World War II are less likely to support Leave (Antonucci et al., 2017). This may be behind the finding in Zhang (2018) that age is not a significant determinant of the proportion of voters choosing Leave in a constituency, rather than there being a more fundamental miscalculation on the role of age. In their comprehensive analysis Becker et al. (2017) conclude that age and education can explain 80% of the variation in sub-regional voting behaviour, whilst economic factors can only explain 70%, and exposure to the EU is only responsible for 50% of the overall variation.

Another of the big questions around EU membership has come from the role of free movement of labour in unemployment figures. Unemployed individuals are found to be much more likely to vote for Brexit (Antonucci et al., 2017; Crescenzi et al., 2018, amongst others). Occupation more generally is found to be significant in Harris and Charlton (2016) and Matti and Zhou (2017) with those with lower level occupations more likely to vote Leave. Sampson (2017) also picks up this association as indicative of poor economic outcomes driving desire for change. Such arguments produce an interaction between age, education, employment and the regional context in which the individual exists. Such interactions are lesser explored.

Crescenzi et al. (2018) analysis sites these works within the broader notion of corporate beneficiaries and a working age population who believe they are far from the gains that proponents of EU membership speak of. Bromley-Davenport et al. (2018) exploration of the story behind the Brexit vote in Sunderland, a former ship-building powerhouse in North East England, paints a clear picture of exclusion and detachment from the pecuniary advantages of the EU. A belief in being "left-behind" is a common theme in the narrative of not only Brexit, but the wider gains made by popularist politics (Inglehart and Norris, 2016). Lee et al. (2018) takes a detailed look at the concept of missing out on the gains, noting that being immobile and detached from the growing centres of urban agglomeration is problematic for many. These gains from agglomeration, described neatly in (Chetty et al., 2014), are by no means the only factor effecting the notion of place (Lee et al., 2018), but they do go some way to explaining the link between the "left behind" and the education and age profiles that are also heavily cited as drivers of Brexit. Older individuals, the unemployed and those with lower levels of education are far less likely to move areas, or find new work to empower mobility; these demographics are precisely those shown to favour Brexit in demographic analyses. Data on perceptions of immobility is necessarily absent from geographical aggregations, but the links between constituency make

it. Leave is a contraction of leave the EU and Remain is a contraction of the statement that the UK should remain a part of the EU. The term UK is used as a contraction of the United Kingdom of Great Britain and Northern Ireland and, in keeping with the literature "Britain" and the UK are used interchangeably. Seemingly irrational as later discussed is tied here to the financial elements of the cost of Brexit on those communities that voted to Leave. However, such voting behaviour cannot be dismissed as irrational since financial considerations are only part of a bigger set of inputs to the utility function.

up and perceived immobility may be loosely made following the lead of Lee et al. (2018).

A primary reaction to the 2016 referendum result was that it was irrational, and that in the main it was the very people who benefit the most from EU membership who had voted to reject it. This hypothesis developed in Los et al. (2017) and others runs contrary to the perception of it being the metropolitan elites that were the main beneficiaries of being in the EU that Lee et al. (2018) and others exposit. As Los et al. (2017) contends the proportion of exports from London to the EU are ten percentage points lower than they are in any other UK region. Billing et al. (2019) notes this as the starting point for the regional responses that will help those regions to mitigate the additional impact Brexit may have upon them. In the misunderstanding of likely rationality in voting behaviours there is a message that feeds back to politicians belief that a referendum would be winnable for Remain (Bailey, 2018).

As well as the inherent role of demographics on individual decisions and voting behaviours there is a need to consider the influence of others on the decisions made by individuals (Liberini et al., 2019). There are many complex reasons why voters behave as they do, with the influence of the media (Jackson et al., 2016) and social media in particular (Lopez et al., 2017; Gorodnichenko et al., 2018) being of high relevance. Messages received by voters are thus subject to a number of distortions, a problem which plagues even the most seemingly neutral sources and outwardly honest predictions of the future possibilities (Cipullo and Reslow, 2019). It is perhaps unsurprising that many turn to their families for advice; Fox et al. (2019) study of parent-child behaviours shows a strong transmission of beliefs on Brexit between generations. Through these influence channels there are powers that can override the basic demographic correlations and introduce complexities and non-linearities of association that require further investigation.

Discussions implicitly blend spatial aggregation and individual level data, with many papers taking advantage of both. Individual data has the inherent advantage of being able to get at social attitudes, such as the conservatism linked to the Leave vote (Lee et al., 2018). However, from a representational perspective aggregation to either the voting districts used in the 2016 referendum, or the constituencies used in Hanretty (2017), Thorsen et al. (2017) and others, has power to enable the study of the whole population. This study belongs in the aggregated class, its data being that used in the Thorsen et al. (2017) work. The aim is to consider how local demographic characteristics combine to explain the voting behaviours observed at the constituency level.

Whilst the decision to leave the EU was taken at the national level much has been made subsequently of the differential voting patterns across the UK. Harris and Charlton (2016) compares expected leave support with that observed in the actual vote to determine a critical role for East England, the East Midlands and the South East in influencing the overall result. Much is made in the wider discussion of the post-industrial North, the "left-behind" areas that have yet to see real replacements for the heavy industry that is no more (Bromley-Davenport et al., 2018; Lee et al., 2018, and others). However, their voting for Brexit was as foreseeable as it was that London would vote strongly to remain. A subsequent narrative about those who supported Brexit being the very regions that would lose the most (Los et al., 2017; Billing et al., 2019) reinforces the aggregation of demographics and economic characteristics at the regional level. Through constituency level data this paper produces a representation of Brexit voting that can be viewed through the regional lens. It is demonstrated that many of the important margins identified in Harris and Charlton (2016) can be understood empirically with the established demographic drivers.

Regional disparities are tied in to the variations in referendum voting patterns amongst supporters of the major political parties. Links between the post-industrial North of England and the Labour Party are well understood (Harris and Charlton, 2016). Alabrese et al. (2019) asks whether it is possible to classify individuals as Leave voters using their demographics, finding that the prediction accuracy is affected by political allegiance. Where Conservative voters are more likely to vote Leave than their age, employment, education and gender may suggest, Labour voters are more likely to vote Remain. In essence whilst Labour voters in the North have the demographics to support Brexit more, the referendum result was moderated by their political allegiance. Such an observation is consistent with the surprisingly high Leave vote in regions that are more Conservative, such as the East of England. Links between political allegiance and voting behaviours at the constituency level are naturally dependent on the aggregate vote shares of the parties, this is shown to hold in the referendum where responses to the big two parties also differ. In Labour and Liberal Democrat marginals remain is far more likely than in Conservative and Liberal Democrat marginals, the latter being where the Brexit "surprise" reported in Harris and Charlton (2016) was strongest. Complexities in the political landscape are also represented conveniently in the TDA Ball Mapper approach employed here.

From the literature it is clear that there are many factors combining to produce the observed Leave vote. Such interactions have challenged the construction of the narrative but for constructing an evidence base they pose important statistical challenges that must be overcome. To this end analyses premised on linear relationships, or bivariate considerations, are prone to missing important elements of the story. Indeed the inherent multicollinearity within the use of demographic data with spatial aggregation means that studies are forced to find solutions, such as focusing on just a subset of the qualification levels or social classes, to obtain any model validity. Zhang (2018) for example contracts to just the percentage in the upper social classes, the percentage with degrees and the percentage who are unemployed within any given local authority area. Such remedies to modelling problems enable analysis but leave questions around the omitted characterstics and the granularities within the combined categories.

Topological data analysis (TDA) is a data-driven approach from the physical sciences that treats data as a point cloud and studies the topology thereof. Born of work by Carlsson (2009) TDA is well adopted in the physical sciences but is yet to take hold in the social sciences. It is free from assumptions about relationships and can, through the ability of computers, capture coordinates on any number of axes to fine grain all possible interactions. Against the discourse in the literature such an ability is vital to facilitate a strong evidence base. Whenever inference is driven from the data upwards in this way there is a suspicion around the degree to which results are representative of the population, small perturbations of the dataset may produce very different outcomes. Against this critique TDA has an important robustness that distributional methods such as regression do not. Movements on any of the axis of a point cloud simply distort the shape, they do not change the ordering of points or any conclusions drawn therefrom.

What follows is premised upon observed proportions within parliamentary constituencies, proportions which are not highly variant (Carl et al., 2019). Such invariance lends a stability to the analysis enabling focus to move firmly to the way voting behaviour varies across the point cloud. Representing the point cloud in a readily interperable manner becomes the aim, the TDA Ball Mapper algorithm of Dłotko (2019) being the answer. An exposition of the approach follows, the intuition being that any multidimensional dataset can be visualised in two dimensions by considering the strength of all co-locations within the point cloud that represents that dataset. Herein data on the 2016 EU referendum is converted to a point cloud and the topology studied for links to the outcome. Complex interactions that have challenged analysis to date are thus reviewed in a topologically faithful fashion that preserves every element of the underlying data.

The remainder of the paper is organised as follows. Section 2 takes a detailed look at the data that forms the axes of the point clouds, considering from a univariate perspective how values differ in Leave and Remain constituencies. Analysis proceeds using TDA and the TDA Ball Mapper algorithm; Section 3 introduces these. As a first stage voting patterns in the 2015 election that preceded the 2016 referendum are considered as an illustrative example of what TDA Ball Mapper can do. Section 4 indicates the roles that competition between political parties plays, and how this transposes to the Leave percentages in each constituency. Working systematically through the demographic make-up of constituencies Section 5 evidences many of the nuances of the relationship that to date have been consigned to the discursive literature, and also shows new patterns missed in work to date. Combining the dataset Section 6 looks at the full demographic dataset, giving some thought to regional disparities therein. Section 7 draws together the lessons from the empirical work and comments further on the impact for the understanding of the seemingly irrational UK 2016 EU Referendum result.

2 Data

Parliamentary constituency data is taken from the British Election Study and is as compiled by Professor Pippa Norris for work in Thorsen et al. (2017)[2]. For each constituency election results are recorded for the 2010, 2015 and 2017 general elections, including candidate details, voting numbers and percentages. In this way it is possible to chart the comparative performance of the major parties before, and after, the result of the 2016 Brexit referendum. Because the constituencies do not correlate directly with the counting districts used in the referendum Hanretty (2017) constructs an estimate of the percentage of voters who selected leave

[2]This data may be freely downloaded from `https://www.pippanorris.com/data`.

Table 1: Summary Statistics and Univariate Tests

Ques	Variable	Mean	s.d.	Min	Max	Leave v Remain			Strong Leave v Strong Remain		
						Leave	Remain	Diff	Leave	Remain	Diff
2015 Vote (%)	Labour	32.35	16.5	4.51	81.3	31.65	33.58	-1.93	34.18	35.37	-1.19
	Conservatives	36.66	16.16	4.67	65.88	39.6	31.46	8.14***	37.04	27.01	10.03***
	Liberal Democrats	7.82	8.36	0.75	51.49	6.96	9.34	-2.38**	4.62	9.5	-4.88***
	Others	23.17	11.85	6.09	65.33	21.78	25.62	-3.83**	24.16	28.12	-3.96*
Tenure	Own Outright	31.05	7.85	6.72	50.12	32.91	27.75	5.16***	32.09	25.82	6.27***
	Own on Mortgage	33.01	5.67	11.92	45.02	34.13	31.03	3.1***	34.1	29.95	4.15***
	Shared Housing	0.71	0.52	0.11	7.11	0.64	0.83	-0.19***	0.55	0.83	-0.29***
	Social Rent	17.99	7.8	4.59	50.63	16.73	20.21	-3.48***	18.17	22.04	-3.87***
	Private Rent	15.9	6.41	5.55	42.1	14.26	18.8	-4.54***	13.77	20.01	-6.24***
	Rent Free	1.35	0.43	0.57	4.01	1.33	1.38	-0.04	1.32	1.35	-0.03
Household	Live Alone	30.52	4.12	20.86	50.01	29.6	32.16	-2.57***	29.77	33.17	-3.39***
	Married	33.33	5.76	14.63	46.33	34.43	31.39	3.04***	33.73	29.97	3.76***
	Cohabiting	9.74	1.49	3.5	13.82	10.05	9.19	0.86***	10.5	9.27	1.23***
	Lone Parent	10.67	2.6	6.21	20.84	10.69	10.62	0.07	11.21	10.78	0.42
	Other	7.5	4.19	3.04	26.06	6.21	9.78	-3.57***	5.9	10.65	-4.75***
	All Students	0.57	1.27	0	9.86	0.23	1.18	-0.95***	0.12	1.42	-1.3***
	All aged 65 plus	0.28	0.09	0.12	0.75	0.29	0.27	0.03***	0.29	0.25	0.03***
No of Cars	0	25.54	11.57	7.86	66.7	22.84	30.32	-7.49***	24.84	33.83	-9***
	1	42.3	2.99	28.24	50.25	42.78	41.46	1.32***	43.12	41.26	1.85***
	2	24.76	7.71	3.78	42.53	26.35	21.95	4.4***	24.72	19.66	5.06***
	3	5.48	2.29	0.44	10.8	5.94	4.66	1.28***	5.47	3.95	1.52***
	4	1.92	1.05	0.14	5	2.1	1.61	0.49***	1.86	1.29	0.56***
NSSEc Status	Higher Managerial	2.28	0.87	0.58	6.07	2.21	2.39	-0.18*	1.95	2.29	-0.34***
	Higher Professional	7.69	3.24	2.35	19.57	6.46	9.86	-3.41***	5.17	10.12	-4.95***
	Lower Managerial	20.75	3.82	9.62	31.99	19.9	22.25	-2.34***	18.26	22.21	-3.95***
	Intermediate	12.82	2.02	6.98	19.93	13.23	12.11	1.12***	13.16	11.71	1.45***
	Small Employer	9.32	2.62	4.04	18.48	9.7	8.65	1.04***	9.19	8.08	1.11***
	Lower Supervisory	7.17	1.56	2.92	11.44	7.77	6.11	1.65***	8.33	5.99	2.34***
	Semi Routine	14.4	3.08	5.84	21.64	15.62	12.24	3.38***	16.95	12	4.94***
	Routine	11.42	3.57	3.62	20.76	12.63	9.27	3.36***	14.51	9.27	5.24***
	Never Worked	3.7	2.35	0.95	18.85	3.53	4.01	-0.48*	3.86	4.22	-0.36
	Long-term Unemployed	1.72	0.64	0.66	3.94	1.74	1.68	0.06	2.01	1.76	0.25***
Qualifications	None	23.32	5.83	9.57	42.48	25.06	20.22	4.84***	28.19	20.15	8.04***
	Level 1	14.29	3.68	5.68	29.28	14.48	13.96	0.52	15.25	14.29	0.96
	Level 2	15.29	2.18	7.26	18.56	16.31	13.47	2.84***	16.56	12.76	3.8***
	Apprentice	3.17	1.18	0.73	8.48	4.14	2.55	1.59***	4.23	2.08	2.15***
	Level 3	12.08	2.45	6.41	27.65	12.04	12.15	-0.11	11.65	11.96	-0.31
	Level 4	26.73	8.32	12.07	57.39	23.03	33.28	-10.25***	19.16	34.54	-15.37***
	Other	5.49	2.35	2.97	16.66	4.98	6.68	-1.70***	5.02	7.50	-2.48***
Self-rated Health	Very Good	47.49	4.01	35.22	60.4	45.51	50.99	-5.48***	43.81	51.85	-8.04***
	Good	33.62	2.13	26.81	38.38	34.45	32.15	2.3***	34.76	31.51	3.24***
	Fair	13.22	1.92	8.23	20.05	14.04	11.77	2.28***	14.83	11.5	3.33***
	Bad	4.38	1.23	1.92	9.05	4.65	3.89	0.77***	5.12	3.9	1.22***
	Very Bad	1.3	0.4	0.52	2.97	1.35	1.21	0.14***	1.48	1.23	0.25***
Deprivation	Level 0	42.14	6.98	22.21	59.71	41.39	43.47	-2.08***	38.48	42.7	-4.22***
	Level 1	32.56	1.75	28.14	38.38	32.66	32.37	0.3	32.91	32.43	0.48*
	Level 2	19.48	4.01	10.25	30.82	20.23	18.15	2.08***	22.17	18.41	3.76***
	Level 3	5.3	2.23	1.51	14.16	5.25	5.37	-0.12	5.93	5.75	0.19
	Level 4	0.53	0.33	0.1	2.28	0.46	0.64	-0.18***	0.51	0.72	-0.21***

Notes: Variables organised by question with totals for each constituency on each question being 100%. All variables relate to 2011 Census apart from the 2015 vote percentages. NSSEC is the National Statistics Socio-Economic Classification. Leave v Remain segregates constituencies on whether the Hanretty (2017) estimated leave percentage is greater (smaller) than 50%. Strong Leave v Strong Remain compares constituencies in the upper quartile of Hanretty (2017) estimated leave percentages with those in the lowest quartile. In both cases the difference is augmented by the significance of a two-sample t-test for equality of means. Data from Thorsen et al. (2017). Significance given by * - 5%, ** - 1% and *** - 0.1%.

and remain in each constituency. Data is also bound with the 2011 census to give a better picture of the economic make-up of each region.

An advantage of using constituency data is that it becomes possible to compare results from the previous years general election with the referendum vote. Average votes to leave in the Labour voting constituencies were slightly lower than those to remain, 31.6% estimated to favour out whilst 33.6% are considered to have voted remain. Conservative supporting constituencies had an 8 percentage point higher average vote for Brexit.

All questions relate to the 2011 census information and are incorporated here on the assumption that the demographic make up of each constituency did not vary greatly during the period since. Given the available information this is a reasonable imposition, and is commonplace in the literature. 2015 represents the last election prior to the referendum and is used to give a reflection of the political leanings of the constituencies. In news commentary it is often remarked that the leave vote was rooted in the Labour heartlands of Northern England, and hence the first line of Table 1 may represent a surprise. Using the 50% cutoff the average Labour vote share is higher in those areas where remain got the majority. Only when comparing the upper and lower quartiles of the estimated leave vote percentages from Hanretty (2017) does Labour achieve a greater electoral performance in leave constituencies. In both cases there is no significance to the results. For 2015 election winners, and incumbent government, the Conservatives the vote percentage was on average eight percentage points higher in leave constituencies. That difference rises to ten percentage points when only the highest and lowest groups are compared. Former coalition partners, and at the time of writing the only party openly committed to remain, the Liberal Democrats, have an average vote percentage which is 50% higher in remain voting constituencies. This figure rises to more than 100% higher when the quartiles are tested. For most purposes other represents the two national parties of Wales and Scotland, Plaid Cymru and the Scottish National Party. Both have spoken openly about supporting remain and hence the statistics support the broad association between higher voting percentages for other occurring in remain constituencies. Overall the results on individual parties suggest constituencies where the majority are conservative will be pro-Brexit, those returning Liberal Democrat or other MPs would be pro-remain. Labour supporting constituencies have an ambiguity to which the TDA analysis returns.

First of the questions from the 2011 census looks at home ownership and the Brexit vote. There are many contending perspectives on the impact that would be expected here. Table 1 informs the highest percentages of homes are owned either outright or on a mortgage, almost two-thirds of properties are in this set. There are slightly more social renters than private on average with the mean sharing and rent-free percentages both being very low. This is not true for all constituencies the minimum ownership and mortgage percentages are both much lower than the national average, while some constituencies have rental percentages around 50%. Ownership percentages are higher in Brexit voting constituencies, a significant difference of three percentage points in the majority comparison and over four percentage points in the quartile comparison. Rental figures go the other way with differences of around four percentage points in the straight majority comparison and more than eight percentage points in the quartile comparison. Such is consistent with the demographic discussion; younger voters being more likely to be in rental accommodation. Consequently constituencies dominated by rental accommodation are likely to have lower Hanretty (2017) estimated leave vote percentages.

Household constitution can also be seen to have strong association with the estimated leave percentages from Hanretty (2017). Living alone and being married are the two most common responses. Households comprising only students represent just 0.5% on average but there are some constituencies in university cities where the proportion is almost twenty times this. There is also considerable variation in cohabiting and households classified as lone parent. T-tests show Brexit voting constituencies to have higher proportions of married couples and over 65's, though the latter has a very small difference. Lower rates of living alone and reporting status "other" are seen in the leave constituencies. Extending the comparison into the quartiles reinforces the messages from the majority comparison with the magnitudes of the differences being around one percentage point higher. This differential is not as pronounced as it was for tenure but it is notable nonetheless. In looking at households which are all students the lowest quartile of leave percentages comprised 1.4% such households compared to just 0.12% in those constituencies with the strongest vote to leave. Such observations are consistent with the literature on education and voting patterns; they also recur in the discussion of qualifications here.

Number of cars is a useful proxy for income, and hence the next variable considered is the proportion of households which have a given number of cars. Highest proportion on average is one car with very few having either three or four. There are a large proportion of households with no access to cars as well. Comparisons between leave and remain voting constituencies show that average proportion of households without a car to be almost eight percentage points higher in the latter. Meanwhile properties with two are more cars are all higher in remain voters. Extending into the quartile comparison the difference in no car households rises to nine percentage points. Much of this can be linked to younger adults not having cars, but there will be much more laying behind this. Similarly the quartile comparisons show larger proportions with one, two, three or four plus cars in those contsituencies voting to leave by more than the 75th percentile. Of all of the variables used in the analysis this is the one where the proportions would be expected to be most correlated; large numbers of households with four or more cars would be expected alongside large numbers with three, rather than large numbers with one or none.

Social status, as captured by the National Statistics Socio-Economic Classification (NSSEC), has the most different axes of all of the questions considered in this paper. Ten different levels are included ranging from Higher Managerial down to those who have never worked. Average proportions in each group are highest for the Lower Managerial, Routine and Semi-Routine. There is great variation in all of the levels, as evidenced in Table 1. In the leave versus remain t-tests Higher Managerial, Higher Professional, Lower Professional and those who have Never Worked are all found to be higher in constituencies estimated to have a higher remain percentage by Hanretty (2017). Moving into the quartile comparisons the first three of these become even more pronounced, but the proportion of households in a constituency where a respondent has never worked is no longer significant. Here the "middle" effect comes through less strongly. Higher estimated leave percentages match with higher proportions of households classified as Small Employers, Lower Supervisors, Semi-Routine and Routine roles. Comparing the highest quartile of leave votes with the lowest again emphasises the contrast more, the largest difference being for Routine and Semi-Routine job classifications.

Broken down into seven groups qualifications are grouped as either apprenticeship, levels one to four, or other. There is also an option for having no qualifications. Levels are based upon the UK national framework. Across the whole UK having at least an undergraduate degree is the most common highest qualification level, but having no qualifications is reported by almost the same proportion of households. Some constituencies have more than 40% of their households with no qualifications, whilst others have almost 60% with degrees. In the t-tests of Leave versus Remain those with no qualifications represent an average 25% of pro-Brexit constituencies but just 20% of Remain. For the quartile comparison the margin rises beyond eight percentage points. At the other end of the scale those with degrees are found far more in Remain constituencies, comprising a third of all households there in. In the upper quartile of Hanretty (2017) estimated leave percentages the qualification level 4 percentage is below 20%. In the mid-range of qualifications there are higher levels of apprenticeships and Level 2 qualifications in the Brexit favouring constituencies. Summary statistics presented in Table 1 are in concordance with the literature on education and the Referendum result. Alignment between degrees and the student household proportions can be seen.

Within the population of constituencies the proportion of Census 2011 respondents reporting good, or very good, levels of health exceeds 80% with bad, and very bad, totalling a little over 5%. Great variation exists, but primarily it is still the higher levels that dominate. Maximum values for the proportion of household representatives reporting bad health is 10% and very bad is just 5% as a maximum. T-tests between Leave and Remain constituencies show the proportion reporting very good health to be six percentage points higher where the vote was for Remain. All other levels of self-reported health are higher in Leave constituencies, particularly fair and good levels. Extending to the comparison between the upper and lower quartiles of the Hanretty (2017) estimated leave percentages the same pattern continues with high significance on the differences. For the very good health level the difference between top and bottom quartiles is eight percentage points; for good and fair health levels the gap is four percentage points. Self-reported health is a very subjective measure, but the presence of a clear split between very good and the rest chimes with a narrative around disaffection and Brexit voting.

Last of the Census 2011 variables considered in this analysis is deprivation. Measured against four criterion the total level ranges from zero to four. Proportions of households in a constituency classifying at each level are then recorded as characteristics for analysis. Firstly where a household has a member who is either unemployed or long-term sick they are considered to be employment deprived. Secondly, where no person in the household has exceeded level two as their highest qualification, and there is no-one in the household working towards level three, then deprivation on the education measure is recorded. Thirdly, if there is a member of the household who has a bad, or very bad, health level, or who has a long-term health related problem, then that household is regarded as health deprived. Finally, housing deprivation is classified as the home being overcrowded or lacking central heating. Households may thus achieve a maximum score of four, but it is the lower numbers that appear in the largest proportion in all constituencies. For the full sample of constituencies level 4 is only recorded for 0.5% of households on average. In the t-tests of Leave versus Remain the least deprived proportion is two percentage points higher in the Remain constituencies, but level two deprivation is two percentage points higher in the Leave constituencies. Other levels have negligible differences. Moving to the quartile comparison the magnitude of the differences increase beyond four percentage points, but the conclusion that remain constituencies are less deprived remains.

Across the full dataset it can be seen that the broad averages are in concordance with the narrative around the Brexit vote. Lower education levels, higher levels of deprivation, lower social-class and poorer health are all part of an exclusion story that brought those who felt left behind to change the course of UK history. These are the factors consistently identified in the Brexit literature verifying the constituency level data as representative against the local authority area and individual level aggregations. This paper moves beyond these headlines to understand what is really going on behind the projected aggregate picture.

3 Topological Data Analysis

This paper focuses on the characteristics of constituencies incorporated as axes in a point cloud. Here the theoretical underpinnings that permit the exploration of such clouds are discussed together with an intuitive exposition of the way properties from the cloud aid understanding of data. Following the construction of a TDA Ball Mapper graph there are a number of further tools to deepen appreciation of the messages that emerge from the data. A visualisation tool is outlined that can then be employed to study how the demographics of a constituency link to voting behaviours.

3.1 Theoretical Method

Topological Data Analysis in general, and the TDA Ball Mapper algorithm in particular, is designed to answer the following question: what is the shape of a given collection of points X? Quite often we consider X equipped with a characteristic function $f : X \to \mathbb{R}$. For example, in the context of this work, X gathers various characteristic of UK parliamentary constituencies and f is the average Leave vote in the 2016 Brexit referendum in each of them. For X considering two or three characteristics, in which case it can be formally said that X is embedded to $\mathbb{R}^2$ or $\mathbb{R}^3$, the shape of X can be readily assessed by making a scatter plot of X coloured by the values of f. However, capturing the shape of X is becoming more difficult when X contains more characteristics and therefore is contained in a higher dimensional space. In that case, the challenge is to build landscapes of high dimensional data.

To efficiently solve this problem the TDA Ball Mapper algorithm (Dłotko, 2019) will be used. This approach may be briefly explained as follows. The only parameter of the algorithm is a positive constant ϵ. It should be thought of as a distance unit; all features of X that are smaller than ϵ will be disregarded in the analysis that follows. In the first step a subset X' of X is selected having the following property: every point x in X is at most ϵ away from some point in X'. Note that this condition implies that once balls of a radius ϵ centered in each point of X' are placed, they will contain all points in X. The points in X' will be referred to as *landmark points*. The way to think about X' is that it is typically much smaller collection of points, that have the same overall shape (up to the unit ϵ) as the whole X. One can obtain X' by construction so called ϵ-net. Please consult Haussler and Welzl (1987) for further details.

TDA Ball Mapper provides an abstract graph, G, referred to as a *TDA Ball Mapper graph*, that will capture the shape of the point cloud X. The vertices of G correspond to the landmark points in X'. It is worth noting that they typically should not be thought of as the points in X' which may be of a very high dimension, but rather abstract vertices. In this case the vertex v is a *representative* of all the points in X that are not farther away than ϵ from the point in X' corresponding to the vertex v. Two vertices v_1 and v_2 are joined with an edge in G if, and only if, the balls of radius ϵ centered in the corresponding points of X' both contain some points in X.

There is an obvious weighting associated with the vertices of G; Vertex v in G can be weighted by the number of points in X contained in the ϵ radius ball centered in the vertex corresponding to v. In what follows the weighting of vertices of G will be visualized by varying the size of vertices. The TDA Ball Mapper graph constructed in this way gives an idea about the geometric landscape of X.

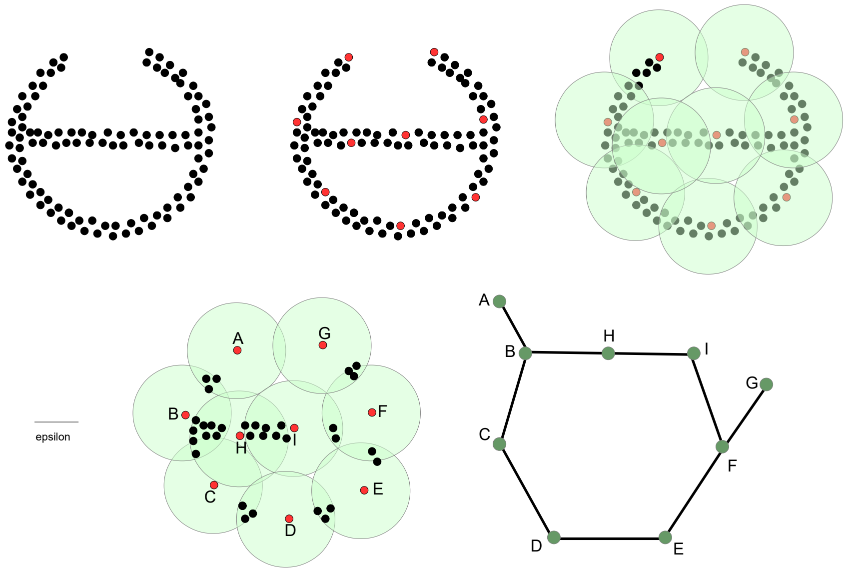
In addition the vertices of G can be coloured using the values of the function f in the following way: The colour of each vertex v of G corresponds to an average value of function f on all points of X in the ϵ radius ball centered in the vertex corresponding to v.

3.2 An Artifical Illustration

To visualise the process through which TDA Ball Mapper creates a cover of the space consider the flow illustrated in Figure 1. This image features a circle with a bar across it and a gap in the circle at the top. The shape is formed from a series of points in the two dimensions, horizontal and vertical. It is a two-dimensional point cloud. Any representation of this would need to have these three features and to preserve the two semi-circular white space between the bar and the perimeter points on the circle. Through the work-flow it can be seen that such a representation emerges.

Firstly a point is selected at random from the full set of data points. A ball of radius ϵ is drawn to surround it. A process of point selection and ball drawing continues until all of the points in the cloud are

Figure 1: TDA Ball Mapper Process



Notes: Schematic illustration of the TDA Ball Mapper process as constructed in Dłotko (2019) and implemented using *BallMapper* (Dlotko, 2019).

covered by at least one ball. At the point at which the coverage is completed the top right image of Figure 1 is arrived at. TDA Ball Mapper graphs have edges connecting points wheresoever there are points in the intersection of two balls. The first image on the second line shows the points that are in the intersections of the balls. Hence A is connected to B but not G; despite the balls overlapping there are no points in that intersection. A connects to B but also has no points in its intersection with H. Working around the shape B is further connected to H and C, C connects to D, D to E and E to F. Meanwhile H is connected to I and I to F; this forms the bar across the centre of the shape. Finally F is connected to G. The resulting shape is shown in the final image.

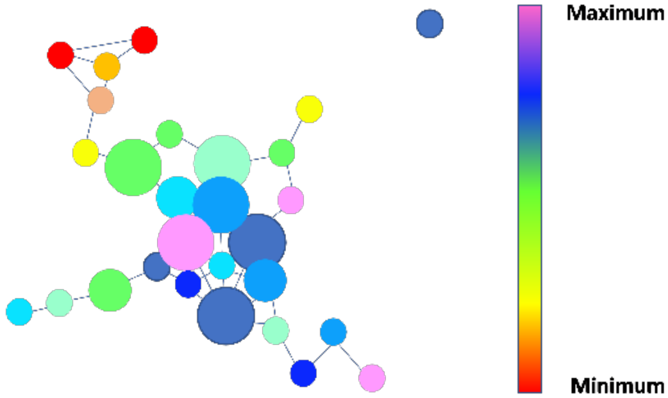
TDA Ball Mapper graphs are two-dimensional representations of a multi-dimensional space and to show this the final figure in Figure 1 is deliberately distorted. In distortion it no longer appears that A and G are close to each other on the circle, but the topological information about their connection to B and F respectively is preserved. The final shape may be recognised as the one that was started with in the top left panel. In the same way the visualisations of multivariate data in the analysis that follows continues to be faithful to the underlying dataset, but nothing may be read into the distances between points in the diagram. Like A and G points shown as unconnected simply inform on a lack of proximity.

3.3 Augmenting Mapper Plots

TDA Ball Mapper plots of the type constructed in Figure 1 are abstract two-dimensional representations of multidimensional point clouds. They may be usefully augmented to convey information about the vertices that lie within the graph. For example unless the data is evenly spread through the space it follows that there will be differences in the number of points covered by each ball. As noted ball size is a function of the ϵ radius parameter and so it is useful to have the visualisation reflect how many observations are fit within each ball. Ball size is accounted for using the size of the ball in the plot. Through the paper it will be seen that some balls are much larger, these are the ones with the most constituencies within them.

A second important functionality is the ability to colour the balls according to some factor of interest. In the simplest case this will be according to the outcome but any measure for which all points in the cover

Figure 2: Augmenting TDA Ball Mapper plots



Notes: Artifical example presuming data on multiple axes. Ball size represents the number of observations within the ball. Colouring variable runs from *Minimum* to *Maxiumum* with colours expressed on a uniform scale between these values.

have a value may be used. Figure 2 shows how colouration can be understood from a TDA Ball Mapper plot, the scale on the right indicating how the colours relate to actual values of the colouration variable. Figure 2 shows how the lowest values of the outcome appear in the upper left of the representation, whilst the highest can be found in the centre-right. Arms sticking off the shape to the left appear to increase in outcome moving away from the main shape, whilst that moving to the upper right falls in value. Recalling that the construction of a TDA Ball Mapper diagram requires that balls which are not connected are sufficiently different in at least one characteristic, seeing patterns like this would be evidence of non-linearity. Following the intuition demonstrated by the lack of connection between points A and G in the artificial example of Figure 1 it cannot be assumed that the four arms that go to the four corners really end in opposite areas of the parameter space.

Colouration is an immediate informant on non-linearity, but it is also a visual aide to highlighting interesting stories that may lie within the data. To the upper right of the shape we see a very high outcome mauve ball connected to a turquoise mid-range ball. These balls are smaller but their connection means that there must be strong similarities on all of the axes. Understanding how their outcome is so different would be an obvious step for the analyst viewing such a picture. When the outcome is a variable of interest, such as the Brexit Leave vote, the researcher should seek to ascertain why there was such a different answer to the referendum question. Should the outcome be the residual from a model then instead the question becomes why does that model fit so differently on such similar observations. In each case augmenting the TDA Ball Mapper graph with colour advantages understanding. Combining with ball size as an indicator of relative size of the comparison Figure 2 highlights two important insights offered from the *BallMapper* (Dlotko, 2019) package that are not readily given by conventional analyses.

3.4 Post Mapper Analysis

Because each TDA Ball Mapper diagram is abstract it is not possible to read quantitatively from the diagrams. However, all of the topological information about the points is preserved.

Immediate measures available include the number of vertices, the number of points covered by each vertex, the number of edges and the average number of edges. These are readily extracted from the graph created by the TDA Ball Mapper algorithm. Further using the unique identifiers of the observations within each ball it is possible to compute summary statistics by matching back to the main dataframe where the ball number can sit alongside the main data. Augmenting the dataset with the ball membership in this way means also allows colouration of the TDA Ball Mapper graph according to any function of the input information. For example the standard deviation could replace the mean, or the proportion of constituent points with a particular binary characteristic. Functionality on alternative colouration is employed widely in the analysis that follows, not least to colour the plots according to the proportion of observations in a ball from a given geographic area.

Further analysis on TDA Ball Mapper output can inform on the distances between points and a set of data points considered to be of interest. Consider the case where points appear unconnected, as A and G do in the artificial example, it would be possible to find out which dimensions prevented them from being joined. In Figure 1 A and G are of very similar vertical coordinate so it is the horizontal distance that is of interest. By contrast G and H are not joined by virtue of the distance diagonally across both axes. A further possibility is that, like points H and I in the artificial example, two points are connected but subsequent relation to an outcome variable reveals them to have very different values. In such cases understanding the differences in the characteristic space might illuminate more on why particular outcomes are observed. In practical terms wherever there is an interest in the relationship between two points the topological information allows computation of the distance through the parameter space. Distances are calculable without a requirement for TDA Ball Mapper, the role of the coverage is to identify cases to measure and to see patterns that might otherwise have been obscured by either reduced set comparisons, or because the link from input to output is not fully understood.

For any given vertex v in G the coordinates of all points within the ball that surrounds v may be averaged to produce a single set of coordinates $x_1^v, x_2^v, ..., x_d^v$. d here simply denotes the number of axes of the TDA Ball Mapper graph. From these values the absolute distance of each and every point in a ball i from another ball k may be measured using:

$$abdist_{i,k} = \sum_{j=1}^{d} |x_{i,j}^v - x_{k,j}^v|$$

This measure may be distorted by variables that have a larger standard deviation and so a normalised version is used with:

$$dist_{i,v} = \sum_{j=1}^{d} \frac{|x_{i,j}^v - x_{k,j}^v|}{\sigma_j} \tag{1}$$

Here σ_j is the standard deviation of axis j. In what follows σ_j this would be the value reported for variable j in Table 1. This functionality can aid the understanding of distances to areas of interest within the point cloud, or to compare against specified coordinates. From a distributional perspective it may be interesting to see the distance from the mean. From an application perspective the distance to the constituency with the closest margin between Brexit and Remain might be of interest. Within a TDA Ball Mapper graph the function can be used to understand how separated components become connected, for example where would an outlier group connect into the main shape[3].

4 2015 Election Voting

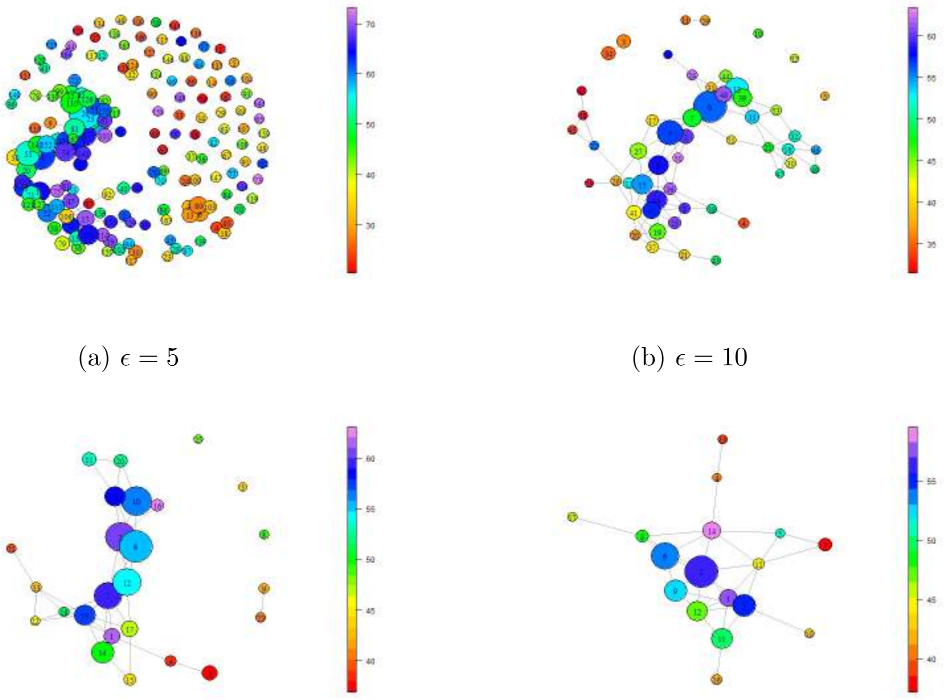
In the discourse on Brexit much is made of the position of the major political parties. In constituencies represented by a Labour MP the average Leave vote is 52.32% compared to 51.92% in other constituencies. By contrast the Conservatives position is much more straightforward. An average of 55.45% voted Leave compared to 49.45% as the average from other constituencies. This difference is six percentage points and is significant at the 0.1% level. A deeper exploration of the role of 2015 election results in understanding the Brexit vote now follows, highlighting how TDA Ball Mapper can bring neat insight into the information nested within the data.

4.1 Ball Radius (ϵ) Selection

Within the TDA Ball Mapper process there is one key parameter, the radius of each ball ϵ. Careful consideration must be applied to the selection of ϵ since too low values will mean too many clusters and a lack of

[3]Because of the abstract nature of the plot it does not follow that the closest point in the plot is actually the closest point to the outlier. This is shown later for proportion of households with a given level of qualifications within constituencies.

Figure 3: Mapper Plots for Election 2015 Voting



(a) $\epsilon = 5$ (b) $\epsilon = 10$

(c) $\epsilon = 15$ (d) $\epsilon = 20$

Notes: ϵ reports the radius used within the mapper algorithm. Plots generated using Dłotko (2019). Red represents the lowest leave percentages, blue the highest. All other colours are a spectrum between the two limits. Labelling shows the process through which clusters become subsumed within others, or, where multiple are identified, sit on the intersection between balls in larger radii. Data from Thorsen et al. (2017).

clarity in the message that is understood from the data. Choosing a value which is higher will mean more points in each ball, more points in the intersection and hence fewer balls with increased connectivity. These larger clusters will be understood by their outcome averages, which in turn will be the contraction of more data. Consequently, as is now demonstrated, the first phase of any development should be the identification of a suitable ball radius.

Consider the vote shares for the three largest political parties and a fourth axis that represents the total proportion of the vote for all other parties. Figure 3 shows how increasing the radius reduces the number of balls and leaves a graph which is easier to interpret. In panel (a), with $\epsilon = 5$ there are 165 balls, many with only one constituency. Few balls join meaning that there are large numbers of outliers floating in the space to the top right of the plot. As ϵ rises to $\epsilon = 10$ the number of outliers is drastically reduced, with most points becoming part of the connected shape. At this radius there are 53 balls. By panel (c), where $\epsilon = 15$, there are very few outliers and the number of arms from the main shape has reduced. Essentially the connected shape has become a "T" with low Brexit votes in the two arms. There are 25 balls in panel (c). Increasing the radius from here does not produce as large reductions in the number of balls. With $\epsilon = 20$ there are 17 balls. As panel (d) of Figure 3 informs, there are no longer any unconnected constituencies.

Table 2 reports a subset of the estimations to demonstrate how as the ball radius increases the number of balls falls rapidly in the first stage. Numbers then fall much slower as the ball radius continues to increase. The number of points within the balls necessarily increases, but it does not follow that the balls are evenly sized, the standard deviation of the ball size is also rising through the increasing ϵ. Number of edges vary,

Table 2: Radius and Ball Size: 2015 Vote Percentages

ϵ	Balls	Size (Mean)	Size (sd)	Edges	ϵ	Balls	Size (Mean)	Size (sd)	Edges
1	585	1.084	0.301	0.005	15	25	51.08	52.23	1.720
2	420	1.709	1.238	0.190	20	17	78.29	73.56	1.882
5	165	6.412	7.618	1.067	25	11	113.8	114.9	1.818
7	98	12.29	15.21	1.602	30	8	144.1	138.6	2.125
10	53	23.25	26.30	2.057	35	7	185.3	154.5	2.143
12	35	33.46	39.47	1.686	40	5	240	137.0	1.800

Notes: Table reports results from construction of a TDA Ball Mapper graph using *BallMapper* (Dlotko, 2019). Size (Mean) gives the average number of points within the ball, with Size (sd) being the standard deviation of the sizes. Edges is the number of edges in the graph divided by the number of balls. Data from Thorsen et al. (2017).

as the number of balls falls so the numbers of edges begins to hold close to 2[4].

4.2 TDA Ball Mapper Analysis

Having explored the role of radius using the voting percentage data it is determined that $\epsilon = 12$ represents a good balance between detail and readability. Figure 4 shows four clear arms stemming from a central body. The arms represent areas where one of the main parties performs much better/worse than average. Much of the Brexit support is through the centre of the plot with the top arms and others recording the lowest estimated percentages. Some variability is seen in the lower arms. To determine more about the characteristics of the balls through the shapes panels (b) to (e) show vote percentages for each of the parties. There are five outliers, three clustered as a triangle and two others which stand alone. All five are remain voting.

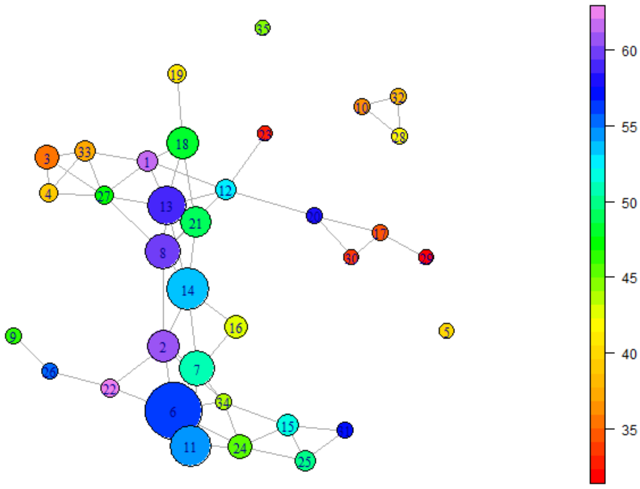
Panel (b) reveals that Labour is strongest towards the top of the plot, competing in the two remain supporting arms with the Liberal Democrats and the Other parties. Conservatives, plotted in panel (c) meanwhile are strongest in the balls to the lower end of the plot, competing in the centre, ball 14 in particular, with Labour. Within the main central body of the plot Leave percentages are all above 45% with most above 50%. Both parties have core balls where their support is strong but the colouration of panel (a) informs Leave percentages to be lower. Ball 16 is of particular interest in this regard. Panels (a) to (c) thus show well how the battle between the main two UK political parties is being fought in both Remain and Leave constituencies.

To see the roles of the arms a look at panels (d) and (e) confirms that the right extensions are where the Liberal Democrats are strong, whilst the left arms are where the Scottish National Party and Plaid Cymru have their highest percentages. The top arms are where these parties are competing with Labour, and the lower arms where their main opponent is the Conservatives. Given traditional poor Conservatives performance in Scotland, and the low showing of Labour in the South West of England the upper arm has more variation on the left, and the lower splits out more on the right. Where the Liberal Democrats are strong the Leave percentages are much lower, especially in balls 17, 29 and 30. Balls 3, 4 and 33 also have very low Hanretty (2017) estimated leave percentages. Panels (d) and (e) also help with the appreciation of the outliers. These are areas where the Liberal Democrats are competing with the other parties, a situation which is typically only found in northern Scotland. Ball 35 has a swing between Conservatives and Liberal Democrats but has a much lower Labour and Other percentage than those in the lower arm.

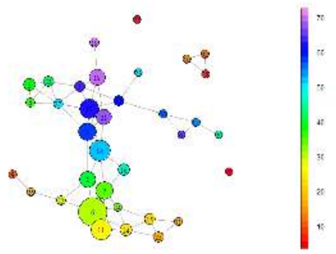
A first comparison is drawn between balls 1 and 33 which sit at the top left of the plot, the former having a Leave percentage well in excess of 60% and the latter being less than 40%. That two constituencies with such similar voting behaviour produce such different outcomes carries much interest. Table 3 shows the biggest difference is the change from Labour to Other, the latter seeing an increase in share of twenty-five percentage points on average. Ball 16 stands out in the centre of the plot as a having a low leave percentage but being attached into ball 14 where Leave percentages are estimated by Hanretty (2017) as being much higher. In ball 16 there are lower votes for Labour and Other but a much higher vote for the Conservatives. None of the changes are large in terms of number of standard deviations but the outcome is very different. This

[4] Full results are omitted for brevity but are available on request.

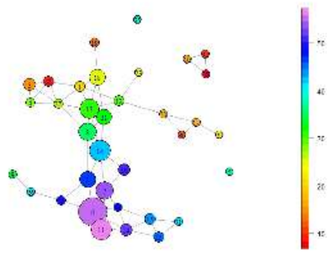
Figure 4: 2015 Voting ($\epsilon = 12$)



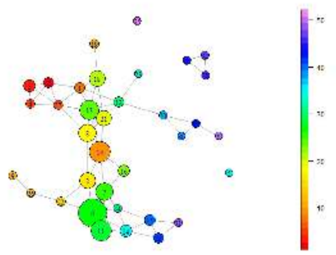
Hanretty (2017) Leave Percentages



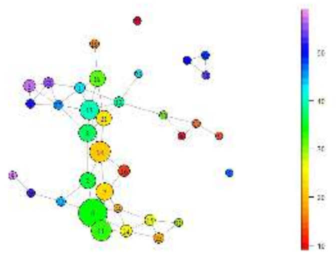
(b) Labour



(c) Conservatives



(d) Liberal Democrats



(e) Other

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (e) are coloured according to votes for other parties. Plots only cover constituencies with percentages recorded for all four parties. Data from Thorsen et al. (2017).

Table 3: 2015 Election: Selected Ball Comparisons

Variable	Ball 1	Ball 33	Diff	Std	Ball 16	Ball 14	Diff	Std	Ball 16	Ball 14	Diff	Std
Labour	50.26	32.51	17.75	1.08	36.30	39.94	-3.64	-0.22	10.16	11.40	-1.24	-0.02
Conservative	14.88	9.38	5.51	0.34	47.40	40.58	6.82	0.42	36.51	41.75	-5.24	-0.32
Lib Dems	3.54	1.93	1.61	0.19	4.29	3.61	0.68	0.08	35.83	30.22	5.61	0.67
Other	31.32	56.18	-24.87	-2.10	12.01	15.87	-3.86	-0.33	17.50	16.63	0.87	0.07
Size	24	24			32	107			6	32		

Notes: Table reports average 2015 voting percentages for each party within the balls indicated in the header. Lib Dems is used in short form for the Liberal Democrats. Diff reports the difference in percentage points between the two means. Std divides the difference by the population standard deviation for that variable for a standardised perspective. Size is a measure of the number of constituencies within each each ball, though it must be remembered that the existence of an edge means some balls appear in both constituencies. Data from Thorsen et al. (2017).

highlights the ability of TDA Ball Mapper to identify interesting cases within the data. A third comparison looks into the right hand end of the lower arm in which Conservatives and the Liberal Democrats perform best. Here there is a notable contrast between the high leave percentage of ball 31 and the pro-remain 15 and 25. Here the main difference is a swing of five percentage points from Conservative to Liberal Democrats. There are just 6 constituencies in ball 31, and 32 in balls 15 and 25 combined, but the message taken from these points is still useful to understand the effect of competition between parties.

Other interesting comparisons can be found between balls 12 and 23, the latter being a remain voting outlier loosely connected with the main shape through ball 12. Differences there are driven by a larger other party vote and lower share for the main two parties. Three outliers cluster together, balls 10,28 and 32, with strong support therein for Liberal Democrats and Others. Closest within the main group is ball 23, but the difference on both the Liberal Democrat and Other axes is almost two standard deviations. Labour poll almost 40% in ball 23, but receive less than 9% on average in the cluster of three. Such comparisons, and others, may be usefully explored using the functionality of *BallMapper* (Dlotko, 2019).

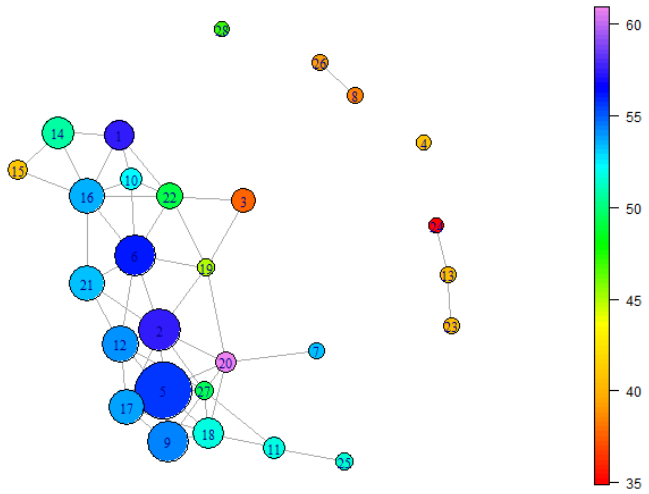
A challenge for the mapping of vote shares is that there are many constituencies where only three candidates stand. In these cases the four axis analysis above is forced to drop them. Maintaining the constituencies and assuming that the vote for others is zero would lead to connections where a three axis consideration would not form one. A similar consideration must be made about whether to recalculate vote shares in constituencies where another candidate did stand. However, doing so would risk losing information about voting patterns and so in such constituencies the actual percentage polled continues to be used. Figure 5 shows the three axis analysis with $\epsilon = 12$.

In this three axis case the majority of the balls have an average Hanretty (2017) estimated leave percentage above 50%. Only a subset of the balls toward the top of the plot are shown to favour remain. There are also a number of outliers, all of which are remain. Looking at panels (b) to (d) the dominance of Labour and the Conservatives remains evident. Labour performs best towards the top of the connected shape, including the areas where Remain is identified as the preferred referendum option. Recall that much of this may be attributed to Scotland where the 2016 vote was much more skewed towards Remain. The influence of the Liberal Democrats on the Conservatives can be seen towards the bottom right of the plot as the light blue colouring indicates much lower Leave voting than in other areas where the Conservatives are strong.

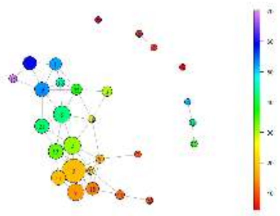
A first interesting case within the Conservative favouring balls lies in the remain voting ball 27 versus its heavily Leave voting neighbour of ball 20. Table 4 considers this contrast, as well as looking at the broader comparison of 27 with all of those it is connected to. The differential within vote shares for the main parties is very small and it is the strength of the Liberal Democrats which stands out as the difference. Given the positioning of the Liberal Democrats on Brexit it is unsurprising that the difference be in this dimension; it is confirmatory that TDA Ball Mapper has identified this group as being different however.

Through a review of the voting behaviour in the 2015 General Election this section has shown how TDA Ball Mapper can identify marginal constituencies and the effect that the respective rivalries have on the Brexit vote. Within the Labour vote large differentials were identified with only those where the Conservatives were also strong having average Hanretty (2017) Leave percentages greater than 50%. The Conservatives also cover a heterogeneous area, not least where they are competing with the Liberal Democrats. As the Labour party wrestles with its position on Brexit, these plots are a timely reminder of the challenge facing the leadership navigating that particular issue. It has also been seen that constituencies with very similar voting

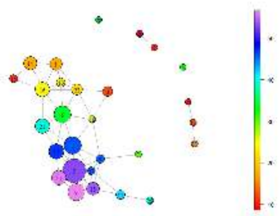
Figure 5: 2015 Voting: Three Parties



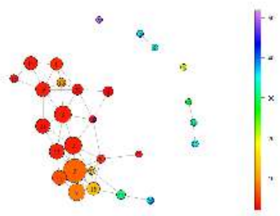
(a) Hanretty (2017) Leave Percentages



(b) Labour



(c) Conservatives



(d) Liberal Democrats

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being in the transition from green to light blue. Panels (b) to (d) are coloured according to votes for other parties. Plots cover all constituencies with percentages recorded for the other parties omitted. Data from Thorsen et al. (2017).

Table 4: 2015 Election 3 Axis: Selected Ball Comparison

Variable	Ball 27	Ball 20	Diff	Std	Balls 2,5,9,11,12,18,20	Diff	Std
Labour	22.19	17.53	4.66	0.28	21.87	0.31	0.02
Conservatives	45.20	45.17	0.03	0.00	49.67	-4.47	-0.28
Liberal Democrats	14.58	4.79	9.79	1.17	8.90	5.68	0.68
Constituencies	16	27			338		

Notes: Table reports average 2015 voting percentages for each party within the balls indicated in the header. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection. Data from Thorsen et al. (2017).

Table 5: Selected Comparisons: Tenure ($\epsilon = 10$)

Characteristic	Ball 13	Ball 17	Diff	Std	Grp A	Grp B	Diff	Std	Ball 14	Grp C	Diff	Std
Outright	7.59	20.17	-12.58	-1.60	24.68	28.21	-3.52	-0.45	39.78	36.60	3.17	0.40
Mortgage	16.18	14.33	1.85	0.33	31.42	34.96	-3.54	-0.63	37.43	35.14	2.29	0.41
Shared	2.24	0.83	1.41	2.73	0.70	0.64	0.06	0.12	0.59	0.65	-0.06	-0.12
Rent Social	50.37	23.84	26.53	3.40	24.52	22.47	2.15	0.28	9.67	12.73	-3.06	-0.39
Rent Private	22.41	37.95	-15.54	-2.42	17.38	12.59	4.79	0.75	11.23	13.49	-2.16	-0.34
Other	1.21	2.88	-1.67	-3.89	1.29	1.23	0.06	0.14	1.21	1.39	-0.19	-0.43
Ball Size	2	3			117	164			78	313		

Notes: Group A comprises balls 2, 5 and 16. Group B comprises balls 1 and 20. Table reports average proportion of households having each tenure form within the balls indicated in the header. Outright is a contraction of Owned Outright, whilst Mortgage is a contraction of Owned with a Mortgage. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017).

behaviours can have very different outcomes in the 2016 Referendum. Value of the TDA Ball Mapper method in showing relationships away from the linear, and to evidence more complex patterns in a simple way, is thus shown.

5 2011 Census Variables

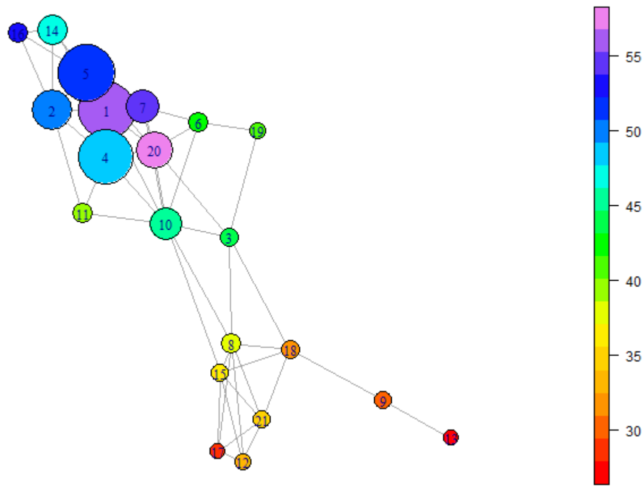
This study is focused on the links between characteristics of the UK Parliamentary constituencies and the 2016 EU Referendum outcome. Working systematically through questions from the 2011 Census this section charts the links between demographics and Brexit voting.

5.1 Tenure

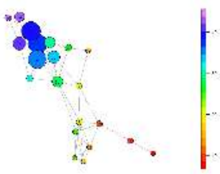
Figure 6 plotted with $\epsilon = 10$ is dominated by large Brexit supporting constituencies to the top left. In the colour by variable plots, panels (b) to (g), the alignment between the Leave vote and the two ownership measures is clear. However, amongst this lies ball 14 with its estimated leave percentage closer to 42%. The existence of these contrasts within such plots demonstrate the value of TDA Ball Mapper as a representation of a dataset. To the lower end of the plot the rental variables dominate with shared and rent-free highest further down the plot. In each case it is interesting to note variation in the non-ownership percentages amongst that top Brexit favouring set. Ball 1, the second highest Brexit vote, after ball 20, has a relatively high proportion of rent free, it is blue in panel (g). Ball 3 immediately adjacent is a darker orange in the two rental variables and yet produces a lower estimated leave percentage on average. Given the comparatively low percentages of shared and rent-free that variation in these does not drive the overall Brexit vote is not surprising. These contrasts with balls 1 and 3, as well as those between 1, 2 and 7.

Using the functionality of *Ball Mapper* (Dlotko, 2019) it is possible to drill further into some of the variations across the plot. First, to highlight some of the division between the Remain voting constituencies a comparison is drawn between the lowest Hanretty (2017) estimated Leave percentages from the ends of the split at the base of the shape. Thus ball 13 is compared with 17. These are small balls, containing just 2 and 3 constituencies respectively, but they are very different on almost all categories. Ball 13 is dominated by rentals, particularly social rentals, whilst ball 17 has far more private rentals. Both are different from the average constituency since their ownership percentages are far lower than the national average. Ball 17 has more outright ownership but is still ten percentage points lower than the national average. A second comparison between Group A (balls 6 and 10) and Group B (balls 7 and 20) contrasts constituencies on the edge between Remain and Leave. Indeed Ball 20 has the highest average Leave percentage, of around 57%, whilst balls 6 and 10 are both closer to 45% according to the Hanretty (2017) estimates. In this contrast the balls are not particularly different, a slightly higher rental percentage in the Remain pair constrasts with slightly higher ownership percentages in Group B. However, as Table 5 confirms none of these differences are more than one standard deviation. A similar lack of differential is found when looking at the Remain

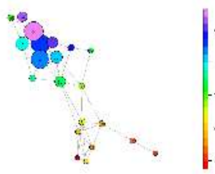
Figure 6: Tenure ($\epsilon = 10$)



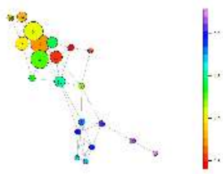
(a) Hanretty (2017) Leave Percentages



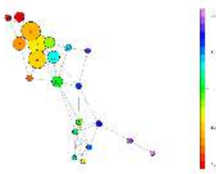
(b) Outright



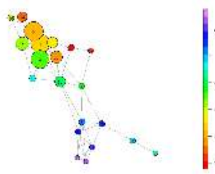
(c) Mortgage



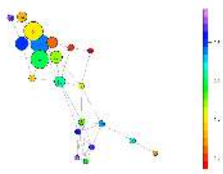
(d) Shared



(e) Social Rental



(f) Private Rental



(g) Rent Free

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (g) are coloured according to the proportions of households in each constituency of each tenure. Outright and Mortgage constitute ownership, Social Rental and Private Rental are the primary rental categories. Shared ownership is uncommon in the sample, but as with Rent Free is non-zero in many constituencies. Data from Thorsen et al. (2017).

Table 6: Selected Comparisons: Tenure ($\epsilon = 9$)

Characteristic	Ball 3	Grp A	Diff	Std	Ball 11	Grp B	Diff	Std	Ball 9	Ball 1	Diff	Std
Alone	26.33	29.52	-3.19	-0.77	23.01	28.04	-5.03	-1.22	32.58	30.16	2.42	0.59
Married	40.57	35.24	5.33	0.93	35.26	36.71	-1.45	-0.25	27.95	34.21	-6.26	-1.09
Cohabit	9.42	9.83	-0.40	-0.27	4.90	9.74	2.82	1.09	9.84	9.96	-0.12	-0.08
Lone Parent	8.13	10.30	-2.17	-0.84	12.86	10.03	2.82	1.09	12.15	10.72	1.44	0.55
Other	5.01	5.98	-0.95	-0.23	19.42	6.33	13.09	3.12	11.81	6.05	5.76	1.37
All Students	0.05	0.22	-0.17	-0.14	0.69	0.18	0.50	0.40	1.77	0.24	1.53	1.21
All 65-plus	0.27	0.29	-0.02	-0.17	0.30	0.28	0.02	0.23	0.27	0.29	-0.02	-0.37
Ball Size	101	511			8	359			90	430		

Notes: Group A contains balls 1, 4 and 16. Group B features balls 4 and 5. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017).

average calculated for ball 14 against its neighbours in Group C (balls 2, 5 and 16). In this case ball 14 has greater ownership percentages than the average of its neighbours in the representation; these would associate normally with Brexit supporting rather than being in the direction evidenced. Again there are no differences coming close to one standard deviation for the variable being contrasted.

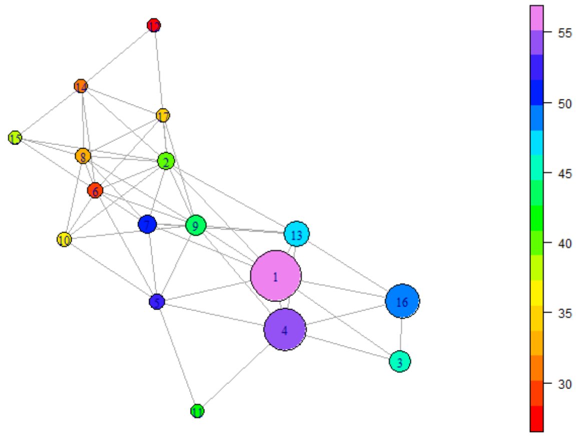
Overall these comparisons demonstrate why TDA Ball Mapper can be useful for highlighting combinations of characteristics within the space. Ball 14 would sensibly be a high leave vote case with the ownership being so high. Plots in Figure 6 show the ownership proportions rising toward the top of the plot, whilst the rentals rise moving from top to bottom. A linear relationship would not produce ball 14 and yet TDA Ball Mapper highlights it clearly. Contrasts between group A and B were also interesting in that there were no large variations in the axis variables but there was a combination that produced the highest Brexit vote very close to balls with large Remain majorities. Such contrasts would again defy the idea of a linear relationship.

5.2 Household Composition

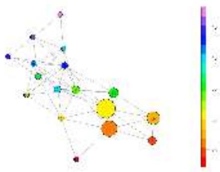
Figure 7 shows the largest numbers of constituencies to be in the lower right and an area dominated by balls 1, 4 and 16 and their high average Leave percentages. To the left of the plot there are a number of interconnected smaller balls, many of which have average Leave voting percentages close to 30%. This spread is consistent with the message from Figure 6 where there was also a spread of smaller Remain balls. Through colouration by axes in panels (b) to (h) it can be seen that these large balls correspond with high proportions of married household representatives and low numbers of one person households. Within the plot there is a stronger sense of correlation between the axes variables and the outcome but there is a contrast around ball 6 that is worth further thought. Balls 3 and 11 toward the bottom of the plot also stands out as points of interest in an otherwise pro-Brexit space.

Table 6 considers three comparisons from Figure 7. First it is explored how the Remain favouring ball 3 differs from the three Brexit voting balls to which it is connected. Ball 3 has a higher proportion of married household heads and lower proportions of people living alone and students. These three are counter to the suggestion from the t-tests earlier that suggested all would bring a higher leave vote, not the lower one seen in ball 3. Balls 1, 4 and 16 do contain, on average more single parents and more households where all residents are aged 65-plus, but these differences are small in absolute terms and in relation to the standard deviations of those variables. Here TDA Ball Mapper is revealing a set of points that do not conform with the interpretation from univariate tests. A second comparison between a small Remain favouring ball, 11, and the two Leave favouring balls to which it is attached has some similar counterintuitive proportions on single person households and married couples. However, in this comparison ball 11 has more students and this is consistent with a remain vote. Ball 11 also has a very high percentage of respondents reporting their household composition to be other, more than three times the figure for the two Brexit balls; more investigation of this with data that had more information on what was meant by "Other" would be suggested. Finally a look is taken at a pair of points from the centre of the main shape, balls 9 and 1 which have a very stark contrast in the Hanretty (2017) estimated Leave percentages. The former is around 40% while the latter is the highest at above 55%. Here the rankings are more consistent with Table 1 as Ball 1 has

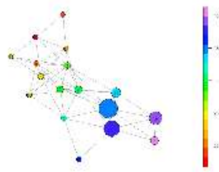
Figure 7: Houeshold Composition ($\epsilon = 9$)



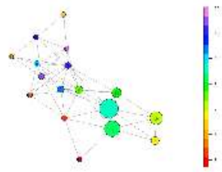
(a) Hanretty (2017) Leave Percentages



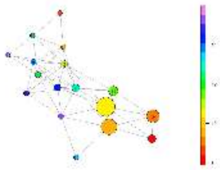
(a) Alone



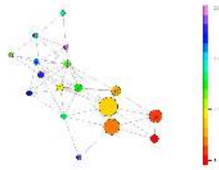
(c) Married



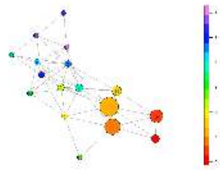
(d) Cohabit



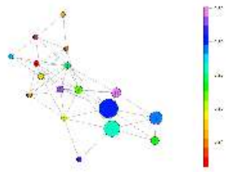
(e) Lone Parent



(f) Other



(g) All Students



(h) All 65-plus

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (g) are coloured according to the proportions of households in each constituency of each constitution. Alone refers to a single person household. All Students covers student houses or other accommodation where all residents are students. All 65-plus covers any household where all of the residents are aged 65 or over. Data from Thorsen et al. (2017).

Table 7: Selected Comparisons: Tenure ($\epsilon = 9$)

Number of Cars	Ball A	Grp B	Diff	Std	Ball 10	Ball 3	Diff	Std	Ball 3	Ball 4	Diff	Std
None	31.88	21.08	10.80	0.93	10.06	12.37	-2.39	-0.20	12.37	16.21	-3.82	-0.33
1	43.71	43.75	-0.04	-0.01	36.58	39.51	-2.93	-0.98	39.51	42.20	-2.70	0.90
2	19.47	27.06	-7.59	-0.98	38.80	35.41	3.39	0.40	35.41	31.37	4.05	0.53
3	3.81	6.02	-2.21	-0.96	10.27	9.04	1.23	0.54	9.04	7.43	2.61	0.70
4 or more	1.13	2.07	-0.97	-0.93	4.29	3.64	0.65	0.62	3.64	2.79	0.85	0.81
Ball Size	146	282			12	81			81	226		

Notes: Group A contains balls 8 and 13. Group B features balls 5 and 12. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017).

fewer people living alone, more married couples, fewer students and more in the 65-plus range.

Through consideration of household composition it has again been seen that TDA Ball Mapper can help identify combinations of characteristics that run counter to assumed relationships. A tight shape here is evidence of stronger correlation between the variables and the links between the various characteristics and Brexit are understood at a summary level from the lower plots of Figure 7. Understanding how points in the cloud that are within the ball radius produce such different outcomes is a challenge for researchers. In these isolated cases there are many more covariates to introduce that would change the picture greatly. Again it is seen that the Remain vote is spread across the space and the Leave vote is far more concentrated.

5.3 Number of Motor Vehicles

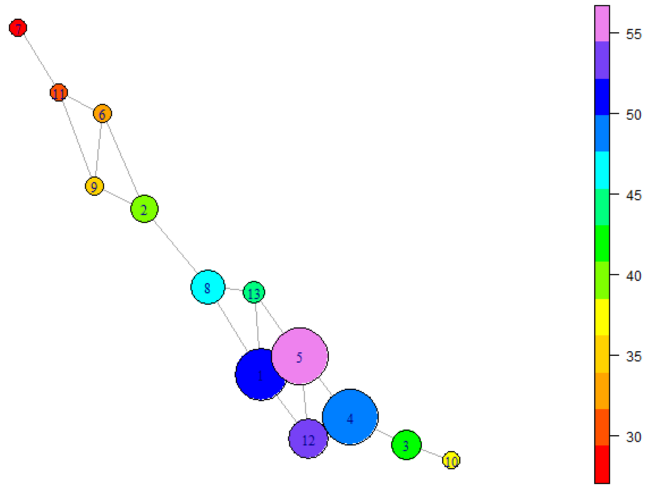
Correlations between the proportions are seen strongly in the TDA Ball Mapper plots of Figure 8. This manifests as an almost straight line through the space with all observations gathered closely to it. Brexit percentages are highest towards the lower end,with a small tail of low percentages extending out at the very extreme. Comparing with the colour by axes plots in panels (b) to (f) the correspondence between balls sees leave votes highest where the average number of cars are between 1 and 3. Percentages rise with the number of cars from the top left of the plot, increasing almost monotonically through the balls to ball 5. From there an effect from some of the richest respondents with multiple cars being remain voting brings the average down again tailing down to the end monotonically. Here the TDA Ball Mapper informs on a “u-shaped” relationship between motor vehicle access and the Leave vote.

Table 7 contains a comparison of the margin at the top end of the Brexit favouring set, and then features two comparisons from the lower tail. Balls 8 and 13 are connected to ball 12, with its leave support on average laying just above 50%. Ball 13 is also connected to ball 5, which has the highest Leave percentage; ball 8 does not connect to 5 however. Averages for the two groups confirm the discussion from Table 1, with the Remain favouring balls 8 and 13 having higher numbers with no cars, similarity on the proportion with one car and then fewer households with all of the higher numbers of cars. Conversely ball 10, with a lower Hanretty (2017) estimated Leave percentage has higher proportions in the highest number of cars than ball 3. In turn ball 3 has higher values in these groups than the larger ball 4 that connects it to the shape. All of the comparisons in Table 7 are consistent with the strong correlation between proportions that have created the narrow shape. Hence all elements of the outcome to characteristic mapping are indicative of the inverted “u-shaped” relationship between proportion of households in a constituency with higher numbers of cars and the leave percentage estimated for that constituency.

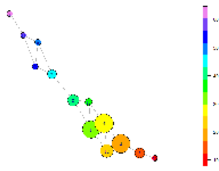
5.4 NSSEC Classification

Nine classifications from the National Statistics Socio-Economic Classification (NSSEC) are used in this analysis, the most of any of the single question studies. To aid the readability of the output a higher value for the radius is used, $\epsilon = 7$, producing 22 balls. Figure 9 shows the balls are all connected, with the constituencies favouring Leave in the middle of the plot. To the top of the plot the estimated leave percentage from Hanretty (2017) falls away. Violating this trend balls 3 and 21 sit close to the dark blue

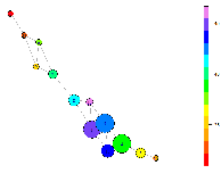
Figure 8: Number of Motor Vehicles ($\epsilon = 7$)



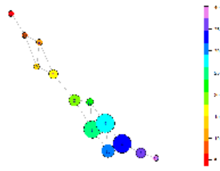
(a) Hanretty (2017) Leave Percentages



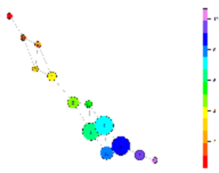
(b) 0



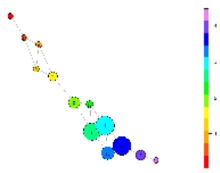
(c) 1



(d) 2



(e) 3



(f) 4 or more

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (f) are coloured according to the proportions of households in each constituency with access to the given number of motor vehicles. Data from Thorsen et al. (2017).

Table 8: Selected Comparisons: NSSEC Classification ($\epsilon = 7$)

NSSEC Class	Ball 21	Ball 7	Diff	Std	Ball 3	Grp A	Diff	Std	Ball 11	Ball 10	Diff	Std
Higher Manager	1.12	1.42	-0.30	-0.34	1.63	1.96	-0.34	-0.39	0.75	1.09	-0.33	-0.38
Higher Pro.	6.17	5.00	1.17	0.36	6.25	5.80	0.45	0.14	3.18	5.85	-2.67	-0.83
Lower Manager	14.49	16.17	-1.68	-0.44	18.03	19.16	-1.14	-0.30	10.92	14.95	-4.03	-1.05
Intermediate	9.84	12.27	-2.44	-1.21	12.37	-12.78	-0.41	-0.20	9.54	9.79	-0.25	-0.13
S Employer	6.17	7.94	-1.27	-0.68	6.73	9.44	-2.71	-1.04	8.01	8.53	-0.52	-0.20
Lower Sup.	5.74	7.19	-1.45	-0.93	7.71	8.04	-0.33	-0.21	5.71	5.33	0.37	0.24
Semi-Routine	12.73	15.92	-3.19	-1.04	15.58	16.32	-0.74	-0.24	14.06	12.44	1.61	0.52
Routine	11.81	13.82	-2.00	-0.56	13.75	13.49	0.25	0.07	13.02	10.10	2.91	0.82
Never Worked	7.71	7.30	0.42	0.18	3.88	3.84	0.03	0.01	17.31	12.28	5.03	2.15
LT Unemp.	2.37	2.49	-0.11	-0.18	2.03	1.87	0.16	0.25	3.18	2.62	0.56	0.87
Ball Size	20	62			49	387			2	10		

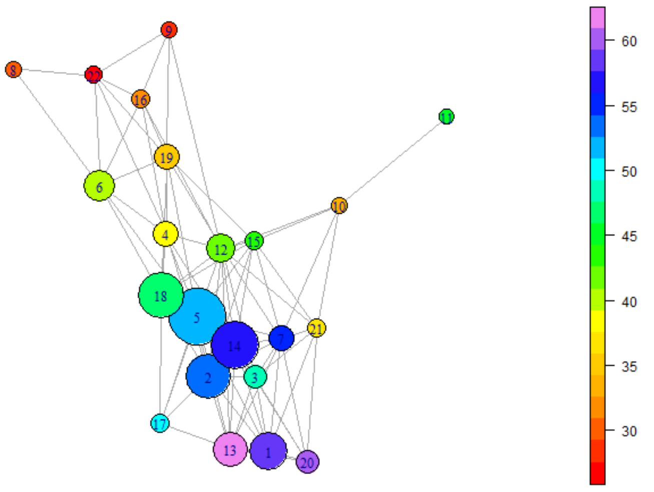
Notes: Pro is used in place of professional, Sup. is used for supervisor. LT Unemp represents Long-term Unemployed. Group A contains balls 1, 2, 5, 10, 14 and 20. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017). NSSEC is a contraction of the National Statistics Socio-Economic Classification

mass, whilst ball 11 towards the top right actually has a higher Leave percentage than ball 10 that connects it to the main shape. Multi-dimensionality can be seen in the way that balls are connected to many more balls than in other plots, consider the number of edges eminating from ball 13 as an example.

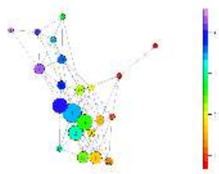
Higher classifications are shown by panels (b) to (c) to lay primarily at the top of the plot, a section shown in panel (a) to have very low estimated Hanretty (2017) leave percentages. Relatively low proportions of these classifications in the overall economy mean that the mauve shading only corresponds to 4.5% in the Higher Managerial group. Moving through the levels shows the intermediate levels, panels (e) and (f) laying in the middle of the shape. Balls here are typically seen as pro-Brexit. Lower Supervisor, Semi Routine and Routine ,panels (g) to (i), lay to the bottom of the plot, these balls are all estimated as having high leave votes. Finally the two categories relating to unemployment, Never Worked in panel (j) and Long-term Unemployed in panel (k) are all at their highest on the right of the plot. From the colouration of panel (a) it would appear the link between these final two dimensions and voting for Brexit is limited. Deeper analysis of the precise patterns is merited.

As a first comparison consider balls 21 and 7, which sit to the immediate right of the central Brexit favouring balls. Ball 21 has a large Remain majority, whilst ball 7 heavily favours Brexit. Such polar opposites connected in the point cloud naturally spark interest. Table 8 shows that Ball 21 has higher proportions of higher professionals and those who have never worked, therein lies little so suggest a particularly different Brexit vote. However, ball 7 does have more of the mid-to-low classifications which have all been associated with Leave voting, including in Table 1. Here it seems demographics in ball 7 are more consistent with accepted Leave motivations, and that hence it is this which creates the observed differential. Secondly ball 3, another Remain set of constituencies, is connected with a large group of Brexit supporters. Group A contains six balls, 1, 2, 5, 10, 14 and 20, all of which had Leave majorities and are connected to ball 3. Again there are a higher proportion of higher professionals in the Remain favouring ball and the leave groups have higher proportions around the middle to lower classifications. However, contrary to the message from the t-tests there are large proportions of the very lowest classifications in Ball 3; this is evidence running counter to the “left behind” voting Leave notion. Ball 11 represents an interesting case stuck out on an arm away from the main shape, and in a strong Remain part of the plot. However, the Leave vote in ball 11 is much higher, almost reaching 50% compared to below 40% in its immediate neighbour ball 10. Here the differences are more consistent with the theory; greater proportions from the lower NSSEC classes appear in ball 11 whilst the higher classes are in ball 10; although the differences are small enough to see the edge form the difference is consistent with the observed voting patterns.

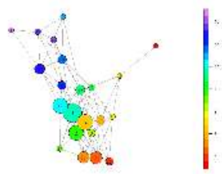
Figure 9: NSSEC Classification



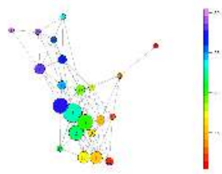
(a) Hanretty (2017) Leave Percentages



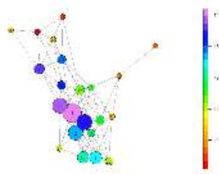
(b) Higher Manager



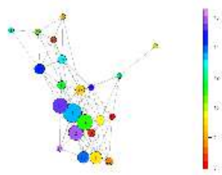
(c) Higher Professional



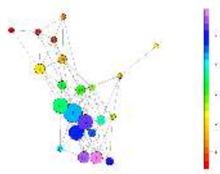
(d) Lower Manager



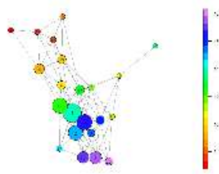
(e) Intermediate



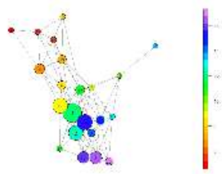
(f) Small Employer



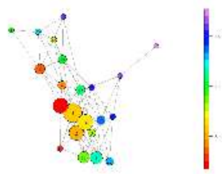
(g) Lower Supervisor



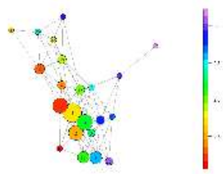
(h) Semi-Routine



(i) Routine



(j) Never Worked



(k) Long-term Unemployed

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (k) are coloured according to the proportions of households in each constituency of each National Statistics Socio-Economic Classification (NSSEC) classification. This question uses ten levels, with the highest group being split into Higher Managerial and Higher Professional accordingly. Data from

5.5 Qualifications

Figure 10 presents the TDA Ball Mapper graph for the four levels of qualification, apprentices, others and the proportions of households having no qualifications. At $\epsilon = 6$ there is a large connected shape with large leave favouring balls at the left hand end. A number of outliers sit above the main shape, but as will be later demonstrated this is only an abstract feature; most are more closely linked to the three green balls on the underside of the big group. As with other plots there is great diversity among the balls with the net average Remain vote. Lower levels of educational attainment have been consistently linked with Leave voting in the literature, but the immediate correspondence from the main plot is with Level 2 and Apprenticeships. These are the levels with significant differences in the t-tests reported in Table 1. A lack of qualifications does apply at the left end of the connected shape among the highest Leave percentages, but a lack of education also appears heavily within the outliers. On the other end of the scale Panel (g) depicts the highest proportions of respondents to the 2011 Census who had at least an undergraduate degree to be in the lower right of the shape. This is a firmly remain area of the characteristic space as depicted in panel (a). Levels 1 and 3 have partial correlations, their inference sitting between Level 2 and then Levels 0 or 4 respectively.

Table 9 reports summary statistics for the 43 balls that appear in Figure 10. Numbering under TDA Ball Mapper is not related to any of the variables, but is used as the sorting factor to better compare quickly with the graphs. Variation in the size of the balls can be seen immediately, with only a few balls covering more than 10% (61) of the constituencies. All those large balls have average Leave percentages well above 50%. Tables like this are useful reference points when seeking to understand the point cloud, but it is the direct comparisons that give the most readily interprable output from the TDA Ball Mapper process.

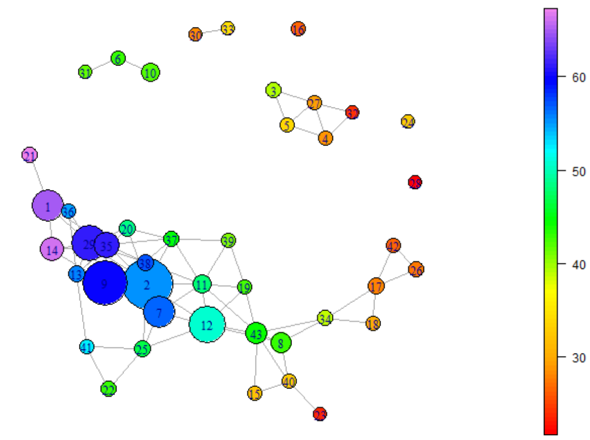
To the lower right end of the main shape there are two arms that end in balls with Hanretty (2017) leave percentages below 30%. Contrasting ball 21 with 26 and 42 the first columns of Table 10 show that the main differences are in levels 3 and 4, the arm down to ball 21 is dominated by level 3, with more graduates in balls 26 and 42. A second comparison is made between ball 12 and two Brexit voting balls, 2 and 7, to which it is attached. On this margin between Remain and Leave there are very few significant differences despite the large difference in the outcomes. The remain favouring ball 12 has a higher proportion of higher qualifications, in keeping with the understood relationship between education level and being pro-EU, but the differences are small relative to others and the proportions at these levels are much lower than those in the first comparison columns. Finally two of the similarly coloured outlier groups are compared. Balls 30 and 33 are termed group C for Table 10 and then the five connected balls (3, 4, 5, 27 and 32) are termed group D. These balls are not connected and it can be seen from the table that the big differential is in the proportion with no qualifications. The smaller group has the bigger proportion of unqualified household heads. Through the three pairings it is possible to get a stronger sense of the overall pattern in the data that is picked up in the subpanels of Figure 10.

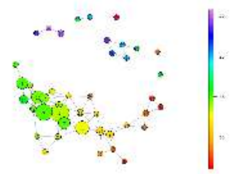
Broad correlation between qualification levels and favouring remain exists across the plot, including the outliers. This is important in a TDA Ball Mapper plot because the placement of the balls is abstract. Using equation (1). Figure 11 presents plots for three of the outliers, namely the string of three balls that have a slight preference to Remain, the cluster of five connected balls that are much stronger Leave voters, and ball 28 which has the highest Leave percentage recorded for any of the balls.

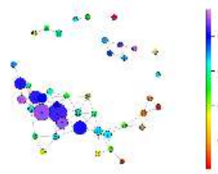
Panel (a) of Figure 11 reveals the string of three balls to be most alike to the group of five and Ball 28; these are the subject of panels (b) and (c). Within the main connected shape it is the marginal balls to the bottom right of the plot that have the greatest similarity to the string of three. Many of the other remain balls that group to the upper side of the connected shape, including 11, 19 and 39, are amongst the most different to the string. Likewise the Remain areas of the space to the right of the plot are also found to be further from the characteristics of the string. Such results underline again the diversity among balls where the member constituencies voted, on average, to Remain in the EU. Panel (b) shows that for the group of five ball 25 is the closest, but again the set of three balls below the main Brexit balls are the ones to which the outliers would attach if radii were expanded enough. In the case of the group of five there is more similarity with the other balls that had average Hanretty (2017) estimated Leave percentages in the 40% to 50% range. Both panel (a) and (b) reveal the heavily Remain ball 23 to be one of the furthest away in terms of characteristics, once more a message of diversity is given.

As a further underlining of this point consider the outlier ball 28. With the highest estimated Leave percentage it might be expected this ball would share characteristics with the left of the connected shape.

Figure 10: Highest Qualification Level ($\epsilon = 6$)





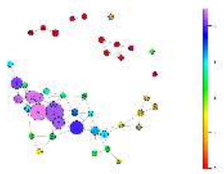


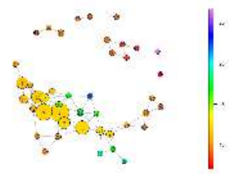
(a)

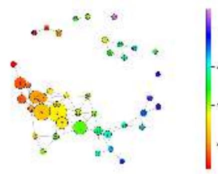
(b) No Qualifications

(c) Level 1

(d)Level 2



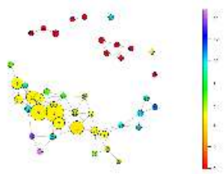




(e) Apprenticeship

(f) Level 3

(g) Level 4



(h) Other

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (g) are coloured according to the proportions of households in each constituency with each qualification level as the highest obtained by a resident therein. Level 1 corresponds to below GCSE level. Level 2 is obtaining a satisfactory level of performance in compulsory education. Level 3 is obtaining two or more A-levels or equivalent. Level 4 is an undergraduate degree or above. Data from Thorsen et al. (2017).

Table 9: Summary Statistics for Qualification Coverage ($\epsilon = 6$)

Ball	None	Level 1	Level 2	Apprentice	Level 3	Level 4	Other	Leave (%)	Size
1	31.49	15.09	15.09	4.06	11.29	16.92	4.98	64.48	88
	(2.31)	(0.88)	(0.8)	(0.76)	(0.83)	(1.95)	(1.22)		
2	22.67	13.74	13.74	4.19	12.13	26.37	4.55	55.34	181
	(2.13)	(0.98)	(0.89)	(0.72)	(0.88)	(2.47)	(0.87)		
3	24.15	23.55	23.55		9.53	27.88		39.29	13
	(1.66)	(1.45)	(1.06)		(0.96)	(2.12)			
4	17.98	19.92	19.92		8.85	38.35		28.91	4
	(0.95)	(1.76)	(1.23)		(1.02)	(1.68)			
5	19.48	21.46	21.46		9.21	34.18		33.95	4
	(1.33)	(1.74)	(1.54)		(1.04)	(1.38)			
6	33.68	25.61	25.61		10.44	17.05		42.96	8
	(2.13)	(1.29)	(0.8)		(0.85)	(1.31)			
7	20.94	14.77	14.77	4.17	12.34	25.93	4.93	56.31	87
	(1.66)	(1.16)	(0.72)	(0.76)	(0.83)	(2.16)	(1.09)		
8	15.47	11.22	11.22	2.98	11.94	39.04	4.79	42.81	36
	(1.35)	(0.9)	(1.09)	(0.56)	(1.13)	(2.33)	(1.19)		
9	24.29	14.86	14.86	4.43	12.18	22.65	4.68	59.65	152
	(2.05)	(1.25)	(0.72)	(0.73)	(0.73)	(2.2)	(0.83)		
10	29.86	24.97	24.97		10.24	21		41.87	26
	(2.46)	(1.78)	(0.79)		(1.11)	(2.16)			
11	20.39	11.75	11.75	3.28	17.58	27.69	5.31	47.92	27
	(1.82)	(1.12)	(1.26)	(0.83)	(2.36)	(1.85)	(1.16)		
12	18.69	12.94	12.94	3.88	12.35	31.77	4.28	50.63	110
	(1.74)	(1.03)	(0.88)	(0.76)	(0.96)	(2.35)	(0.8)		
13	26.86	14.59	14.59	2.95	10.91	20.76	9.42	55.7	15
	(2.88)	(0.67)	(0.97)	(0.9)	(0.95)	(1.65)	(1.68)		
14	28.63	16.18	16.18	4.27	11.5	17.43	4.94	66.1	48
	(2.13)	(1.06)	(0.65)	(0.7)	(0.71)	(1.64)	(0.99)		
15	13.92	9.51	9.51	2.26	18.6	38.61	5.87	32.11	4
	(1.29)	(0.91)	(1.24)	(0.42)	(2.15)	(1.21)	(1.33)		
16	10.34	6.22	6.22	1.01	9.52	54.87	9.62	26.83	6
	(0.68)	(0.44)	(1.04)	(0.24)	(0.75)	(1.33)	(1.85)		
17	15.86	8.54	8.54	1.05	10.14	45.03	9.85	27.84	16
	(2.22)	(0.84)	(1.14)	(0.17)	(1.45)	(2.88)	(1.52)		
18	18.91	9.64	9.64	1.12	10.08	40.21	9.54	30.48	7
	(1.25)	(0.73)	(1.55)	(0.31)	(0.87)	(1.9)	(0.47)		
19	20.33	11.02	11.02	2.79	16.03	31.67	4.93	42.28	8
	(1.32)	(0.73)	(1.4)	(0.82)	(2.69)	(1.58)	(1.15)		
20	25.86	13.2	13.2	2.79	11.56	25.9	6.5	48.31	14
	(1.39)	(0.99)	(1.36)	(0.93)	(1.34)	(1.94)	(1.92)		
21	37.16	15	15	2.93	9.88	13.53	6.43	67.07	9
	(1.67)	(0.61)	(0.73)	(0.67)	(0.66)	(1.01)	(1.64)		
22	20.22	11.69	11.69	1.4	9.85	30.34	14.66	43.04	12
	(1.32)	(1.08)	(0.93)	(0.36)	(0.32)	(2.01)	(1.67)		
23	12.31	7.05	7.05	1.46	18.36	46.35	5.34	23.96	3
	(2.25)	(0.26)	(0.51)	(0.42)	(1.86)	(1.14)	(1.06)		
24	14.51	8.19	8.19	2.07	26.56	33.12	4.86	33.11	3
	(1.04)	(0.49)	(1.32)	(0.65)	(0.95)	(0.46)	(1.47)		
25	19.11	13.27	13.27	2.32	10.97	31.06	8.89	46.46	14
	(1.45)	(1.12)	(1.35)	(0.75)	(0.86)	(1.66)	(2.83)		
26	12.86	7.27	7.27	0.98	9.78	49.2	11	27.77	10
	(1.41)	(0.72)	(0.81)	(0.17)	(0.73)	(2.33)	(1.63)		
27	19.2	19.72	19.72		8.73	35.83		29.33	7
	(1.21)	(1.41)	(1.38)		(0.99)	(2.03)			
28	14.57	15.67	15.67		7.15	48.21		21.97	2
	(0.16)	(0.2)	(2.93)		(1.05)	(1.84)			
29	28.02	14.43	14.43	4.34	11.78	20.5	4.66	61.2	107
	(1.99)	(0.76)	(0.78)	(0.72)	(0.75)	(1.95)	(0.92)		
30	25.84	17.04	17.04		9.7	33.01		28.47	2
	(1.6)	(2.59)	(1.96)		(0.02)	(2.2)			
31	40.77	23.7	23.7		8.85	15.23		41.79	3
	(2.75)	(1.7)	(0.72)		(0.61)	(2.17)			
32	18.73	15.68	15.68		8.69	39.42		24.24	2
	(1.17)	(3.02)	(1.54)		(1.32)	(1.63)			
33	29.03	19.62	19.62		9.33	29.03		34.53	3
	(1.78)	(0.88)	(0.68)		(0.35)	(2.52)			
34	16.18	10.65	10.65	1.71	10.87	38.68	9.4	38.8	13
	(1.49)	(0.83)	(1.19)	(0.53)	(1.01)	(2.04)	(2.34)		
35	27.97	14.31	14.31	4.3	12.5	20.04	4.96	61.01	58
	(1.52)	(0.81)	(0.99)	(0.83)	(1.38)	(1.48)	(1.18)		
36	31.98	14.16	14.16	2.53	10.37	18.3	8.86	54.71	7
	(1)	(0.61)	(0.94)	(0.65)	(0.75)	(1.36)	(1.67)		
37	23.87	11.84	11.84	2.49	16.53	25.6	6.87	44.93	13
	(2.17)	(0.72)	(0.99)	(0.91)	(2.29)	(1.53)	(1.63)		
38	22.61	13.43	13.43	4.21	15.33	23.45	4.97	56.83	15
	(1.91)	(1.19)	(1.18)	(0.63)	(2.58)	(1.43)	(0.73)		
39	20.94	9.85	9.85	2.35	21.64	28.14	5.22	40.69	7
	(1.03)	(0.89)	(1.25)	(0.76)	(2.93)	(1.65)	(1.28)		
40	13.47	8.78	8.78	2.33	16.72	42.14	4.33	33.14	5
	(0.91)	(1.05)	(1.83)	(0.73)	(2.09)	(2.08)	(0.97)		
41	21.2	13.88	13.88	2.1	10.22	25.93	13.61	52.3	6
	(1.18)	(1.28)	(1.01)	(0.71)	(0.69)	(2.78)	(1.56)		
42	12.45	7.55	7.55	1.18	10.34	50.42	8.32	26.92	5
	(1.01)	(0.93)	(1.25)	(0.37)	(0.88)	(1.83)	(1.52)		
43	16.55	11.45	11.45	3.14	12.58	36.79	4.86	44.15	40
	(1.32)	(0.82)	(1.18)	(0.68)	(1.69)	(2.28)	(1.19)		

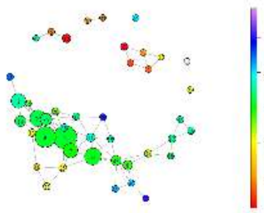
Notes: Ball numbers correspond to those in Figure 10. Figures reported are mean proportions of the Census 2011 respondents for a constituency who had each qualification level. Gaps indicate that there are no constituencies within the ball who have a value for this particular variable. This happens primarily due to the classification of qualifications in Scotland. Leave (%) is the average Hanretty (2017) estimated Leave percentage across constituencies within the ball. Size reports the number of constituencies within the ball. Data from Thorsen et al. (2017).

Table 10: Selected Comparisons: Highest Qualification Level ($\epsilon = 6$)

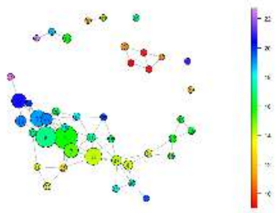
Level	Ball 21	Grp A	Diff	Std	Ball 12	Grp B	Diff	Std	Grp C	Grp D	Diff	Std
None	12.31	12.86	-0.55	-0.10	18.69	22.37	-3.68	-0.63	27.95	21.90	6.05	1.04
Level 1	7.05	7.47	-0.42	-0.11	12.94	13.95	-1.01	-0.27	18.52	21.83	-3.32	-0.90
Level 2	9.13	9.26	-0.13	-0.06	16.10	16.42	-0.32	-0.15	13.70	15.43	-1.73	-0.79
Apprentice	1.46	1.07	0.40	0.26	3.88	4.19	-0.31	-0.20	0	0	0	0
Level 3	18.36	9.92	8.44	3.45	12.35	12.18	0.16	0.07	9.42	9.29	0.13	0.05
Level 4	46.35	49.09	-2.75	-0.33	31.77	26.26	5.50	0.66	30.41	31.55	-1.14	-0.14
Other	5.34	10.33	-4.99	-1.81	4.28	4.63	-0.35	-0.13	0	0	0	0
Ball Size	3	12			110	201			4	22		

Notes: Group A contains balls 42 and 26. Group B contains balls 2 and 7. Group C contains balls 30 and 33. Group D contains balls 3, 4, 5, 27 and 32. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017).

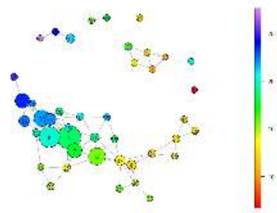
Figure 11: Distance from Outliers: Qualifications with $\epsilon = 6$



(a) Balls 6, 10 and 31



(b) Balls 3, 4, 5, 27 and 32



(c) Ball 28

Notes: TDA Ball Mapper graphs constructed using the six axes summarised in Figure 10 with $\epsilon = 6$. Colouration is according to the average distance from the points within any given ball to the average coordinates of the balls named below the panels. Distances in each dimension are normalised by the population standard deviation following equation (1). Colouration is by the total number of standard deviations difference across the full set of variables. Data from Thorsen et al. (2017).

Table 11: Selected Comparisons: Deprivation Level ($\epsilon = 5$)

Level	Ball 13	Ball 5	Diff	Std	Grp A	Grp B	Diff	Std	Ball 15	Ball 1	Diff	Std
Level 0	46.46	51.29	-4.83	-0.69	49.21	34.90	14.31	2.05	27.23	32.24	-5.01	-0.72
Level 1	32.36	31.11	1.25	0.72	31.85	34.04	-2.18	-1.25	30.08	31.72	-1.65	-0.94
Level 2	17.02	14.42	2.60	0.15	15.27	22.42	-7.15	-1.78	28.55	26.29	2.26	0.56
Level 3	3.82	2.91	0.90	0.41	3.33	7.70	-4.37	-1.97	12.68	8.97	3.71	1.67
Level 4	0.35	0.27	0.09	0.26	0.34	0.95	-0.61	-1.82	1.47	0.78	0.69	2.08
Ball Size	206	126			199	93			5	47		

Notes: Group A contains balls 3, 12 and 13. Group B contains balls 7, 8 and 14. Diff reports the difference in percentage points between the two means. Std divides this difference by the standard deviation of the variable from the full dataset. Size is a measure of the number of constituencies within the ball(s). Again it must be recalled that the existence of a connection between these balls means there must be points in the intersection that are counted in the averages for both balls. Data from Thorsen et al. (2017).

However, as panel (c) reveals, it is in fact ball 43 which is closest. Further, the right hand part of the shape is more similar to ball 28 than is the left half. By looking deeper into the constituencies that are in ball 28 it can be found that the average proportion with level 4 qualifications is 46.2% and that this is significantly above the overall average for residents with degrees (22.6%). By all measures this should be a Remain constituency pair, more research into this ball is needed.

In the study of qualifications the functionality of *BallMapper* has been exploited further to reiterate the point that often it is the surprise results that inform the most about overall outcome patterns. Here the outliers often connected via parts of the main shape where the Brexit voting behaviours were very different. Such does not explain the fact the points are not connected, but does cast further light on the challenges presented by the desire to see monotonic relationships between characteristics and outcomes.

5.6 Self Reported Health

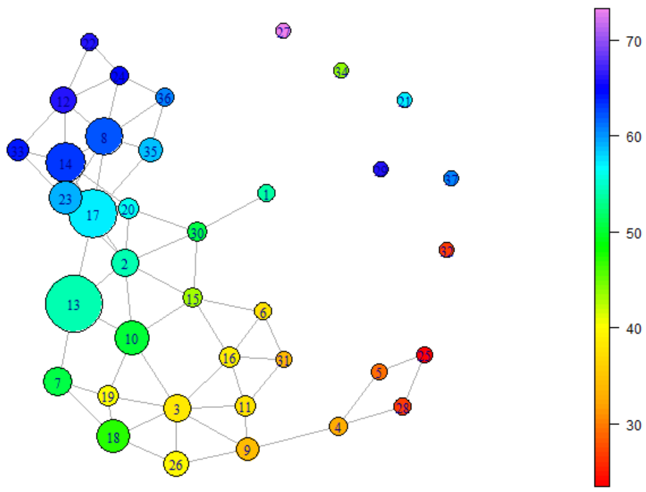
TDA Ball Mapper with $\epsilon = 2$ shows a dominance of Brexit supporting constituencies within a tight top area of a large connected component. Through the middle there are a number of marginal balls that then interlace with the Remain supporting constituencies to the bottom right of the plot. There are six outliers with varying degrees of estimated (Hanretty, 2017) Leave percentages. Panels (b) to (f) show the strong links between very good self-reported health and voting behaviour. Those red and yellow balls, with Leave percentages below 40% are all shown to be dominated by very good health. Good levels are reported in the groups to the left of the plot, including some which sit in the middle of the voting range. There are only a few in the Census 2011 data who report as lower levels of health and so the scales on panels (d), (e), and (f) do not extend as far. High levels are seen more into the outliers and in the extremes of the Brexit favouring region of the plot. Through the exploitation of the axes plots it is apparent that outliers will connect either to the top of the shape or, in the case of ball 32, to the tail of the connected shape. As the relationship is more monotonic there are a few interesting comparisons to run and the analysis of self-reported health is kept short.

5.7 Deprivation Level

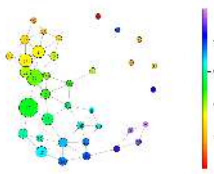
Deprivation is studied with $\epsilon = 5$ resulting in a highly packed shape with just a sole outlier. Remain can be seen through the top and left of the plot with Brexit favouring constituencies running down the right hand side of the shape. There is a segregation between the Remain favouring balls 3, 12 and 14 at the top and 7, 8 and 14 in the lower cetnre. Colouration by axes in panels (b) to (e) identifies the balls with high proportions of zero deprivation households to be towards the upper part of the plot, whilst the lower set is linked more to deprivation Levels 1 and 4. In the summary statistics Level 2 was strongly linked to Brexit voting and in Balls 1, 4, 6 and 10 the correspondence is clear. These balls are also high in either Levels 3 or 4. This plot features a number of cases where the colouration foes not fully align with the average message comming from the summary statistics. Ball 5, dominated by Level 0 and amongst the lowest average values of Level 2 has a Leave percentage of 52%. TDA Ball Mapper works very effectively to identify such contradictions.

As a first comparison consider balls 13 and 5 which sit at the top of the main shape and have Leave

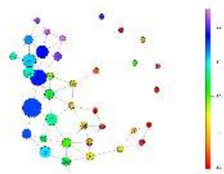
Figure 12: Self-Reported Health



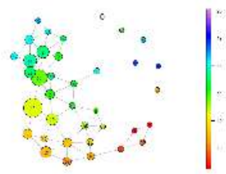
(a) Hanretty (2017) Leave Percentages



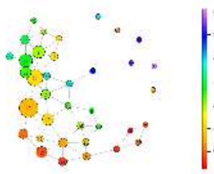
(b)Very Good



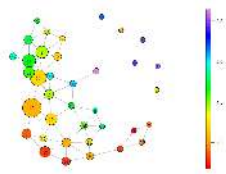
(c) Good



(d) Fair



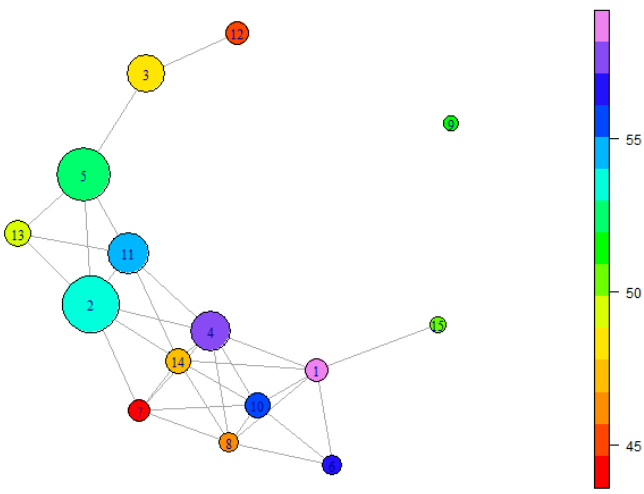
(e) Bad



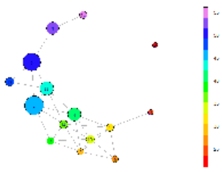
(f) Very Bad

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (g) are coloured according to the proportions of individuals in each constituency with each self-reported health level. Data from Thorsen et al. (2017).

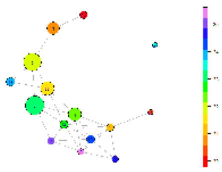
Figure 13: Deprivation Levels ($\epsilon = 5$)



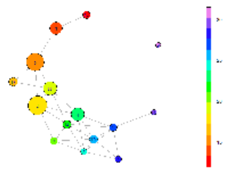
(a) Hanretty (2017) Leave Percentages



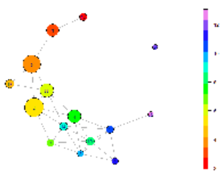
(a) Level 0



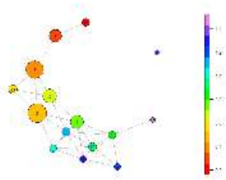
(b) Level 1



(c) Level 2



(d) Level 3



(e) Level 4

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Panel (a) has colouration by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of dark green. Panels (b) to (g) are coloured according to the proportions of households in each constituency with each reported level of deprivation. Deprivation is qualified across four dimensions, Employment, Education, Disability and Housing. The numbers represent the proportions of households recording each total across these four. Information is not provide upon which dimensions households are classified as deprived. Data from Thorsen et al. (2017).

percentages either side of the 50% margin. Table 11 shows Remain favouring ball 13 has a greater proportion of residents classed as deprived on one or two of the measures, whilst ball 5 has a higher proportion that are not deprived by any measure. This contrasts with the suggestion from Table 1 and is selected for analysis here to evidence again the ability of TDA Ball Mapper to highlight cases that would have otherwise been missed. With both groups containing more than 100 constituencies this a non-trivial example. Secondly, a group of Remain voting constituencies from the top of the plot are contrasted with a set from the lower centre. Formally Group A (balls 3, 12 and 13) is compared with Group B (7, 8 and 14). The former group has much lower levels of deprivation, this building on the first comparison. Indeed the surprise from the first comparison was the presence of a Brexit favouring ball 5 amongst these low levels of deprivation. When contrasted with the strongest Remain constituencies from Group B it is clear that the approach is correctly segregating two very different sets of Remain voters. Finally ball 15 our on the lower arm is contrasted with the Brexit voting ball 1. Ball 15 is shown in Table 11 to have much higher levels of deprivation than any of the other balls, including the Leave voting ball 1. That the average vote in ball 15 is for Remain, despite these high deprivation levels is the biggest challenge to the “left-behind” hypothesis within the data. It must be stressed that there are only five constituencies in ball 15 and more work would be needed to understand why these particular constituencies behaved as they did.

5.8 Summary

Through consideration of data from seven questions from the 2011 census, as aggregated at the Parliamentary constituency level, systematic evidence of three key benefits of TDA Ball Mapper has been provided. Firstly, the lack of monotonicity in relationships between characteristics and outcomes is identified strongly; this brings into immediate question the practice of treating characteristics separately rather than as part of a bigger dataset. Secondly, there is a consistent message that the Leave vote is diversely spread across parameter spaces rather than being concentrated as the Brexit support is. In many cases tests have revealed just how spread across the axes the Remain voters are. Finally, the ability to get a quick oversight of the data in a way that can be quickly interpreted is exploited in every section. This ability is a strong bonus over two axis, and three axis, comparisons to which established visualisations are limited.

From the plots it is not possible to infer causality, but given the strong correlations in many of the variables it also holds that there would be limited validity to causality derived through linear regression. TDA Ball Mapper is a tool to inform the practitioner, derive understanding of a complex point cloud, and determine directions for further investigation. Taking the advantages of the approach to full dataset yet greater appreciation of the role of demographic combinations is gained. Attention now turns to such generalisations.

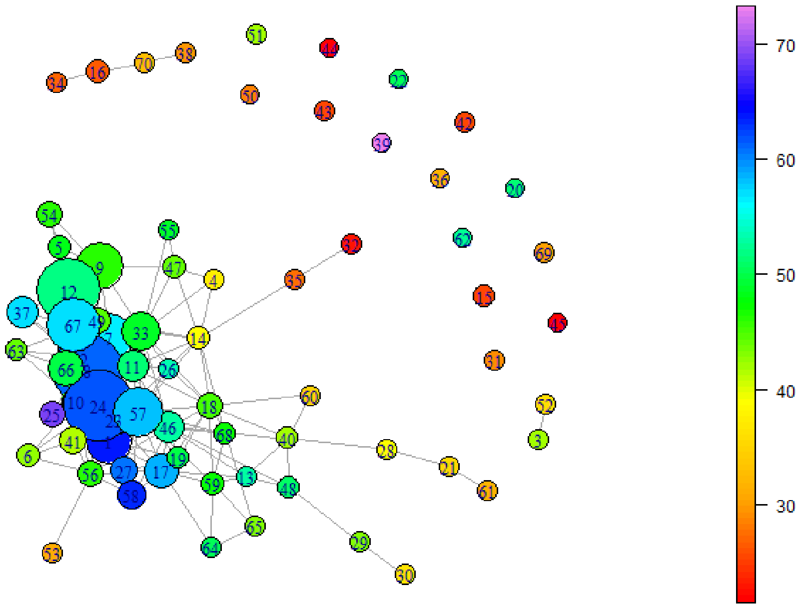
6 Full Census 2011 Question Set

Breaking down by question is informative on the detailed patterns within the data, but the 2016 referendum results are the consequence of the combination of all of the factors. In this regard full refers to using all seven of the questions discussed question-by-question in the analysis. TDA Ball Mapper is a big data approach and can readily extend to the 45 values that have been studied in the previous section and beyond. Through a consideration of all 45, then only those where the average proportion is 20% or higher attention is now given to that combined effect. Regional variations are also explored using the reduced dataset. In each case the ability to visualise data in two dimensions is invaluable in evidencing the spread of the Remain vote in comparison to the commonality of constituencies in the Leave set.

6.1 All Variables

As a first exercise on utilising the whole dataset to understand the links between constituency characteristics and the 2016 Brexit Referendum result a TDA Ball Mapper plot is estimated. The large number of variables means that a larger ball radius is needed, here $\epsilon = 18$ is selected as it maintains a number of interesting sets that stem out from the central core. Increasing ϵ will reduce the number of balls, and see a number of the outliers connect. Results of an exercise on radius choice on the full set are not reported here for brevity. Figure 14 shows the resulting TDA Ball Mapper plot with the big block of large blue coloured balls clearly

Figure 14: Full Dataset: $\epsilon = 18$



Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Colouration is by Hanretty (2017) estimated leave percentages with the 50% cut off being at the upper end of the green shading. Data from Thorsen et al. (2017).

located in the centre of the connected shape. Surrounding this paler blue balls, with leave percentages in the mid 50's are seen. Marginal constituencies then surround this before the expanse of Remain favouring balls. Amongst the anti-Brexit set there are a series of arms that reach out in the point cloud. A number of interesting comparisons thus suggest themselves as it is sought to understand how precisely these shapes derive.

Table 12 presents three comparisons based upon the TDA Ball Mapper graph in Figure 14. For brevity only those characteristics where the difference between the balls is greater than one standard deviation are included in the table. First consideration is given to the string of two red balls that extend from the connected set in the centre of the figure. These balls, numbers 32 and 35 have average Hanretty (2017) leave percentages below 30% and connect into the shape through ball 14, which has an estimate Leave vote of just below 40%. Following the discussion of tenure the two strongest remain constituency balls have a lower outright ownership, and higher incidence of private rental, than Ball 14. Differences on household characteristics are found to be large for the "Other" classification with both balls having proportions living in this type well above the national average. As noted, those in the highest NSSEC classes are found in greater numbers in Remain voting constituencies, and this is seen in the larger numbers in balls 32 and 35 in comparison to ball 14. Likewise, higher levels of qualification are more strongly associated with Remain voting; the higher proportion of Level 4 qualified in balls 32 and 35 is in concordance with this. This snapshot fits the narratives picked up from the full analysis of each question, reminding too that there are a large number of characteristics for which these balls are very similar.

A second comparison is drawn between ball 53 and ball 56, which is a pair sitting to the lower left of the connected set. Despite being of similar referendum result to the previous chain there is no connection

Table 12: Selected Comparisons Full Dataset

Question	Characteristic	First Set	Second Set	Difference	Std
Panel (a): Balls 32 and 35 (4 Constituencies) versus Ball 14 (9 Constituencies)					
Tenure	Owned Outright	20.64	28.93	-8.29	-1.06
	Private Rental	31.57	23.21	8.36	1.30
Household Constitution	Other	17.57	12.21	5.36	1.28
NSSEC Status	Higher Professional	14.67	10.36	4.31	1.33
Qualifications	Level 2	9.20	12.67	-3.47	-1.49
	Level 4	44.20	34.24	9.96	1.20
Panel (b): Ball 53 (2 Constituencies) versus Ball 56 (16 Constituencies)					
Household Constitution	Alone	42.22	35.01	7.21	1.75
Deprivation	Level 4	1.22	0.78	0.44	1.33
Panel (c): Ball 14 (9 Constituencies) versus Ball 26 (5 Constituencies)					
Household Constitution	All Students	3.56	1.93	1.63	1.28
	All 65-plus	0.26	0.38	-0.12	-1.38
NSSEC Classification	Semi-Routine	11.65	14.74	-3.09	-1.00
Qualifications	Level 2	12.67	15.72	-3.05	-1.4
	Level 3	18.42	15.29	3.13	1.28
Self-Reported Health	Very Good	51.01	46.83	4.18	1.04
	Fair	11.35	13.32	-1.97	-1.03
Deprivation	Level 1	31.76	33.50	-1.74	-1

Notes: Ball numbers relate to those in Figure 14. All means are reported as percentages within that particular question. For full details of the characteristics see Table 1. NSSEC is the National Statistics Socio-Economic Classification. For brevity only those comparisons where there is an absolute difference between the balls of one standard deviation, or greater, for any given characteristic are included. Data from Thorsen et al. (2017).

between the two. Only two differences of more than one standard deviation are found, in the proportion of households with just one person living there and in the proportion of households classified with all four types of deprivation. In ball 53 there are more living alone and more high levels of deprivation than in ball 56. Ball 53 has the lower leave percentage, running contrary to the idea that it was always the poorer parts of the UK that voted for Brexit. It is worth noting that there are only two constituencies in ball 53, but this does highlight how the TDA Ball Mapper representations preserve the information in the dataset.

A third comparison, which has more variation seeks to understand the contrast moving closer to the common mass of Leave constituencies, is made between ball 14 and ball 26. The latter has a Hanretty (2017) estimated leave percentage of around 55%, around 15% higher than ball 14. In the contrast there are eight characteristics for which the differential is more than one standard deviation. Ball 14 has more students, 3.56% of households being student only, compared to just 1.93% in ball 26. A reverse is seen for households where all residents are of retirement age, although the absolute percentages are much smaller (0.26 in ball 14 and 0.38% in ball 26). In keeping with the summary statistics the proportion of residents who are classed as having semi-routine jobs under the NSSEC system is higher in ball 26, reaching almost 15% of the population compared to just 11% in ball 14. Qualification levels are higher in the Remain constituencies, more having achieved A-Levels than have in ball 26. Health is also higher in ball 14, the proportion reporting very good health being more than four percentage points higher. This transition from 14 to 26 can be contrasted with that from balls 32 and 35 to 14 as a way of understanding how the TDA Ball Mapper representation in Figure 14 faithfully represents the topology of the space.

There are many other considerations that could be made, including the reasons why there are so many arms extending off the main connected shape or what it is exactly that binds the Brexit voting balls so tightly. Regional variations and the role of voting patterns could all be introduced into the set of colouration variables here. TDA Ball Mapper is a tool for understanding the shape of the point cloud and informing about the variation of other properties across that space. There are though a number of challenges presented by a dataset like the one used here that add qualification to the results. Hence this paper proceeds with a limited set of characteristics for the next steps of analysis.

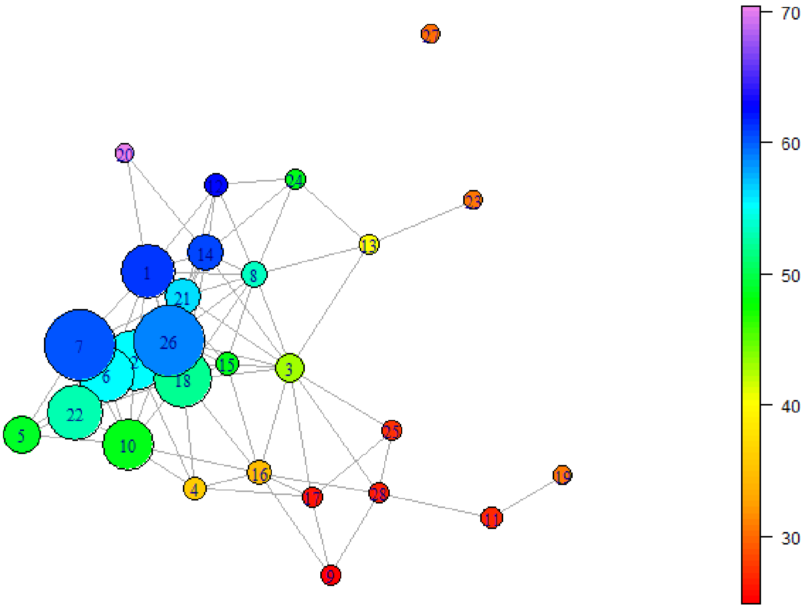
6.2 Reduced Variable Set

A challenge of dealing with the very low proportions is that they will often create connections as the balls encompass their full range long before the other characteristics. This can be circumvented using normalisation, but that then risks other complications. A natural extension of the analysis is to drop those with low proportions from the dataset and re-estimate the TDA Ball Mapper coverage. This is done with a cut off of 20% on average. So doing leaves outright ownership and ownership on a mortgage as the only two tenure characteristics. Single person households and married couples are the only household constitutions included. Car access is either to 0, 1 or 2 cars, meaning that the very richest are excluded on this measure. Only Lower Managerial of the NSSEC classifications has sufficient proportions to be included in this reduced set. From the qualifications set having no qualifications and having achieved Self reported health levels of very good and good are included, as are the lowest two deprivation levels. Amongst this set are those positively associated with the Leave vote and many, like qualification level 4, which are negatively associated.

Figure 15 provides a TDA Ball Mapper graph with $\epsilon = 18$ showing an almost heart shaped plot with lines sticking out from the two halves. There is also a heavily Brexit favouring ball, number 20, sticking out on the top. The scale shows 50% to be in the upper end of the green meaning that all Leave constituencies are in the upper left part of the plot. It is also immediate that the biggest balls correspond to those voting to leave the EU, and that those wishing to remain part of the union are more spread out. That the Brexit favouring constituencies are more similar than others comes through strongly in the plot.

As a further exercise consider the balls organised along the upper line, this has ball 23 at it's tip, ball 13 at the top of the heart and ball 8 immediately inside the main shape. Ball 23 has a Hanretty (2017) estimated leave percentage of less than 30%, whilst ball 13 is closer to 40% and for ball 8 the figure is above 50%. Within this short range the prediction has risen considerably and hence the characteristic changes along this line have interest. Picking out those comparisons where the difference is more than one standard deviation reveals two common transition. Ball 23 has fewer households inhabited by married couples than ball 13, which in turn has fewer than ball 8; this is fully in keeping with the positive association between marriage and the Brexit vote. Ball 23 has less households classed as deprived on just one of the measures

Figure 15: Reduced Set ($\epsilon = 18$)



Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Colouration is by Hanretty (2017) estimated leave percentages with the 50% cut off being towards the upper end of the green shading. Data from Thorsen et al. (2017).

Table 13: Selected Comparisons Reduced Set ($\epsilon = 18$)

Question	Characteristic	Grp A	Grp B	Diff	Std	Ball 20	Grp C	Diff	Std
Tenure	Owned Outright	13.78	15.87	-2.09	-0.27	39.69	29.79	9.90	1.26
	Owned with Mortgage	18.17	21.29	-2.12	-0.37	28.19	32.79	-4.60	-0.81
Household Composition	Alone	37.73	36.64	0.89	0.22	31.79	31.42	0.37	0.09
	Married	19.88	21.41	-1.53	-0.27	30.40	31.82	-1.42	-0.25
Number of Cars	None	52.72	52.00	0.72	0.06	27.08	29.78	-2.70	-0.23
	One	36.77	37.39	-0.62	-0.21	45.24	43.11	2.13	0.71
	Two	8.72	8.82	-0.10	-0.01	21.03	21.43	-0.40	-0.05
NSSEC Status	Lower Managerial	14.62	24.67	-10.05	-2.63	15.62	17.20	-1.58	-0.41
Qualifications	None	23.71	16.81	8.90	1.53	36.05	29.76	6.29	1.08
	Level 4	27.48	45.60	-18.12	-2.18	14.02	18.76	-4.84	-1.21
Self-Reported Health	Very Good	48.24	53.77	-5.53	-1.38	39.13	44.00	-4.84	-1.21
	Good	32.67	31.06	1.61	0.76	30.14	36.01	-5.87	-0.84
Deprivation	Level 0	25.97	41.13	-11.16	-1.6	30.14	36.01	-5.87	-0.84
	Level 1	35.40	33.50	1.90	1.09	33.94	32.42	1.52	0.87

Notes: Ball numbers relate to those in Figure 14. All means are reported as percentages within that particular question. For full details of the characteristics see Table 1. Difference reports the difference in the two means and Std is that difference divided by the standard deviation of the characteristic from the whole population. Grp A comprises balls 13 and 23 on the upper arm of the main connected shape from Figure 15. Grp B comprises those balls on the left of the main connected shape from Figure 15; balls 9, 11, 17, 19, 25 and 28. Grp C comprises the two large Brexit favouring balls 1 and 14. NSSEC is the National Statistics Socio-Economic Classification. Data from Thorsen et al. (2017).

than ball 13, but ball 13 also has more than ball 8 on this measure; a non-linearity is captured. Because these are proportions having more of one characteristic from within a question simply means that other characteristics must have less, but the information in a reduced dataset does not tell us more about where the other changes manifest. The question by question plots tell more about relationships within a particular question. Ball 23 also has more people living alone and less who self report as having very good health than ball 14. Ball 13 has fewer people who own their own home through a mortgage and fewer households with just one car than ball 8. All of these are consistent with the message on association with Leave, except for the health comparison; in the general discussion it was seen that an increasing proportion of Very Good was associated with reduced Brexit support; the comparison of 23 and 13 runs in the opposite direction. Such goes someway to explaining why 23 is not connected with other Remain constituencies.

6.3 Reduced OLS Model

Thus far it has been seen how TDA Ball Mapper may produce informative plots that link the collective characteristics (in this case, the characteristics of various constituencies) with an outcome of interest (in this case, the result of the Brexit referendum). The major selling point being the ability to recognise all interaction terms without reducing the degrees of freedom. Such reductions make it hard to create an ordinary least squares (OLS) model that recognised such interactions[5] Here by focusing on the reduced dataset the issues of multicollinarity are reduced[6] and hence it is possible to create an OLS model to compare against the TDA Ball Mapper insights.

For comparison the following model is estimated:

$$LH_i = \alpha + \beta X_i + \psi_i \tag{2}$$

Where LH_i is the estimated Hanretty (2017) Leave vote in constituency i. X_i is a row vector of characteristics which is multiplied by the parameter vector β. Finally ψ_i is an iid error term with mean 0 and constant variance. The elements in X_i are those constituency characteristics listed in Table 13 from the previous section. Fitting this model it is possible to calculate the fitted residuals $\hat{e_i}$ and hence the absolute fitted residuals $abs(\hat{e_i})$. Table 14 reports the estimates from the model.

Table 14 reveals both models have very large constants. Thus to bring the predicted shares into the range observed it is necessary for the effects of the covariates included to be linked to Remain voting. Indeed, almost all of those coefficients which are significant are negative. Two exceptions in Model 1 are the proportion of households in a constituency where the respondent is married and the proportion of households in a constituency recording a deprivation level of 0. Because of the challenges of multicollinearity in model 1 time is not spent on interpreting the coefficients, save to say that they have the expected signs in almost every case. Model 2 by contrast only uses one variable from each question meaning that it has less information but does not suffer from any statistical issues in its construction. It also omits the Tenure question because of the strong correlation with household composition.

Using TDA Ball Mapper it is possible to see where exactly in the space the fit is good and bad. This is done using colouration using any variable which captures fit. The simplest such variable is the residual, or the absolute residual. Appraisal of the linear models can now be taken back to the TDA Ball Mapper graphs on the reduced data set.

In these plots the shading is performed using the proportion of observations within any given ball which have estimated absolute residuals $abs(\hat{e_i})$ which are greater than 2, 4 and 6. Given the margin of the Referendum was less than 4% residuals of these sizes are of interest. In both Model 1 and Model 2 the largest absolute residuals are bigger than 15% and come from the difficulty both have in predicting the strong Remain sentiment in ball 19. All 6 panels of Figure 16 colour ball 19 as mauve.

Figure 16 reveals that consistently the best fit is to the Leave, and marginal, constituencies to the lower left of the plot. Leave balls in the centre are well fitted, but not quite to the same level. As these are the largest balls it is not surprising that the OLS model does well in these balls such a high R-squared is obtained. However, such a strong fit in a small set of the parameter space is indicative that the model is

[5]A model with four independent variables has six pairwise interactions, three three-way interactions and one four-way. In total there would be an additional ten coefficients to estimate. Increasing the number further adds many more coefficients and hence reduces the degrees of freedom.

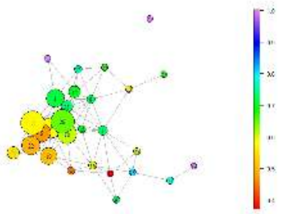
[6]But not totally eliminated.

Table 14: OLS Regression for Hanretty (2017) Leave Percentage: Reduced Data Set

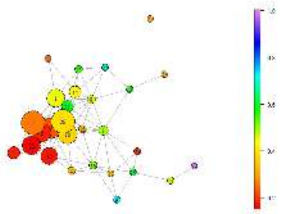
Question	Characteristic	Model 1	Model 2	Question	Characteristic	Model 1	Model 2
Constant		171.03***	117.96***	NSSEC Status	Lower Man.	0.276	-0.111
		(28.51)	(4.46)			(0.141)	(0.122)
Tenure	Outright	-0.097		Qualifications	None	-0.078	
		(0.068)				(0.124)	
	Mortgage	-0.131			Level 4	-1.039***	-0.692***
		(0.082)				(0.104)	(0.057)
Hhold Comp.	Alone	-0.339**		Health	Very Good	-1.575***	-1.635***
		(0.116)				(0.185)	(0.063)
	Married	0.352**	0.365***		Good	0.064	
		(0.061)	(0.057)			(0.263)	
No of Cars	None	-0.537*		Deprivation	Level 0	0.830***	0.448***
		(0.243)				(0.163)	(0.063)
	One	-0.568*	-0.076		Level 1	0.368	
		(0.221)	(0.069)			(0.281)	
	Two	-1.038**					
		(0.060)					
$R-squared$		0.847	0.825				

Notes: Regression of estimated Leave vote percentages from Hanretty (2017) on constituency characteristics from data used in Thorsen et al. (2017). Set of characteristics limited to only those with an average proportion per constituency of 20% or higher. Model 2 limits to the most common characteristic from within each group. Tenure is dropped completely because of correlation with Houshold Composition. NSSEC is the National Statistics Socio-Economic Classification. Data from Thorsen et al. (2017). $n = 631$. Significance given by * - 5%, ** - 1% and *** - 0.1%

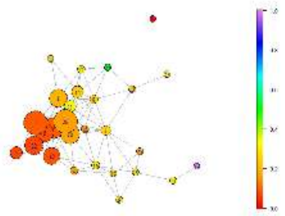
Figure 16: Magnitude of OLS Residuals: Reduced Set ($\epsilon = 18$)



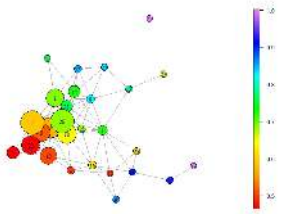
(a) Model 1 $|\hat{e}_i| > 2$



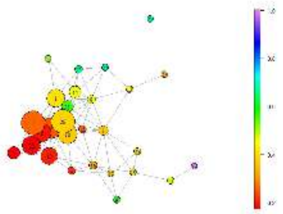
(b) Model 1 $|\hat{e}_i| > 4$



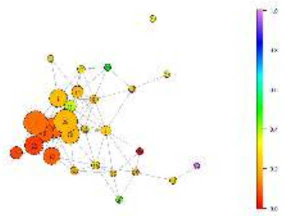
(c) Model 1 $|\hat{e}_i| > 6$



(d) Model 2 $|\hat{e}_i| > 2$



(e) Model 2 $|\hat{e}_i| > 4$



(f) Model 2 $|\hat{e}_i| > 6$

Notes: Model 1 uses the full set of characteristics reduced to only those with an average proportion per constituency of 20% or higher. Model 2 limits to the most common characteristic from within each group. Tenure is dropped completely from Model 2 because of correlation with Household Composition. Data from Thorsen et al. (2017).

not actually doing a good job of representing the driving forces behind the Brexit vote. A strong message comes back from the data to say that understanding of the diversity in the Remain vote has not been strong enough. Here TDA Ball Mapper makes its contribution.

6.4 Regional Effects

Much has been made of the regional differentials within the 2016 Brexit voting patterns; the London versus "not London" divide being a critical part of the immobility discussion seen in Lee et al. (2018), for example. Harris and Charlton (2016) highlighted how it was regions like the East Midlands and East of England, that had delivered more surprising votes to Leave, that were where the referendum was decided. By taking the information about region from within the data this section overlays the concentration of each region onto the TDA Ball Mapper graph constructed for the reduced dataset. In so doing it can be seen that within the central mass of Brexit voting constituencies the proportion from the different regions varies quite considerably. Phrased alternatively the constituencies within each region that voted for Brexit may be found within different parts of the characteristic space.

Figure 17 has twelve panels to demonstrate the heterogeneity within the characteristics of the constituencies around the UK. As eleven regions are included in the Census 2011 classification the first plot is a repeat of Figure 15. Scotland and London are the two regions most associated with Remain and their plots, in panels (b) and (j) respectively, reflect this fact strongly. Only the East of England can be seen to produce anything similar in terms of low intensity within the strongest Leave balls. However, the similarity between London and Scotland is limited when viewing which parts of the Remain characteristic space they occupy. Scottish constituencies make up much of the area to the towards the top of the plot, whilst London constituencies are found almost exclusively in the lower half. There are overlaps, such as some more deprived London areas appearing in balls 8 and 13, and some Scottish constituencies appearing in ball 17. The differences between these halves was shown in Table 13, the upper group having higher incidence of low qualifications, lower level of self-reported health and lower home ownership. What was driving the Scottish remain vote was very different to what was driving that in London.

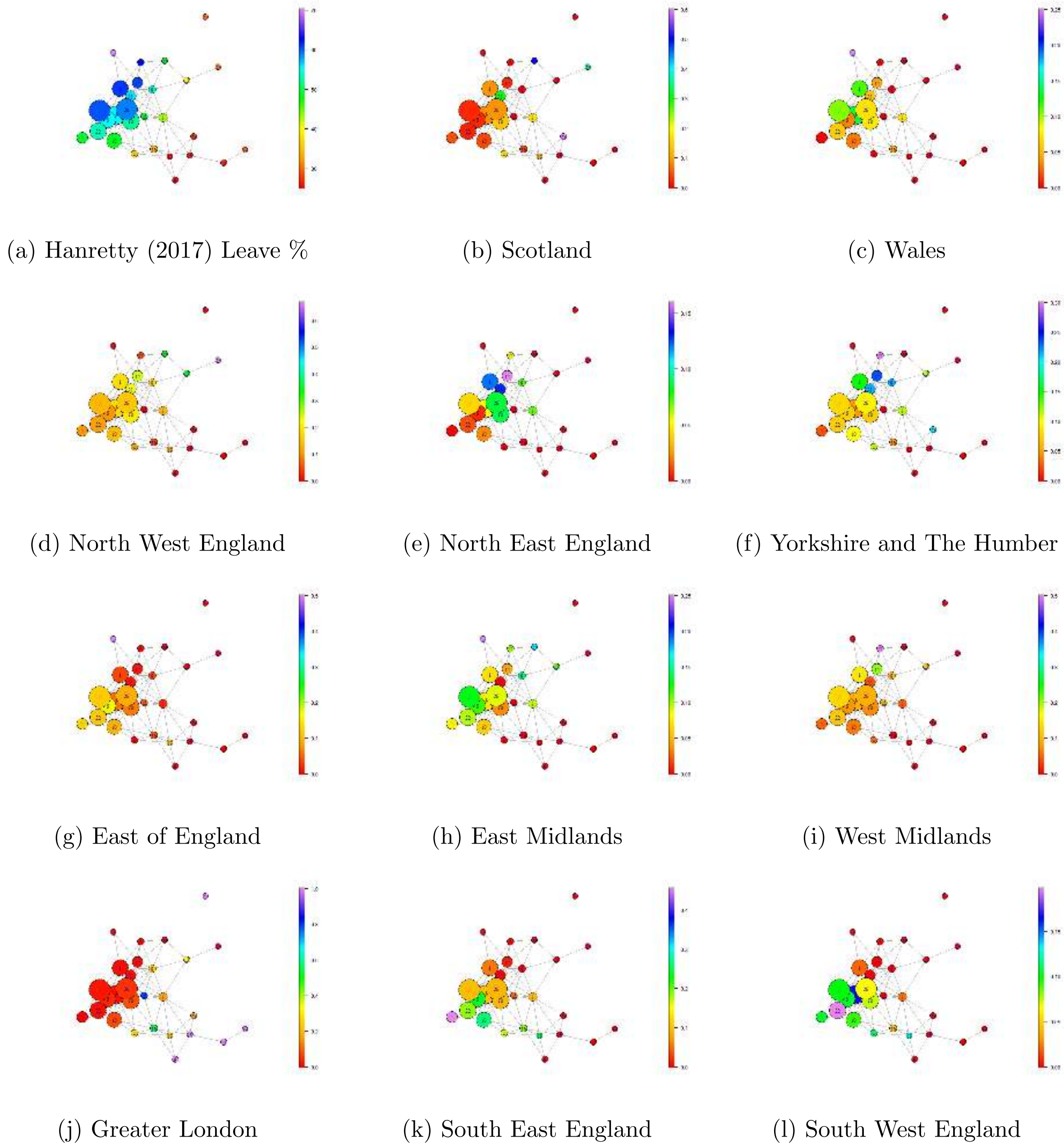
Harris and Charlton (2016) highlights the East Midlands and the East of England as areas where the pro-Brexit vote share came as a surprise to pollsters. The East of England is shown in panel (g) of Figure 17 with the colouration indicating a spread of characteristics through the marginal and leave balls to the left of the plot. However the largest proportion from the East of England is Ball 20, the ball that had the highest average Leave vote. This was an area discussed in Section 6.2, and shown in Table 13, to have a high ownership percentage but very low levels of qualifications and well below average levels of health. In all the discussion of the North these constituencies stand out as ones that were missed and, given the outcome in the East of England, can be seen as pivotal to the overall vote. Constituencies in the East Midlands and Wales also share these characteristics, again not linked strongly with the initial narrative on the post-industrial North of England.

For Wales, panel (c), the North East of England, panel (e), Yorkshire and the Humber, panel (f) and the West Midlands, panel (i), there is a reasonable spread of constituencies across the main Brexit balls and into the marginals to the upper right of the plot. Yorkshire and the Humber in particular makes up a large portion of the very marginal Ball 8. All four regions can be found within Ball 3, which is another mildly Remain ball on the intersection between the characteristics that saw London vote Remain and those shown to prevail in Remain favouring Scotland. These regions show much greater characteristic spread and hence, within them, hold interesting contributions to the overall result. Though not as pronounced as the East of England presence in Ball 20 the message is still one that there exist combinations of demographic characteristics that were fertile ground for the growth of Leave and yet differed from those of the post-industrial North.

For concordance with the narrative on Labour heartlands the North West of England provides a strong barometer. Containing the heavy Remain favouring Manchester and Liverpool the region also has some of the most diverse conditions seen amongst the plots in Figure 17. Comprising a big part of all of the largest balls the North West does features lower in each than at least one other region; for example the proportion of East Midlands constituencies in Ball 7 is higher than that of the North West.

Harris and Charlton (2016) also picks out the South for discussion, this is a region typically dismissed as being Remain because of all of the strongholds of anti-Brexit sentiment therein. However, as the mapping exercise shows there were pockets where the Leave vote did significantly better than expected. For Harris

Figure 17: Reduced Set Coloured by Region



(a) Hanretty (2017) Leave %

(b) Scotland

(c) Wales

(d) North West England

(e) North East England

(f) Yorkshire and The Humber

(g) East of England

(h) East Midlands

(i) West Midlands

(j) Greater London

(k) South East England

(l) South West England

Notes: TDA Ball Mapper diagram constructed using *BallMapper* (Dlotko, 2019). Colouration in panel (a) is by Hanretty (2017) estimated leave percentages with the 50% cut off being towards the upper end of the green shading. Panels (b) to (l) are coloured according to the proportion of observations within any given ball that stem from the specified region. Data from Thorsen et al. (2017).

and Charlton (2016) this was an important oversight, TDA Ball Mapper shows here why that might have happened. Both the South East, panel (k) and the South West, panel (l) make up large proportions of balls 5, 6 and 22 which sit on the margins of Leave in the voting. In the case of the South West almost 20% of consituencies in Ball 22 are from that region.

Analysing the data in this way provides useful insight on the diversity across the UK. Focus on what lies behind voting patterns at the regional level speaks to the purpose of TDA Ball Mapper to unlock the greatest potential from within the data itself. For those seeking to spread their message, be it pro-Leave or pro-Remain this is an important next step in the identification of the issues upon which to play when designing communications.

7 Conclusions

In voting to leave the European Union in 2016 the voters of the United Kingdom sent a major shockwave through the integration agenda, a wave which is still to calm. For many the decision of voters to adopt a position considered to be economically detrimental was challenging, the narrative needing to be phrased in concepts of being "left behind" or in rebellion against the status quo that had made these voters "left behind" in the first place. The empirical analysis that drove these conclusions was simplistic and negated many non-linearities within the data. Using a constituency level dataset this paper has demonstrated the commonalities within the Leave voting areas and, conversely, the diversity across Remain supporting constituencies. Amongst the non-linearities found were potential failures of the "left behind" hypothesis, inconsistencies in the role of social classification and in voting behaviours. To obtain these insights topological data analysis was instrumental, permitting the consideration of all interactions and visualisng the data in ways not yet explored. It has been shown too that there are areas of the data where model fit is better than others, for many Remain areas the fit was particularly poor. TDA Ball Mapper shows how important it is to understand what is going on in these Remain areas; current work is missing an important element of the picture. Heterogeneity across Regions is also usefully represented through the TDA Ball Mapper graph; another useful demonstration of the power of the technique. What has been presented is an exposition of the contribution TDA can make to understanding one of the major political shocks of recent history.

Many critiques of data driven approaches abide, but when trying to understand the way that the final result was arrived at it is useful to dig as deep as possible to avoid the generalisations this paper has shown do not always hold. Likewise the decision to use constituencies is open to discussion given that votes were not reported at that level. As the UK stands on the verge of a potential General Election having information at the constituency level will be valuable to thinking about the likely outcome of that vote. Variable choice is also of importance. Those selected here are ruled by the existing literature and the available data within the readily accessible set of Thorsen et al. (2017). However, the strength of the TDA Ball Mapper algorithm comes from the ability to deal in multiple dimensions. To that end the presentation here can be readily extended and an analysis of any ordinal constituency characteristic incorporated. Notwithstanding these critiques valuable depth has been added to the discussion of the Brexit vote.

From a policy perspective the results inform that the diversity of the Remain vote was always going to be a challenge for mobilisation relative to the concentration of Leave. In determining promotion strategies the Leave message was able to resonate more strongly than the often mis-focused around just part of the overall support for Remain. More broadly this paper demonstrates that Topological Data Analysis and the TDA Ball Mapper algorithm have much to offer in uncovering patterns hidden within the data, or the attempts to generalise therefrom. Next logical steps would see the approach applied to individual level data where there is more heterogeneity and a further interest in the interactions of multiple characteristics. Functionality within the TDA Ball Mapper post-estimation is also developing opening possibilities for further contribution. Harnessing the full power of TDA Ball Mapper for political analysis is an exciting project to come. The depth that TDA Ball Mapper offers, and the value that can bring to understanding results traditional methods do not explain well, stands as a valuable addition in a world still struggling to fully rationalise the political upheavals of recent times.

References

Alabrese, E., Becker, S. O., Fetzer, T., and Novy, D. (2019). Who voted for brexit? individual and regional data combined. *European Journal of Political Economy*, 56:132–150.

Antonucci, L., Horvath, L., Kutiyski, Y., and Krouwel, A. (2017). The malaise of the squeezed middle: Challenging the narrative of the 'left behind'brexiter. *Competition & Change*, 21(3):211–229.

Arnorsson, A. and Zoega, G. (2018). On the causes of brexit. *European Journal of Political Economy*, 55:301–323.

Bailey, D. J. (2018). Misperceiving matters, again: Stagnating neoliberalism, Brexit and the pathological responses of Britain's political elite. *British Politics*, 13(1):48–64.

Becker, S. O., Fetzer, T., and Novy, D. (2017). Who voted for brexit? a comprehensive district-level analysis. *Economic Policy*, 32(92):601–650.

Bell, B. and Machin, S. (2016). Brexit and wage inequality. *VoxEU, August*.

Billing, C., McCann, P., and Ortega-Argilés, R. (2019). Interregional inequalities and uk sub-national governance responses to brexit. *Regional Studies*, 53(5):741–760.

Bromley-Davenport, H., MacLeavy, J., and Manley, D. (2018). Brexit in sunderland: The production of difference and division in the uk referendum on european union membership. *Environment and Planning C: Politics and Space*, page 0263774X18804225.

Carl, N., Dennison, J., and Evans, G. (2019). European but not european enough: An explanation for brexit. *European Union Politics*, 20(2):282–304.

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

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Cipullo, D. and Reslow, A. (2019). Biased forecasts to affect voting decisions? the brexit case. Technical report, Working Paper.

Clarke, S. and Whittaker, M. (2016). *The importance of place: Explaining the characteristics underpinning the Brexit vote across different parts of the UK*. Resolution Foundation.

Crescenzi, R., Di Cataldo, M., and Faggian, A. (2018). Internationalized at work and localistic at home: The 'split' Europeanization behind Brexit. *Papers in Regional Science*, 97(1):117–132.

Darvas, Z. (2016). Brexit should be a wake up call in the fight against inequality. *LSE European Politics and Policy (EUROPP) Blog*.

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

Dlotko, P. (2019). *BallMapper: Create a Ball Mapper graph of the input data*. R package version 0.1.0.

Fox, S., Hampton, J. M., Muddiman, E., and Taylor, C. (2019). Intergenerational transmission and support for eu membership in the united kingdom: The case of brexit. *European Sociological Review*, 35(3):380–393.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2018). Social media, sentiment and public opinions: Evidence from# brexit and# uselection. Technical report, National Bureau of Economic Research.

Hanretty, C. (2017). Areal interpolation and the uk's referendum on eu membership. *Journal of Elections, Public Opinion and Parties*, 27(4):466–483.

Harris, R. and Charlton, M. (2016). Voting out of the european union: Exploring the geography of leave. *Environment and Planning A: Economy and Space*, 48(11):2116–2128.

Haussler, D. and Welzl, E. (1987). ϵ-nets and simplex range queries. 2:127–151.

Hobolt, S. B. (2016). The brexit vote: a divided nation, a divided continent. *Journal of European Public Policy*, 23(9):1259–1277.

Inglehart, R. F. and Norris, P. (2016). Trump, brexit, and the rise of populism: Economic have-nots and cultural backlash.

Jackson, D., Thorsen, E., and Wring, D. (2016). Eu referendum analysis 2016: Media, voters and the campaign.

Lee, N., Morris, K., and Kemeny, T. (2018). Immobility and the brexit vote. *Cambridge Journal of Regions, Economy and Society*, 11(1):143–163.

Liberini, F., Oswald, A. J., Proto, E., and Redoano, M. (2019). Was brexit triggered by the old and unhappy? or by financial feelings? *Journal of Economic Behavior & Organization*, 161:287–302.

Lopez, J. C. A. D., Collignon-Delmar, S., Benoit, K., and Matsuo, A. (2017). Predicting the brexit vote by tracking and classifying public opinion using twitter data. *Statistics, Politics and Policy*, 8(1):85–104.

Los, B., McCann, P., Springford, J., and Thissen, M. (2017). The mismatch between local voting and the local economic consequences of brexit. *Regional Studies*, 51(5):786–799.

Manley, D., Jones, K., and Johnston, R. (2017). The geography of brexit–what geography? modelling and predicting the outcome across 380 local authorities. *Local Economy*, 32(3):183–203.

Matti, J. and Zhou, Y. (2017). The political economy of brexit: Explaining the vote. *Applied Economics Letters*, 24(16):1131–1134.

Sampson, T. (2017). Brexit: the economics of international disintegration. *Journal of Economic perspectives*, 31(4):163–84.

Thorsen, E., Jackson, D., and Lilleker, D. (2017). Uk election analysis 2017: Media, voters and the campaign.

Zhang, A. (2018). New findings on key factors influencing the uk's referendum on leaving the eu. *World Development*, 102:304–314.