
TRANSNATIONAL SOCIAL CONNECTEDNESS AND THE 2016 REFERENDUM ON THE UK'S MEMBERSHIP OF THE EUROPEAN UNION

SUPERVISED INDIVIDUAL STUDY

Owen Winter
School of Geographical Sciences
Bristol University
Bristol BS8 1RL
owen.winter.2021@bristol.ac.uk

March 17, 2022

Keywords EU Referendum · Euroscepticism · Facebook · Migration · Social Connectedness · Social Media · Voting

1 Introduction

There is a broad literature on the social transmission of social and political values (Huckfeldt and Sprague 1987; Lee et al. 2017; Nurmuhhammad et al. 2016) and the effect of social contact on perceptions of out-groups (Enos 2017). In the digital age, the social connections along which political views might be transmitted can be quantified more easily. This study investigates the effect of transnational social connectedness on voting patterns in the United Kingdom's 2016 Referendum on membership of the European Union, using Facebook's Social Connectedness Index (SCI) (T. K. Bailey R. Cao and Wong. 2018).

The research question of this study is **Did transnational Facebook connections with individuals in other EU member states reduce the likelihood of voting to leave the EU?**

2 Hypotheses

Testable hypotheses are set up to elucidate the research question. These hypotheses are based on theorised relationships between transnational social connections, migration and EU referendum vote choice (Figure 1). While the main purpose of this study is to consider the effects of transnational social connections on EU referendum voting, individual-level data on these connections is not available. This means using aggregated data, which is closely related to other variables which may affect voting patterns.

As illustrated in Figure 1, Facebook connections between UK and EU residents are shaped by migration, because it is expected that migrants have high levels of connectedness to origin countries (**H1**).

H1: Areas with higher levels of EU migration tend to have higher levels of social connectedness with EU residents

The effects of migration on voting operate at both the individual level and as a contextual effect. It is expected that EU migrants are less likely to vote to leave the EU because of the risks to continued freedom of movement. In line with Goodwin and Milazzo (2017), it is hypothesised that the presence of EU migrants in an area had a negative effect on the likelihood of voting to leave the EU, both through the individual-level effect and because exposure to migrants reduces anti-migrant sentiment, leading to more positive assessments of membership of the EU (**H2**). Meanwhile, increases in the *rate* of migration locally had a positive effect on likelihood to vote to leave the EU, because of a perception of undesirable change (**H3**).

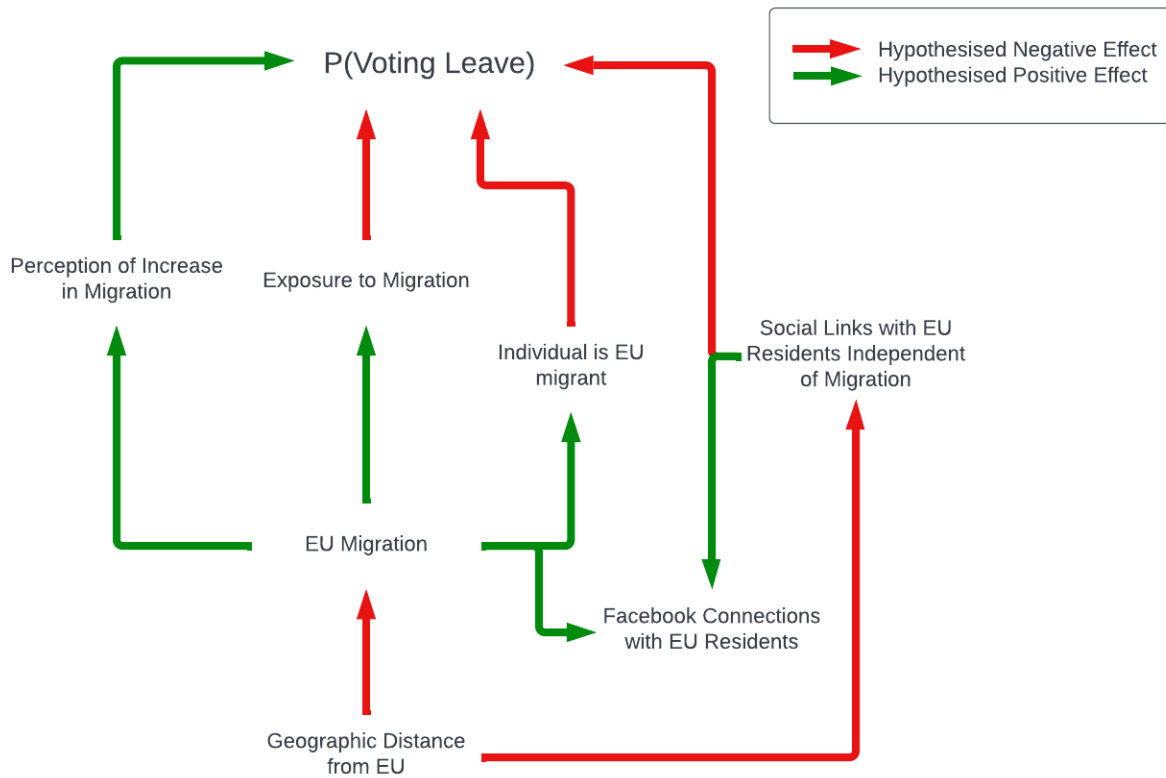


Figure 1: Hypothesised Effects of Migration and Social Connectedness on EU Referendum Voting

H2: The presence of EU migrants in an area has a negative effect on likelihood of voting to leave the EU

H3: Recent increases in the volume of EU migration in an area has a positive effect on likelihood of voting to leave the EU

As a result of these migration-related effects, it is expected that areas which voted to remain in the EU will have higher levels of social connectedness with EU residents (**H4**). This is not necessarily a causal relationship, but reflects the fact that areas with higher levels of EU migration tended to vote to remain in the EU.

H4: Areas which voted to remain in the EU tend to have higher levels of social connectedness with EU residents

Finally, the aim of this study is to isolate the effect of social connectedness with EU residents independent from the individual and contextual effects of migration discussed above. It is hypothesised that social connections will have an independent negative effect on the likelihood of voting to leave the EU, because of the social transmission of positivity towards other EU nations and reduced negativity towards EU migration (**H5**). This is consistent with the findings of M. Bailey et al. (2020), that, across the EU, euroscepticism decreases with the share of a region's Facebook connections that are to regions in a different European country.

H5: Social connectedness with EU residents has a negative effect on the likelihood of voting to leave the EU

3 Data

3.1 Facebook Social Connectedness Index (SCI)

The SCI is published by Facebook (T. K. Bailey R. Cao and Wong. 2018) to study connections across geographic regions. It is produced by assigning Facebook users to geographic areas using profile information, before counting and scaling the number of connections between pairs of areas. Areas with few friendships are dropped and random noise is added to ensure anonymity. The SCI is the number of possible friendships between two regions which are fulfilled, per billion.

$$SocialConnectedness_{i,j} = \frac{FB_Connections_{i,j}}{FB_Users_i * FB_Users_j}$$

The lowest level in the UK which SCI is measured at is GADM regions, which are concurrent with NUTS regions (Global Administrative Areas 2022). Table 1 shows the highest SCI scores including UK regions. As one might expect, the highest scores are within-region, with the highest proportion of friendships fulfilled within UKM66 (Shetland Islands).

Table 1: Highest SCI scores at GADM-level, all pairs including UK regions

Region 1	Region 2	SCI
UKM66	UKM66	61610896
UKM65	UKM65	52441781
UKM64	UKM64	39082660
UKL11	UKL11	15893326
UKN16	UKN16	12226389
UKN10	UKN10	11714844

To study the effect of transnational friendships, SCI scores must be aggregated at higher-levels than GADM1 for regions outside the UK. Because Facebook does not share this information, the closest approximation is a population weighted average. This approach assumes that *Facebook_Users_i* in the equation above is approximately proportional to the population of region *i*. This is necessary to produce coherent units for analysis. In the EU, GADM1 regions are concurrent with NUTS regions, so NUTS population data can be used (eurostat 2021).

Table 2: Highest SCI scores between UK regions and other countries

UK Region	Country	Mean SCI	Population Weighted Mean SCI
UKN11	TLS	106071	106071
UKN10	IE	67299	50373
UKN16	IE	56769	44639
UKJ31	VCT	37680	37680
UKE42	TCA	34643	34643
UKI32	MCO	34200	34200

Table 2 shows UK NUTS regions and their scaled SCI with countries excluding the UK. The most connected regions are UKN11 (Mid Ulster) and TLS (East Timor). An article by The Belfast Telegraph (Peake 2014) suggests that there is a genuine connection, however, the presence of Timor-Leste also shows the limitations of the SCI. The SCI measures the proportion of all possible connections fulfilled, rather than the proportion of residents. In the East Timorese case, it may be that a small minority of Mid Ulster residents have many East Timorese connections, rather than a large proportion of those in Mid Ulster having a connection to East Timor.

Principal components analysis (PCA) can be used for descriptive analysis, showing some of the associations between UK regions and country-level SCI scores. A log transformation is used to scale the data, to reflect

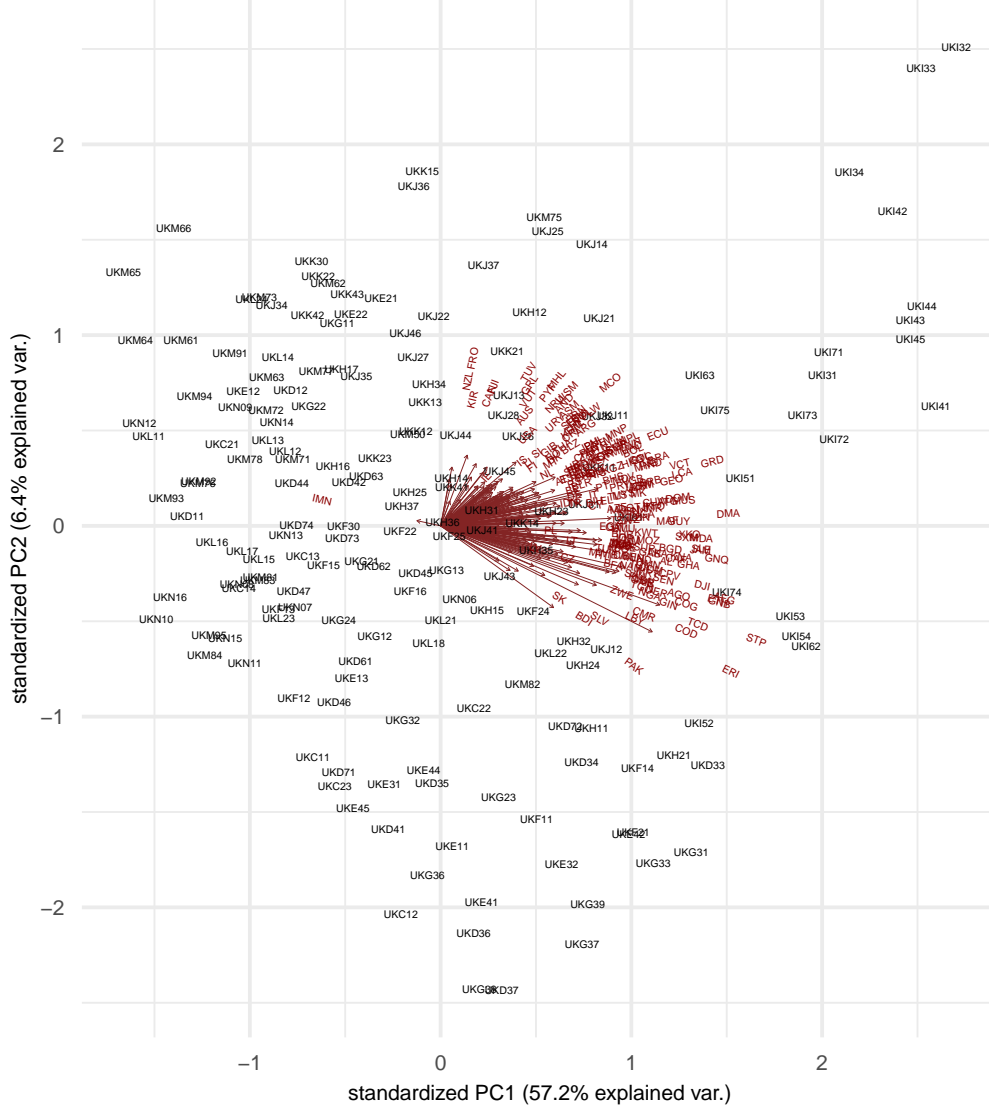


Figure 2: Biplot of country-level SCI scores by UK NUTs area

the exponential nature of social connections (Bettencourt et al. 2007). The results (Figure 2) show how there is significant correlation between countries along one scale (PC1). Areas with more transnational connections tend to have higher SCI scores across the board (with the sole exception of the Isle of Man). PC2 explains much less variation, only 6.4%.

Finally, EU-27 country-level SCI scores are merged using a population weighted average to give each NUTS-3 region of the UK an SCI score for the EU-27. This statistic is plotted in Figure 3 along with internal SCI. That is, the proportion of social connections within the region which are fulfilled. The two indexes are compared in Figure 4.

3.2 British Election Study

The second main source of data is the British Election Study (BES) (Fieldhouse and Eijk 2021). The BES is a large-scale survey of public opinion, with data available at the level of individual respondents. There are wide ranging questions including demographic variables, policy preferences and past votes. BES is conducted using online surveys and a random sample weighted to be representative of the population of Great Britain. Wave 9, which includes the most relevant data for this study, was conducted between 24th June 2016 and 4th July 2016 with 30,036 respondents.

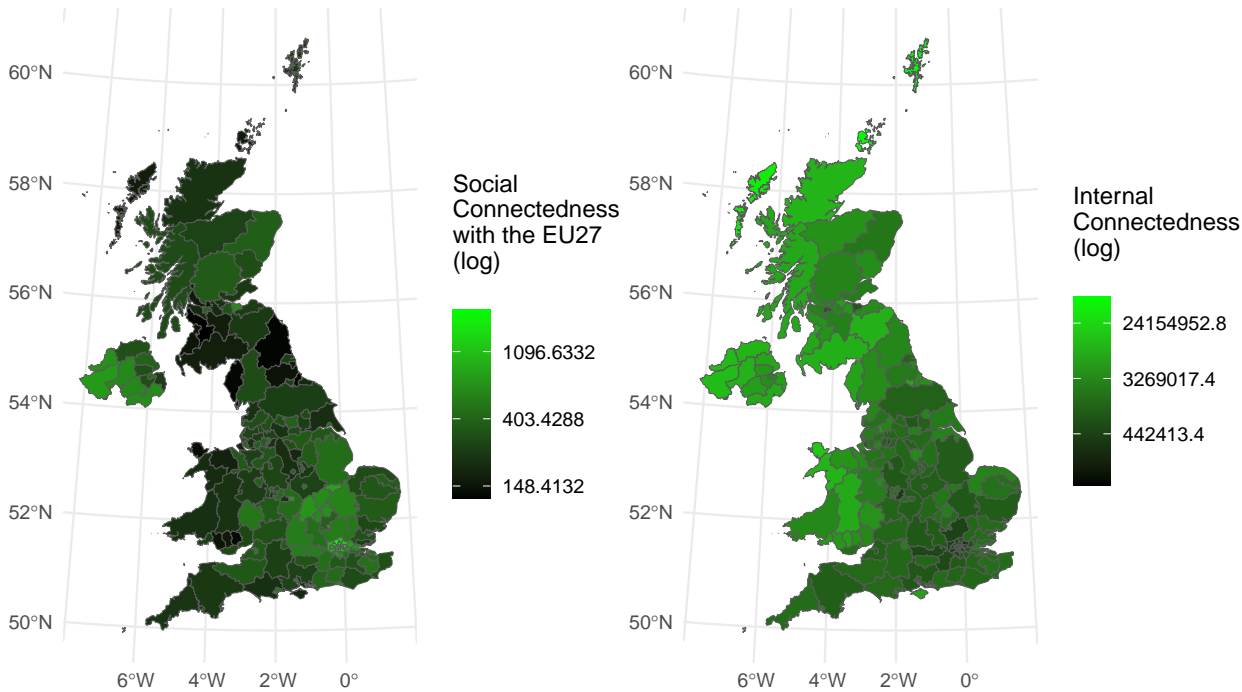


Figure 3: NUTS3 Regions by SCI with the EU-27 and internally

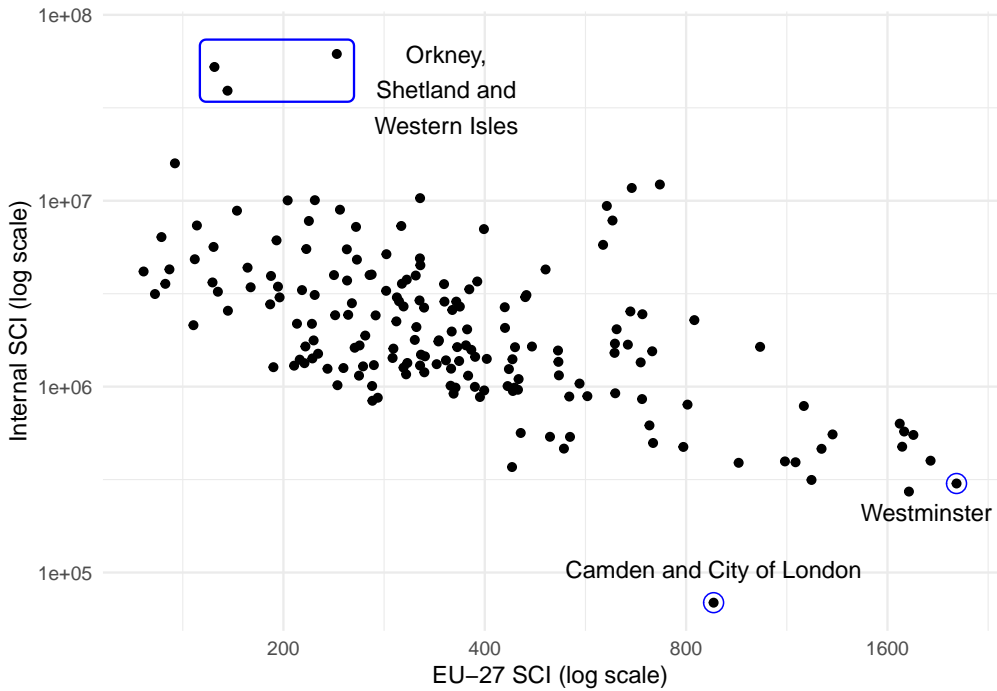


Figure 4: Scatter Plot of Internal and EU-27 SCI Scores

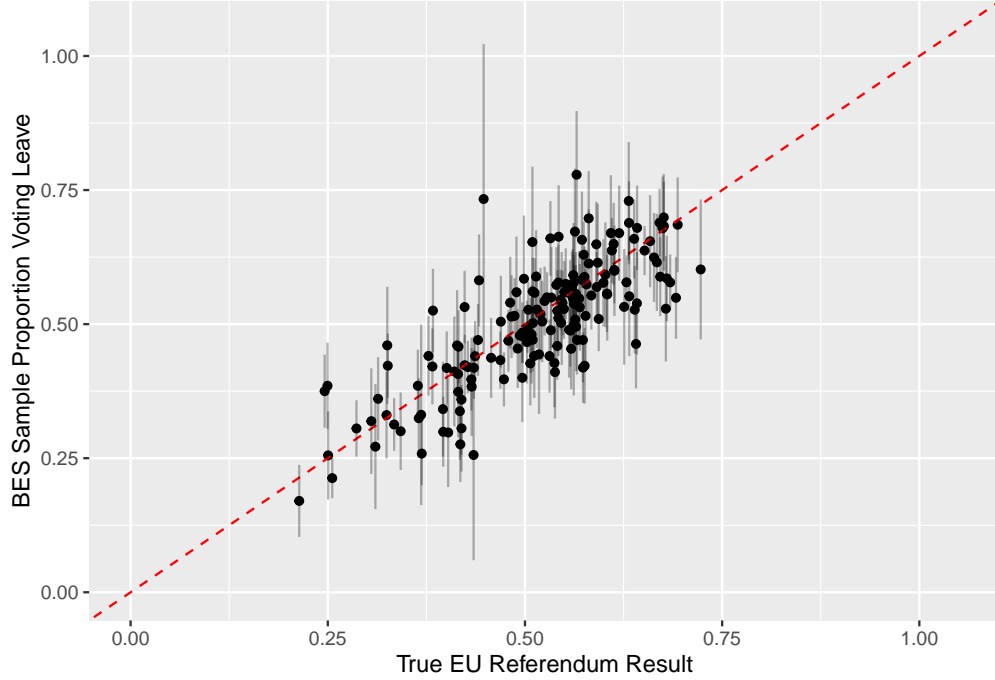


Figure 5: Reported EU Referendum vote (BES) compared with true results at NUTs-level

To check the validity of BES, NUTS-level samples are compared to true EU Referendum results Authority (2018). Figure 5 shows how 76% of NUTS regions have sample proportions within the 95% margin of error of true EU Referendum results for the sample size. This suggests some sampling bias on top of statistical variance.

3.3 NUTS-3 Level Data

Migration data is aggregated at the NUTS-level, using datasets provided at local authority level by the Office for National Statistics Stickney (2021b). This includes the proportion of residents born in the EU, the rate of EU migration as a proportion of current population in 2016, and change in rate of EU migration between 2011 and 2016. Finally, distances between NUTS regions and the EU-27 are calculated. These distances are measured from the centroid of each NUTS regions to the closest border of the EU-27 (Figure 6).

4 Methods

The primary method used is multilevel logistic regression (Gelman, Hill, and Hill 2007). The dependent variable is self-reported vote to leave the EU, of all those who voted in the 2016 Referendum, weighted using BES weights. A multilevel model better represents the hierarchical nature of survey data, with respondents nested within geographic or demographic groups. In each case, respondents are drawn from broader populations and the effect of these group-memberships is drawn from a distribution of possible effects. In practice, this is useful because it allows for partial pooling of the intercept and variance of each group, leading to more accurate estimates including more robust standard errors for fixed effects. In this study, we are most interested in the effect of migration and social connections, so most other variables are included as random effects.

5 Results

The first results to consider are of two correlation tests, between SCI scores and the proportion of people born in the EU-27 (**H1**) and between SCI scores and the proportion of people who voted to leave the EU (**H4**) at NUTS level. In both cases, there is a significant correlation, leading us to accept the hypotheses. The

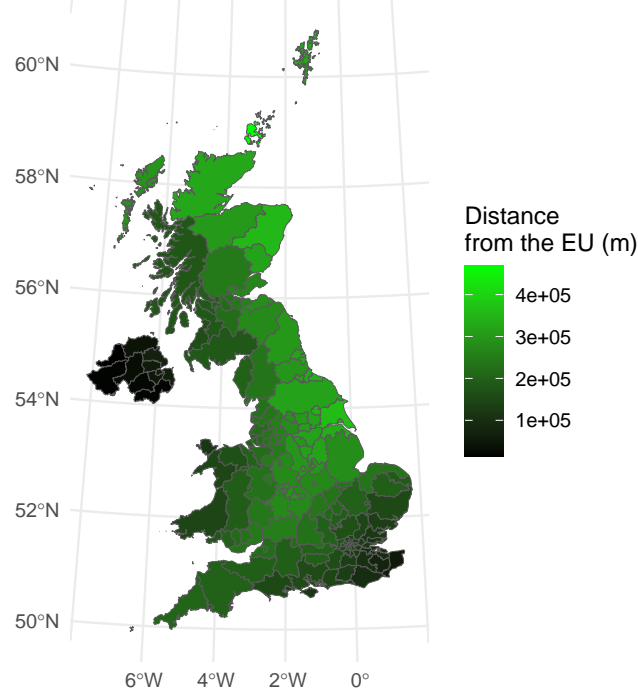


Figure 6: Distance to EU-27

correlation between SCI and EU-born population is particularly high (0.899) while the correlation between SCI and leave-voting is less pronounced but significant (-0.463).

Table 3: Correlation Tests

Pearson's Product-Moment Correlation EU-27 Social Connectedness Index		
	Percentage Born in EU	Leave Vote Percentage
Estimate	0.899***	-0.463***
Degrees of Freedom	165	166
T-Statistic	26.421	-6.739
95% Confidence Interval	[0.866, 0.925]	[-0.575, -0.336]
<i>Note:</i> . p<0.1; *p<0.05; **p<0.01; ***p<0.001		

To consider **H2**, **H3** and **H5**, the results of multilevel logistic regression is presented in Table 4.

In line with **H2** and **H3**, the rate of EU migration has a negative effect on the likelihood of voting to leave the EU, while a change in rate has a positive effect. However, these effects are not significant in models 5 or 6, so there is insufficient evidence to accept the hypotheses. In addition, rather than the expected negative relationship between SCI and voting to leave (**H5**), there is a small positive effect, which is not significant. This is the case across all six models. Again, there is insufficient evidence to accept the hypothesis. Perhaps the most interesting finding is that in both models 3 and 6, the internal SCI of a region has a significant positive effect.

Table 4: Logistic Regression Models

	<i>Dependent variable:</i>					
	Leave Vote					
	(1)	(2)	(3)	(4)	(5)	(6)
Age (Scaled 0-1)	1.789*** (0.059)	1.156*** (0.081)	1.146*** (0.081)	1.150*** (0.081)	1.150*** (0.081)	1.145*** (0.074)
Gender (Female = 1)	-0.087** (0.029)	-0.149*** (0.036)	-0.150*** (0.036)	-0.149*** (0.036)	-0.149*** (0.036)	-0.149*** (0.033)
EU Migration Inflow 2016 (Per Capita)	-16.229* (6.713)			-4.769 (3.942)	-11.670 (7.100)	-7.773 (6.433)
Inflow Change 2011-2016	61.489*** (10.504)			26.790** (9.525)	20.338 (11.474)	13.968 (16.809)
log(EU-27 SCI)	0.148 (0.100)	0.021 (0.080)			0.161 (0.118)	0.204 (0.111)
log(Internal SCI)	0.055 (0.052)		0.165** (0.053)			0.164** (0.054)
Distance from EU (Scaled 0-1)	0.210 (0.207)					0.321 (0.299)
Constant	-2.247* (1.085)	-1.008 (0.580)	-3.230*** (0.833)	-0.835* (0.341)	-1.713* (0.730)	-4.478*** (1.130)
Random Effects σ^2						
NUTS 3	0.106	0.092	0.077	0.087	0.087	0.074
Gross Household Income		0.010	0.010	0.010	0.010	0.010
Gross Personal Income		0.583	0.580	0.582	0.582	0.581
Newspaper Read		0.009	0.009	0.009	0.009	0.009
Ethnicity		0.223	0.224	0.224	0.225	0.226
County of Birth		0.019	0.019	0.019	0.019	0.019
UK Region		0.050	0.083	0.051	0.045	0.066
Social Grade		0.079	0.079	0.079	0.079	0.079
Education Level		0.188	0.189	0.187	0.187	0.187
Observations	28,195	22,803	22,803	22,803	22,803	22,803
Log Likelihood	-14,277	-10,362	-10,357	-10,360	-10,359	-10,355
Akaike Inf. Crit.	28,571	20,750	20,741	20,748	20,748	20,743
Bayesian Inf. Crit.	28,645	20,855	20,845	20,860	20,869	20,880

Note:

. p<0.1; *p<0.05; **p<0.01; ***p<0.001

6 Discussion and Conclusion

This study has not provided conclusive evidence of an effect of transnational social connectedness on voting choices in the 2016 EU Referendum. This may be due to the aggregation of SCI scores, measurement error, confounding variables, or simply that the theorised effect does not occur. One interesting result, which ought to be studied further, is the significant positive relationship between voting to leave the EU and internal SCI. Internal SCI shows connectedness within regions, and tends to be highest in areas which are relatively isolated (especially island regions). This may reflect suspicion of migration and/or transnational institutions in tightly knit communities, or a particular form of social transmission of euroscepticism. Internal SCI is also likely to be closely correlated to areas with fewer links to other areas of the UK, which may be related to narratives of regional inequality and “geography of discontent” (McCann and Ortega-Argilés 2021).

Replication Materials

Winter, O. (2022) Facebook Links Project. GitHub Repository. https://github.com/owenwntr/facebook_links_project

R Packages

- creditmodel (Fan 2022)
- ggbiplot (Vu et al., n.d.)
- ggforce (Pedersen 2021)
- ggplot2 (Wickham 2016)
- haven (Wickham and Miller 2022)
- knitr (Xie 2021)
- labelled (Larmarange et al., n.d.)
- lme4 (Bates et al. 2015)
- matrixStats (Bengtsson 2021)
- readxl (Wickham and Bryan 2022)
- rticles (Allaire et al. 2022)
- sf (Pebesma 2018)
- stargazer (Hlavac, n.d.)
- stringr (Wickham 2019)
- tidyverse (Wickham et al. 2019)

Bibliography

- Allaire, JJ, Yihui Xie, Christophe Dervieux, R Foundation, Hadley Wickham, Journal of Statistical Software, Ramnath Vaidyanathan, et al. 2022. *Rticles: Article Formats for r Markdown*. <https://CRAN.R-project.org/package=rticles>.
- Authority, Greater London. 2018. “EU Referendum Results.” Dataset. <https://data.gov.uk/dataset/be2f2aec-11d8-4bfe-9800-649e5b8ec044/eu-referendum-results>.
- Bailey, Michael, Drew Johnston, Theresa Kuchler, Dominic Russel, Bogdan State, and Johannes Stroebel. 2020. “The Determinants of Social Connectedness in Europe.” Book Section. In *Lecture Notes in Computer Science*, 1–14. Springer International Publishing. https://doi.org/10.1007/978-3-030-60975-7_1.
- Bailey, T. Kuchler, R. Cao, and A. Wong. 2018. “Social Connectedness: Measurements, Determinants, and Effects.” Journal Article. *Journal of Economic Perspectives* 32 (3): 259–80. <https://dataforgood.facebook.com/>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bengtsson, Henrik. 2021. *matrixStats: Functions That Apply to Rows and Columns of Matrices (and to Vectors)*. <https://CRAN.R-project.org/package=matrixStats>.
- Bettencourt, L. M., J. Lobo, D. Helbing, C. Kuhnert, and G. B. West. 2007. “Growth, Innovation, Scaling, and the Pace of Life in Cities.” Journal Article. *Proceedings of the National Academy of Sciences of the United States of America* 104 (17): 7301–6.
- Enos, Ryan D. 2017. “The Space Between Us.” Electronic Book. Cambridge University Press. <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5212975>.

- eurostat. 2021. “Population Change - Demographic Balance and Crude Rates at Regional Level (NUTS 3).” Dataset. eurostat.
- Fan, Dongping. 2022. *Creditmodel: Toolkit for Credit Modeling, Analysis and Visualization*. <https://CRAN.R-project.org/package=creditmodel>.
- Fieldhouse, J. Green, E., and C. van der Eijk. 2021. “British Election Study Internet Panel Waves 1-21.” Dataset.
- Gelman, Andrew, Jennifer Hill, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Book. Analytical Methods for Social Research. Cambridge ; Cambridge University Press. http://catdir.loc.gov/catdir/enhancements/fy0668/2006040566-t.html%0Ahttp://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&doc_number=015428351&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA%0Ahttp://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=015428351&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA%0Ahttp://www.gbv.de/dms/ilmenau/toc/527420107.PDF%0Ahttp://catdir.loc.gov/catdir/enhancements/fy0733/2006040566-b.html%0Ahttp://catdir.loc.gov/catdir/enhancements/fy0668/2006040566-d.html.
- Global Administrative Areas, Database of. 2022. “GADM Maps and Data.” Web Page. <https://gadm.org/>.
- Goodwin, Matthew, and Caitlin Milazzo. 2017. “Taking Back Control? Investigating the Role of Immigration in the 2016 Vote for Brexit.” Journal Article. *The British Journal of Politics and International Relations* 19 (3): 450–64. <https://doi.org/10.1177/1369148117710799>.
- Hlavac, Marek. n.d. “Stargazer: Well-Formatted Regression and Summary Statistics Tables. R Package Version 5.2.3.” Computer Program. <https://CRAN.R-project.org/package=stargazer>.
- Huckfeldt, Robert, and John Sprague. 1987. “Networks in Context: The Social Flow of Political Information.” *American Political Science Review* 81 (4): 1197–1216. <https://doi.org/10.2307/1962585>.
- Larmarange, Joseph, Daniel Ludecke, Hadley Wickham, Michal Bojanowski, and Francois Briatte. n.d. “Labelled.” Computer Program. <https://larmarange.github.io/labelled/>.
- Lee, Francis L. F., Paul S. N. Lee, Clement Yk So, Louis Leung, and Michael Cheming Chan. 2017. “Conditional Impact of Facebook as an Information Source on Political Opinions: The Case of Political Reform in Hong Kong.” Journal Article. *Asian Journal of Political Science* 25 (3): 365–82. <https://doi.org/10.1080/02185377.2017.1352523>.
- Mccann, Philip, and Raquel Ortega-Argilés. 2021. “The UK ‘Geography of Discontent’: Narratives, Brexit and Inter-Regional ‘Levelling up’.” Journal Article. *Cambridge Journal of Regions, Economy and Society* 14 (3): 545–64. <https://doi.org/10.1093/cjres/rsab017>.
- Nurmuhammad, Rizwangul, Heather A. Horst, Evangelia Papoutsaki, and Giles Dodson. 2016. “Uyghur Transnational Identity on Facebook: On the Development of a Young Diaspora.” Journal Article. *Identities* 23 (4): 485–99. <https://doi.org/10.1080/1070289x.2015.1024126>.
- Peake, Gordon. 2014. “Hard-Working, Respectful and Warm... We Could Learn from Our Timorese Guests.” Newspaper Article. <https://www.belfasttelegraph.co.uk/opinion/columnists/hard-working-respectful-and-warm-we-could-learn-from-our-timorese-guests-30377313.html>.
- Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pedersen, Thomas Lin. 2021. *Ggforce: Accelerating 'Ggplot2'*. <https://CRAN.R-project.org/package=ggforce>.
- Stickney, Chris. 2021a. “Local Area Migration Indicators, UK.” Dataset. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/datasets/localareamigrationindicatorsunitedkingdom>.
- . 2021b. “Population of the UK by Country of Birth and Nationality.” Dataset. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/populationoftheunitedkingdombycountryofbirthandnationality>.
- Vu, Vince, Ben Marwick, Jim Hester, and Max Held. n.d. “Ggbiplot.” Computer Program. <https://github.com/vqv/ggbiplot/tree/experimental>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, and Jennifer Bryan. 2022. *Readxl: Read Excel Files*.

Wickham, Hadley, and Evan Miller. 2022. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*.
Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*.