Political Virtual Network: an Analysis of the UK Parliament and the US Congress on Twitter

Claudia Abdallah, Renata Canini, Giovanna Chaves, Lucas Santos

August 10, 2023

Abstract

Twitter has become one of the main channels of communication between elected officials and voters. Legislative representatives, in particular, can use the platform to inform taxpayers about their activity in parliament, promote their campaigns, and interact with fellow politicians. In this project, we aim to understand virtual networks under different political systems and how they contribute to political discourse. Using Twitter data of politicians in the United Kingdom's House of Commons and the United States' House of Representatives, we compare the two network structures and apply community detection algorithms to identify relevant clusters. Our results point to a highly partisan partition of the networks, with the US network's communities concentrated around the two main parties (Democrat and Republican), and with the UK one presenting a hierarchical structure dominated by the two main parties, but more disperse than the US.

1 Introduction

The use of Twitter amongst government and members of parliament (MPs) has cultivated a new online realm of communication, drastically changing the relationship between elected officials across political parties. The nature of social media networks amongst Members of Parliament (MPs) has been extensively researched, particularly with regards to their involvement in communication with the electorate (Graham et al., 2013; Margaretten and Gaber, 2012), information dissemination (Small, 2011), and political polarisation (Buccoliero et al., 2020; Gruzd, 2012). This study primarily focuses on the latter.

There is a growing body of evidence suggesting that social media networks have contributed to increasing levels of polarisation amongst politicians and the general public. This is influenced in particular by the development of "echo chamber" environments within social networks, where individuals are overexposed to information which acts in support of their own opinions. The formation of such isolated networks could consequently result in politicians restricting their online interactions to politicians from the same party, limiting the potential for dialogue between MPs from different political groups. Given the relatively high volume of political discourse conducted on social networks, in particular Twitter, online polarisation has the potential to hamper collaboration within parliament itself. Thus, understanding the nature of online relationships between politicians is key.

This study seeks to identify the structure of political virtual networks under different political systems, using the UK parliament and the US Congress as case studies of interest. Our paper proceeds as follows: first, we provide a review of the existing literature on the use of Twitter and other social networks by political representatives. We then detail the data collection method and descriptive statistics that allow us to generalize our findings of MPs on Twitter to MPs broadly. Furthermore, we compare the network structures of the United States House of Representatives with that of the UK's House of Commons, highlighting where they differ and the implications for cross-party communication. In the Methodology section, we describe the algorithms used to predict missing node attributes and undertake community detection. Finally, we present the results and discuss the differences between the UK and the US when it comes to the interactions of their representatives on Twitter.

2 Literature Review

2.1 The Structure of Political Networks on Twitter

Understanding the structure of online social networks is key to mapping a range of social interactions and information sharing (Yoon and Park, 2014). Since its inception in 2006, Twitter membership has grown exponentially, amassing a total of 350 million active monthly users in 2022 (Statista, 2022). Twitter has subsequently become a representative "networked public sphere", providing a platform for discussions about government, politics, and society to take place amongst elected officials and members of the public (Benkler, 2007). The unique network structures exhibited by political social networks have been widely debated in the literature, an overview of which is presented in Soares et al. (2018). Some studies support the idea that operating in one large public sphere facilitates interactions between groups with different socio-political ideologies, with different communities being connected by key users on their borders (Bastos, 2012; Bruns and Highfield, 2016). Others, however, suggest that these networks are in fact highly fragmented, and that consequently twitter cannot be considered as a large public sphere but rather a collection of smaller, inward-looking clusters. The fragmentation and isolation of groups in this regard threatens the potential for discourse covering all realms of political opinion, and can contribute to the rapid polarisation of views for those at different ends of the political spectrum.

Given that social media platforms often implement algorithms for network filtering (Pariser, 2011), presenting individuals with information tailored their pre-existing preferences and opinions, the development of polarised communities is inherent to the structure of social networks (Soares et al., 2018). This serves to inhibit individuals' access to wider information and perspectives and leads to the development of polarised communities, each of which operate within their own "echo-chamber". Smith et al. (2014) explores this phenomena though the model of the "polarised crowd", a social network structure formed by two polarised groups, to provide a theoretical grounding for the network representation of bipartisan political systems. Recuero et al. (2017) extend the applicability of this model to a multi-party system in Brazil. Thus, the potential for clustering is exhibited across a range of political systems, despite different levels of party concentration.

2.2 Political Polarisation and Democracy

The development of fragmented political networks has significant implications for the evolution of democratic societies. Wojcieszak and Mutz (2009) highlights how deliberative discussions across political communities contributes to individuals being more informed, tolerant, and reflective by provide them with the environment to reevaluate their preconceived opinions. Disagreement is thus paramount for facilitating deliberation between communities, and the functioning of the democratic process as a whole. Consequently, the fragmentation of online political networks hampers the contribution of social media to improving democracy, since individuals located within clustered communities limit their interactions to others with similar views (Mislove et al., 2010). It thus becomes an urgent question as to whether social media promotes public sphere or creates echo chambers (Silva and Proksch, 2017), which is a key question this analysis seeks to address.

2.3 The Use of Twitter by Elected Officials

Twitter provides MPs with the means to directly communicate their views with the general public. In contrast to regulated parliamentary debates, which often restrict the time and gravity permitted to place on certain issues, Twitter provides a platform upon which MPs can express a broader range of opinions (Silva and Proksch, 2022). Whilst Kruikemeier (2014) suggests that politicians use Twitter interactively as a means of interacting with the public, Jungherr (2016) shows that it also plays a key role as a "broadcasting" tool. This provides MPs with a wider grasp of the views held by their peers than they would have had without the presence of social media. It additionally acts as another public platform for MPs to debate amongst each other on political issues, where they can express a broader range of views than they would be able to in a parliamentary setting. The extent to which this contributes to furthering the democratic discourse depends on whether MPs follow their counterparts from opposition parties, or whether they operate in "echo-chambers" of MPs sharing their own political views.

2.4 Synthesis

Identifying the structure of political social networks is key to informing whether social media facilitates or hampers the progression of democratic debate. This issue is especially pertinent with regards to the use of twitter by elected officials in parliament, notably in relation to their interactions with their peers. The presence of a diverse and integrated network structure would suggest that Twitter plays an active role in encouraging MPs to collaborate and exchange views. Meanwhile, a network structure comprised of isolated party-specific clusters would indicate that online political networks exacerbate the divisions between MPs from opposing parties, leading to a concentration of homogeneous perspectives. This study seeks to investigate whether political social networks amongst MPs foster a collaborative or divisive structure. We tackle this question by investigating the structure of Twitter networks between MPs in the UK and Members of Congress in the United States.

3 Data

The United Kingdom has 650 members of the House of Commons, while the United States has 441 members in the House of Representatives (435 voting members and 6 non-voting delegates). As there was no data readily available to undertake the analysis and answer the questions proposed in this work, we constructed our own database from data retrieved using the Twitter API Research Access. During the week of March 27th, we collected information on the members of the 2019 UK Members of Parliament and the members of the 118th United States House of Representatives using the **twarc2** library, which enables command lines via prompt to archive Twitter data in .json files.

The first step was to collect the usernames of current members, which we retrieved through Politics Social for the UK and House of Representatives Press Gallery for the US. Then, through the API, we identified which politicians followed each other. Since this process is computationally expensive, we approached this through the smaller side using their following count, as we expected them to follow a smaller number of people than they had followers. Finally, we filtered the list of followings to keep only the aforementioned politicians and, with the help of the **NetworkX** library, created a dataframe with the structure of a network graph. The data was so collected in 02/04/2023, so the network and results we are going to present are conditioning to the information available on that date.

3.1 Descriptive Statistics

Through the aforementioned process, we were able to retrieve Twitter data for 426 of the 441 representatives of the United States and 581 of the 650 MPs in the United Kingdom. Of the missing politicians, 8 from the US and 60 from the UK were not on Twitter at the time of collection. Data for the remaining 7 American and 9 British representatives was corrupted and therefore could not be accessed.

Table 1 shows the distribution of the politicians we could get data for by political party, while tables 3 and 2 display their distribution by geography. This allows us to compare the sample on Twitter with the general population of MPs and representatives. As we are only missing less than 4% of the US data and around 10.6% of the UK data, we expected the two populations to be relatively similar. Even then, it is important to compare these two to ensure there are not systematic differences between the types of politicians with and without Twitter.

Country	Political Party	# Representatives	% of Twitter	% of House
USA	Republican Party	217	50.94	51.02
	Democratic Party	209	49.06	48.98
UK	Alliance Party of Northern Ireland	1	0.17	0.15
	Conservative	298	51.29	54.46
	Democratic Unionist Party	6	1.03	1.23
	Green Party	1	0.17	0.15
	Independent	10	1.72	2.15
	Labour	192	33.05	30.62
	Liberal Democrat	14	2.41	2.15
	Plaid Cymru	3	0.52	0.46
	Scottish National Party	46	7.92	7.08
	Sinn Féin	7	1.21	1.08
	Social Democratic and Labour Party	2	0.34	0.31
	Speaker	1	0.17	0.15

Table 1: Distribution of Representatives by Political Parties, both on Twitter and in the House of Representatives and House of Commons

Country	Region	# Representatives	% of Twitter	% of House
UK	East Midlands	38	6.54	7.08
	Eastern	53	9.12	8.92
	London	70	12.05	11.23
	North East	26	4.48	4.46
	North West	66	11.36	11.54
	Northern Ireland	16	2.75	2.77
	Scotland	58	9.98	9.08
	South East	72	12.39	12.92
	South West	49	8.43	8.46
	Wales	36	6.20	6.15
	West Midlands	49	8.43	9.08
	Yorkshire & the Humber	48	8.26	8.31

Table 2: Distribution of Representatives by Regions, both on Twitter and in the House of Commons

While we can see that there are some differences in the distribution of political parties, notably that there seem to be many Conservative MPs missing from Twitter, none of these differences are statistically significant. The same can be said for the geographic distribution, which confirms the adequacy of using this data to represent the Houses of the US and UK government.

Country	State	# Representatives on Twitter	% of Twitter	% of House
USA	Alaska	1	0.24	0.23
	Alabama	7	1.64	1.59
	Arkansas	4	0.94	0.91
	Arizona	9	2.11	2.04
	California	51	11.97	11.79
	Colorado	7	1.64	1.81
	Connecticut	4	0.94	1.13
	Delaware	1	0.24	0.23
	Florida	28	6.57	6.35
	Georgia	14	3.29	3.18
	Hawaii	2	0.47	0.45
	Iowa	4	0.94	1.13
	Idaho	2	0.47	0.45
	Illinois	17	3.99	3.86
	Indiana	9	2.11	2.04
	Kansas	4	0.94	1.13
	Kentucky	6	1.41	1.36
	Louisiana	6	1.41	1.36
	Massachusetts	7	1.64	2.04
	Maryland	8	1.88	1.81
	Maine	2	0.47	0.45
	Michigan	12	2.82	2.95
	Minnesota	7	1.64	1.81
	Missouri	8	1.88	1.81
	Mississippi	4	0.94	1.13
	Montana	2	0.47	0.45
	North Carolina	13	3.05	3.18
	North Dakota	1	0.24	0.23
	Nebraska	3	0.70	0.68
	New Hampshire	2	0.47	0.45
	New Jersey	11	2.58	2.72
	New Mexico	3	0.70	0.68
	Nevada	4	0.94	1.13
	New York	26	6.10	5.90
	Ohio	15	3.52	3.40
	Oklahoma	5	1.17	1.13
	Oregon	6	1.41	1.36
	Pennsylvania	17	3.99	3.86
	Rhode Island	2	0.47	0.45
	South Carolina	6	1.41	1.81
	South Dakota	1	0.24	0.23
	Tennessee	9	2.11	2.04
	Texas	35	8.22	8.62
	Utah	4	0.94	1.13
	Virginia	9	2.11	2.49
	Vermont	1	0.24	0.23
	Washington	10	2.35	2.27
	Wisconsin	8	1.88	1.81
	West Virginia	2	0.47	0.45
	Wyoming	1	0.24	0.23
	American Samoa	1	0.24	0.23
	Guam	1	0.24	0.23
	Northern Mariana	1	0.24	0.23
	Puerto Rico	1	0.24	0.23
	Virgin Islands	1	0.24	0.23
	Washington D.C.	1	0.24	0.23

Table 3: Distribution of Representatives by States, both on Twitter and in the House of Representatives

Country	Position	# Representatives	% of Twitter	% of House
USA	Aye	146	34.27	52.64
	No	123	28.87	45.52
	NA	157	36.85	0.93
UK	Leave	85	14.63	22.34
	Remain	242	41.65	69.03
	NA	254	43.72	8.63

Table 4: Distribution of opinions on Trump's Impeachment (USA, 2019) and Brexit (UK, 2016). NAs considered to either have abstained or not in office at the time.

Finally, we also gathered information on the politicians' opinions on controversial topics such as Brexit, in the UK, and Donald Trump's first impeachment, in the USA. This information was retrieved from The Guardian and GovTrack, respectively, and is displayed in table 4.

As expected, the distribution of votes for Trump's impeachment at the time is very different from the distribution of those who were both in office in 2019 *and* are current representatives with Twitter accounts. This is because we have many more nodes for which there is missing information - namely, the representatives who were recently elected to the 118th Congress. This is also true of the opinion MPs expressed at the time of the Brexit referendum back in 2016; many of these lost their seats in the two subsequent elections, and we do not have that information for the politicians who replaced them.

3.2 Network Structure Measures

From the data, we generate two directed graphs, for the UK House of Commons and the US House of Representatives.

In comparing these two networks, some key features stand out. Firstly, we compare the number of edges and nodes of each network. Since the UK House of Commons has more political representatives than the US House of Representatives, the former will evidently have a higher number of nodes. Precisely, the UK network has 581 nodes, each representing a Member of Parliament, while the US one has 428. Perhaps due to this, we find a lower number of edges in the US network, 55,759, than in the UK one, which has 95,916 edges, essentially indicating that the UK politicians have more connections - in other words, there are more links between them, or they follow each other more, than in the case of the US.

The average degree in the UK network is 165.1, compared to 130.9 in the US. Speaker Kevin McCarthy has the highest number of fellow House members following him, with 250 in-degrees, while Congressman Raja Krishnamoorthi has the greatest out-degrees, following 321 representatives. On the other hand, Brad Finstad, Julia Letlow, Val Hoyle and Mark Pocan are not followed by any of their peers, while Tom McClintock only follows 2 of them: McCarty and Steve Scalise, the two leaders of his party. Similarly, in the UK, Boris Johnson and Rob Butler are the most and least followed MPs, respectively, at 361 and 18 in-degrees; Dame Eleanor Laing and Wes Streeting follow most of their peers, with 564 out-degrees, while the only MP that Charles Walker follows is Brandon Lewis. This points to the diversity of relationships within the virtual network and counters the idea that, because of diplomacy, all elected officials would follow each other.

To corroborate our findings, we plot below the degree distribution of both virtual networks. Since we are looking at politicians' relationships by who they follow on Twitter, we have a directed network with two different measures of degree for each node: the in-degree and out-degree. Those measures are the same for our two graphs, so we look at the total degree distributions of both networks, which essentially reflects the frequency of different degree values over the number of nodes. The distribution of the United States network looks slightly skewed to the left (Figure 1), while the UK one has a tail on the right-side (Figure 2). Though both have most of their nodes with average numbers of total degrees, the US House of Representatives seems to have a few more nodes with lower degrees on Twitter, while the UK House of Commons has some nodes with very high degrees.

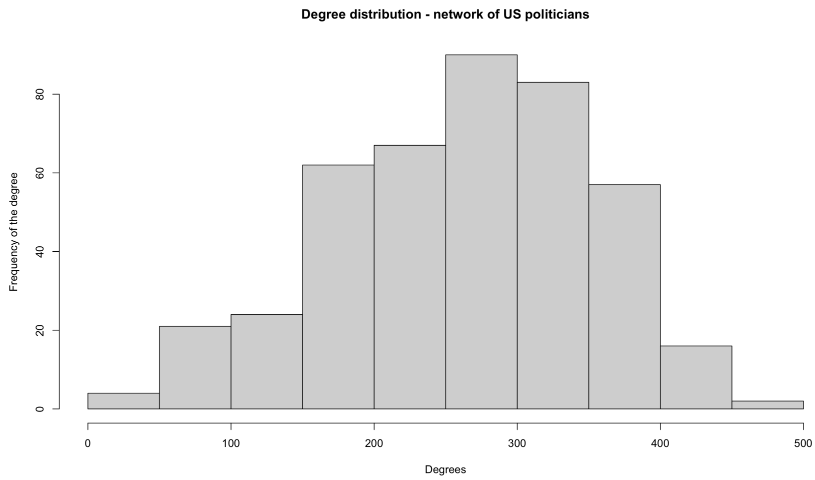

Figure 1: Histogram for the degree distribution of the US network

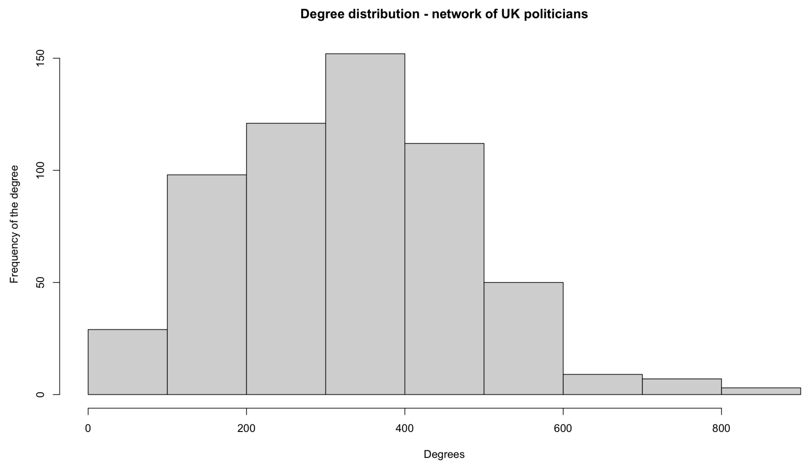

Figure 2: Histogram for the degree distribution of the UK network

We look at the densities to measure the power of the links in the network, or the interconnectivity of the links. Formally, the density of a network is determined by its ratio of links to all the potential links between each pair of nodes, therefore describing the portion of the potential connections in a network that are actual connections. Intuitively, the higher the ratio, the denser the network and therefore the more powerful its network effects are. We find that the interconnectivity in the US network (0.31) is slightly higher than in the UK (0.28), but both demonstrate a relatively average density. That is, there are many potential connections which are not realized in the network.

Another important aspect of node interaction is represented by clustering measures, which also reflect the level of connectivity between nodes. Clustering can be defined informally as "the extent that a friend of a friend is also my friend", or the proportion of closed triads or triplets[1]. At the local level, clustering captures the probability that the neighbours of a particular node are all connected or the number of pairs of neighbors of i that are connected over the number of pairs of neighbors of i. While the local clustering coefficient returns a vector with different values for every vertex, a global clustering coefficient measures the number of closed triplets (or 3 x triangles) over the total number of triplets (both open and closed). For the UK network, the value of the global clustering coefficient is 0.68, while in the US it is marginally higher at 0.72.

[1]In network theory, a triplet is identified whenever three nodes are all interconnected, forming a triangle, meaning there are no "empty" edges between any of them.

For the average clustering coefficient, which takes the mean of all the vertices' clustering coefficients, we also observe a higher value for the US (0.75) when compared to the UK (0.73). We can interpret both values as a reflection of the interconnectivity within the networks, as we know low clustering means the network will be very expansive, and we already know both networks are averagely dense. Moreover, comparing the global and average clustering coefficients allows for additional inferences about the structure of the networks. Because the average clustering coefficient places more weight on the low-degree nodes in comparison with the global coefficient, the fact that it is greater than the global measure indicates that the politicians in both houses tend to form tightly-knit clusters, more so than would be expected based on the total number of triangles in the network.

Additional assumptions can be made on the strength of the links between nodes by looking at the average path length. A path between i and j is a walk (a sequence of links) such that $i_1 = i$ and $i_k = j$ where each node in the sequence is distinct, and the shortest path between two nodes is called the geodesic path. An average path length (or average shortest path length) is therefore defined as the average number of steps along the shortest paths for all possible pairs of nodes. Essentially, it measures how efficiently information flows through a network. Again, we can assume that denser networks will have a shorter average path length because there are more options of edges to walk through. The average path length for the two networks is around 1.73, which is to be expected as the other measures have been relatively similar. The length of the longest geodesic path is 4 for the US and 3 for the UK.

Finally, we look at measures of influence to identify the most important nodes in our networks. The first and most simple one is degree centrality: it is obtained by assigning an importance score based solely on the number of links held by each node. The interpretation is straightforward - for each node, its degree centrality measures how many direct, "one hop" connections each node has to other nodes in the network. Since degree centrality is also a node-level measure, we take the mean of the degree centralities for each network and compute them: while the US has an average degree centrality of 0.62, for the UK we observe a slightly smaller one of 0.57. At the network level, a higher degree centrality suggests that the network is highly interconnected.

After obtaining scalar values for the importance of the nodes, we move to ranking them by their importance, finally identifying who are the "Beyoncés" in our data. Betweenness centrality and eigenvector centrality are two important measures of influence: while the former is a measure of centrality based on shortest paths, the latter takes into account the centrality of its neighboring nodes. Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is therefore used to find nodes that serve as a bridge from one part of a graph to another, with the algorithm calculating the shortest paths between all pairs of nodes in a graph. The ranks of betweeness centrality are presented in table 5.

Raja Krishnamoorthi, a Democrat representing Illinois's 8th congressional district, is the most influential member in terms of betweenness centrality. That means he is well-connected, lying on many of the shortest paths between other nodes. Kevin McCarthy, the Republican representative for California's 20th congressional district and current Speaker of the House, has the second highest betweenness centrality. 7 of the highest ranking members in terms of betweenness centrality are Democrats. The UK's Speaker of the House, Lindsay Hoyle, also ranks highly in this measure, followed in ranking by mostly Conservative MPs and recent Ministers.

The higher the eigenvector centrality of a node, the more influential it is within the network. We use two different measures of eigenvector centrality and rank the politicians by their influence. We start by computing the eigenvector centrality of each node, then compute the normalized and rescaled eigenvector centrality measures by dividing the raw eigenvector centrality by the maximum raw eigenvector centrality, and multiplying it by the number of nodes in the network, respectively. If all nodes in the network have positive values of eigenvector centrality, then the maximum value of eigenvector centrality will be equal to the largest value of raw eigenvector centrality in the network, and hence, the normalized eigenvector centrality will also be equal to the raw eigenvector centrality. In summary, since all the nodes in our graph have positive values of eigenvector centrality, the normalization step does not affect the values of the centrality measures. We therefore keep the normalized eigenvector, the rescaled eigenvector rank the nodes by those two measures, selecting the top 10 nodes which are presented in table 6 below.

Country	Twitter Handle	Betweenness Centrality	Centrality Rank
USA	CongressmanRaja	3265.64	1
	SpeakerMcCarthy	2533.61	2
	RepRussFulcher	2140.06	3
	CongressmanGT	1889.25	4
	RepChrisPappas	1883.81	5
	RepDarrenSoto	1877.41	6
	RepGolden	1684.96	7
	SpeakerPelosi	1676.96	8
	RepTedLieu	1589.84	9
	RepTerriSewell	1550.28	10
UK	LindsayHoyle_MP	9111.00	1
	wesstreeting	8355.63	2
	eleanor4epping	7169.00	3
	JamesCleverly	5931.33	4
	nadhimzahawi	5428.62	5
	darrenpjones	4990.18	6
	LukePollard	4453.74	7
	mattwarman	3867.48	8
	AndrewRosindell	3694.02	9
	TomTugendhat	3264.00	10

Table 5: Rank of Representatives by Betweenness centrality

It is interesting to note that the ranking for the United States looks very different than in terms of betweenness centrality. Nancy Pelosi has highest eigenvector centrality meaning she is connected to other nodes that are themselves well-connected, which makes sense considering she served as Speaker of the House for eight years. Behind her is Steny Hoyer, previous House Majority Leader, and Hakeem Jeffries, current House Minority Leader. Interestingly, the top 10 does not include any Republican politicians and has 6 representatives from California, 3 from the DMV region (DC, Maryland, Virginia) and 1 from New York. Similarly, the best connected node in terms of links to other well-connected people is Boris Johnson, former Primer Minister of the UK. He is followed by Liz Truss, also a recent Prime Minister, Michael Gove, a current Minister, and Rishi Sunak, the current Prime Minister. All of the politicians at the top of the ranking have recently held prominent roles in the government and are from the Conservative party.

At last, we measure reciprocity in the networks, or how much people are followed by those who they follow. In the UK House of Commons, this is 52%, while in the US House of Representatives it is roughly 57%.

From these measures, we can draw the following conclusions: the virtual political networks of the UK and the US are quite similar, despite their different political systems. Though the UK network has more MPs and overall connections, the densities of the two are not dissimilar, with the US having a slight edge. They are both quite clustered and, thus, not that expansive, despite a small average path length and diameter. The reciprocity in both Houses is around 50%, which suggests a diverse range of relationships within the virtual network, contradicting the prior that all elected officials would follow each other.

Finally, individual measures of centrality point to the stark differences between importance in the House and influence on the general population. Popular elected officials such as Alexandria Ocasio-Cortez, Marjorie Taylor Greene, Jeremy Corbyn and Ed Miliband are nowhere to be found in either of the rankings; in fact, the first two place in 280/118 and 403/382 in betweenness and eigenvector centrality, respectively, while the latter are in 62/283 and 120/216.

Country	Twitter Handle	Normalized Eigenvector	Rescaled Eigenvector	Centrality Rank
USA	SpeakerPelosi	1	428.00	1
	RepStenyHoyer	0.90	386.61	2
	RepJeffries	0.88	378.35	3
	RepBarbaraLee	0.86	369.30	4
	RepLindaSanchez	0.86	368.60	5
	RepSwalwell	0.84	357.97	6
	RepRaskin	0.83	355.34	7
	RepSpanberger	0.83	353.89	8
	RepJudyChu	0.82	352.98	9
	RepTedLieu	0.82	352.28	10
UK	BorisJohnson	1	581.00	1
	trussliz	0.96	557.87	2
	michaelgove	0.94	545.88	3
	RishiSunak	0.94	543.17	4
	MattHancock	0.93	540.34	5
	nadhimzahawi	0.93	538.06	6
	JamesCleverly	0.92	535.39	7
	TomTugendhat	0.92	534.32	8
	sajidjavid	0.91	529.20	9
	GregHands	0.90	525.67	10

Table 6: Rank of Representatives by Eigenvector centrality

4 Methodology

In this section, we implement algorithms to predict missing node attributes and identify relevant clusters. For the former, we use the local smoothing algorithm to interpolate political opinions on Brexit and Trump's impeachment. Additionally, we use community detection algorithms such as Infomap and DC-SBM to understand how the nodes' attributes interact with their predicted community structure.

4.1 Local Smoothing Algorithm

The local smoothing algorithm is a method for predicting missing attributes in a network graph. The idea is to use the observed attributes of a node's neighbors to infer the missing attribute value of the node. It works by, first, identifying the neighbors of the node with the missing attribute, and then using the observed attribute values of these neighbors to calculate the mode for the missing attribute value, in the case that it is a categorical variable. We use the local smoothing algorithm to predict Brexit opinion and impeachment votes for the 43.4% and 35.9%, respectively, of elected officials who are missing data on that topic.

To understand whether our local smoothing algorithm is accurately predicting the missing information, we randomly delete a fraction of the observed votes and split the dataset into test and train. At each fraction, we predict the vote classification for the test set. The accuracy of the algorithm for Brexit and impeachment varies depending on the share of missing data – for the UK, where roughly 44% of the attribute is missing, we expect an accuracy of around 96%, with 98% for the US as it has slightly less unavailable data.

4.2 Community Detection

A community is defined, in its most straightforward manner, as any group of nodes that present a higher likelihood of connecting within that group than to nodes from other communities. In the case of our data, it is fair to assume that the network structure is not random - we can think that politicians from the same party will follow each other more than they follow those from opposing parties, likewise with the region. To understand the main determinants of community formation within our available variables, we apply different community detection algorithms: Infomap and Degree-Corrected-Stochastic Block Model (DC-SBM). Additionally, we plot the community structure along with relevant politician characteristics such as region, party and votes in relevant subjects such as Brexit and the impeachment of Donald Trump. Essentially, we want to answer the question of whether certain political attributes of the nodes in our networks are responsible for the composition of communities.

One of the most important concepts to properly understand the community detection methods is the **modularity**. In formal terms it can be defined as the measure that captures systematic deviations from a random configuration, for each possible partition of nodes. If the modularity is high, that partition separates groups with many more internal connections than expected at random. We also compute the modularity for each of the community detection algorithms.

4.2.1 InfoMap

The general idea of the **InfoMap** algorithm is to find communities by minimizing the description length of random walks. This algorithm is based on the idea that the optimal community structure should minimize the amount of information required to describe a random walk on the network. The concept was introduced by Rosvall and Bergstrom (2008) and it identifies communities as groups of nodes that have a higher probability of being visited by a random walker within the same community compared to other communities.

The advantage of this algorithm is that, by minimizing the map equation, it exploits the duality between finding cluster structures in networks and minimizing the length of a random walk (Bohlin et al., 2014). The application in R can be done with the **igraph** package, and the results are detailed in the respective section.

4.2.2 Stochastic Block Models: DC-SBM

The Degree-Corrected stochastic block model (DC-SBM) is an extension of the traditional SBM that incorporates the node degree heterogeneity, correcting the problem that happens when you apply the SBM to networks with skewed degree distributions where the model tends to group vertices by degree. This allows the algorithm to perform better than others in real-world networks that follow a heterogeneous degree distribution. DC-SBM aims to find a partition of the nodes into communities such that the connections between nodes are well-explained by the model, considering both the community memberships and the node degrees.

5 Results

Finally, we present the results from our missing attributes prediction and community detection and discuss the implications for our research question.

5.1 Missing Attributes

The predicted opinions of politicians with regard to Trump's impeachment and Brexit are shown in table 7.

Country	Position	# Representatives	% of Twitter	% of House
USA	Aye	208	48.83	52.64
	No	218	51.17	45.52
	NA	0	0.00	0.93
UK	Leave	124	21.34	22.34
	Remain	457	78.66	69.03
	NA	0	0.00	8.63

Table 7: Distribution of predicted opinions on Trump's Impeachment (USA, 2019) and Brexit (UK, 2016) from the local smoothing algorithm.

We had 146 Democrats voting in favor of Trump's impeachment, 1 voting against, and 62 for whom we had no data on. Similarly, 122 Republicans voted against impeaching Trump and the remaining 95 were missing information as they were not in office at the time. With the current House composition and network structure, however, the local smoothing algorithm predicted that 61 missing Democrats would have voted in favor and 1 against, while 94 missing Republicans would have voted against, with 1 in favor.

On the other hand, the distribution of Brexit opinions for the two largest parties in the UK is a bit different. We had 123 Labours voting to remain in the European Union, while 1 voted to leave and 68 were missing information. The Conservatives were more evenly split: 75 spoke out in favor of remaining, while 78 wanted to leave the EU, with the attribute missing for 145. While the algorithm predicted that all 68 missing from the Labour party would choose remain, it predicted that only 38 Conservative MPs would vote to leave and the 107 left would prefer remain. This suggests that, despite a fifty-fifty split in the party for those whose preferences we could observe, MPs would vote according to their neighbors and a significant share would side with the majority of the House, which implies more cross-party interactions than in the United States.

It is relevant to note that we only used information on the unobserved politicians' political party in the case where none of their neighbors had a voting record. This would imply that, if they could not make up their opinion from their neighbors' opinions, they would follow the party majority. The fact that predicted votes were so evenly split by party lines in the US suggests that the House of Representatives tends to a very polarized virtual network that models reality based on voting clusters. It also suggests that, with the current composition of the 118th US Congress, Trump's first impeachment process would not have passed in the House. Alternatively, MPs in the House of Commons might be interacting more with key representatives online and not only those whose ideology they align with.

5.2 Communities

By creating our data from information from different sources, we were aware that it would not be possible to validate our community assignment algorithms. Therefore, we perform a descriptive analysis of our results for two clustering algorithms: DC-SBM and Infomap. Since the results for the first one were less intuitive, considering the structure of the network, we focus on the InfoMap community structure, delving deeper into its interaction with other political characteristics. We conclude that the InfoMap algorithm is better at predicting the network structure via community assignment, at least according to observed attributes and the observed network statistics.

5.2.1 Degree-Corected Stochastic Block Model

With the DC-SBM algorithm, the results are not in line with our findings on the network structure. We verify our hypothesis by analyzing clustering attributes for DC-SBM, and find that the community assignment is highly sensitive to different specifications. This essentially means that adding a new member could alter the community assignment significantly, therefore altering the observed community structure.

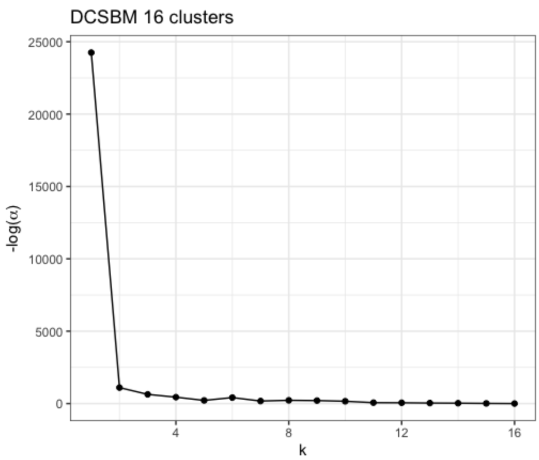

Figure 3: Path for the US network.

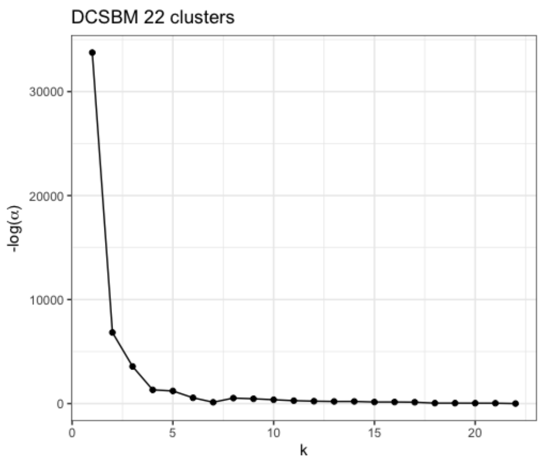

Figure 4: Path for the UK network.

Looking at the path graphs and considering that we have about the possible partitions of the political environment can see that 16 (for US) and 22 (for UK) clusters are overfitted results. For the US, after the second and the third cluster seems the ideal partition, since after that a small change in the regularization would result in the emerge of an extra cluster. For the UK network, the same is true but around the fourth and fifth partitions, which was expected given the more plural political environment of the UK considering the regime.

The suggested partition can be seen in the figure. The US partition reported a Integrated Classification Likelihood (ICL) of -85824.49 and the UK partition reported a ICL of -160382, indicating that the partition found in the US political network is better, since the higher the ICL value, the better the model is at explaining the data.

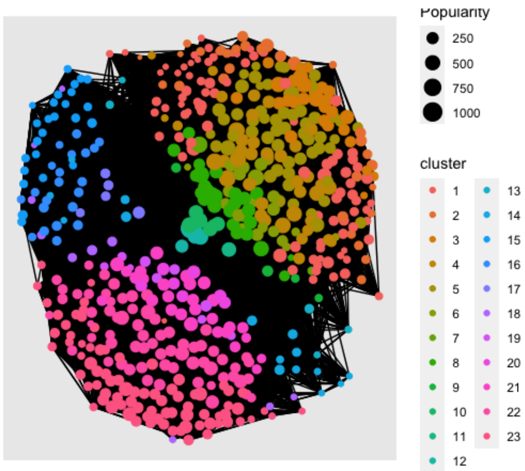

Figure 5: Communities on the UK network with the DC-SBM model.

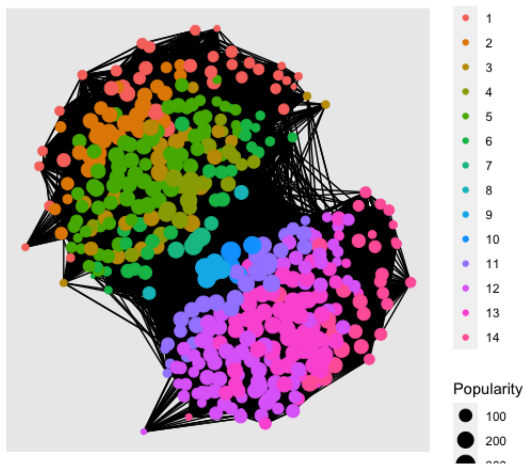

Figure 6: Communities on the US network with the DC-SBM model.

As the results of the DC-SBM indicated several groups that in our intuitive knowledge do not make sense considering the network structure, we chose to not proceed with the interpretation of the results, and move forward by interpreting only our results for the Infomap one.

5.2.2 Infomap

Since this algorithm partitions nodes by minimizing the information needed to describe the movement of a random walker, we can say that the predicted communities reflect how strong the links between nodes are and at the same time how fast information spreads within the network. Additionally, we can compute the modularity for the predicted communities to measure the strength of the division procedure, as we do not have a validation community structure to confirm our community assignment is correct. That being said, it is useful to compare how the community structures behave with respect to other characteristics, such as region, political party and important voting procedures within the two houses.

We compute modularity of 0.429 for the US and 0.376 for the UK. This is in line with our previous results for the networks summary statistics section, which identifies stronger links and a more dense network in the case of the US when compared to the UK. Low modularity would mean that a change in one of the components - in this case, nodes - would alter the structure of the communities. This further corroborates our hypothesis that the US's network is more tightly knit than the one in the UK, which is more disperse. We plot the results of the community detection below:

We find that while the US network is partitioned into two communities, with 208 and 218 nodes respectively, the UK presents 5 partitions, with decreasing sizes. The first UK community has 311 nodes, the second 197, and the third, fourth, and fifth, respectively, have 15, 51 and 7 nodes each. The latter, therefore, might suggest a hierarchical or hierarchical-modular organization, where some communities are much larger than others, potentially indicating a core-periphery structure or other similar organizational patterns. We could think of this structure as either region or party-based: we know that the two biggest parties in the UK are Labour and Conservative, therefore they could be the ones at the top of the hierarchy. The same thing can be said about regions, where the ones closer to London and other major British

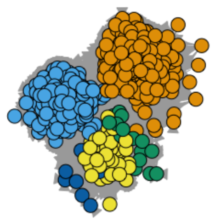

Figure 7: Communities on the UK network with the Infomap model.

Figure 8: Communities on the US network with the Infomap model.

capitals are the ones concentrating the bigger number of nodes. We verify our hypothesis by plotting the community membership along with region and party, for both networks.

We first compare our community structure with the distribution of parties within the two political systems: while both are mainly bipartisan, the UK has a bigger dispersion of votes. This is due to some regional parties being the ones solely responsible for voting on local issues of the different countries included in the parliament (namely Northern Ireland, Wales and Scotland). In that sense, we could attribute the assignment of only two communities for the US when compared to 5 in the UK as a result of stricter bipartisanship. The graphs below confirm our hypothesis, as they reveal a clear split between parties and communities in the US, and suggest a hierarchical organization of the parties as well as the communities in the UK.

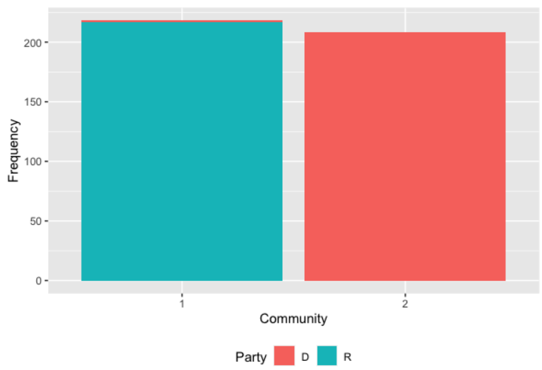

Figure 9: Distribution of parties within predicted communities for the US network.

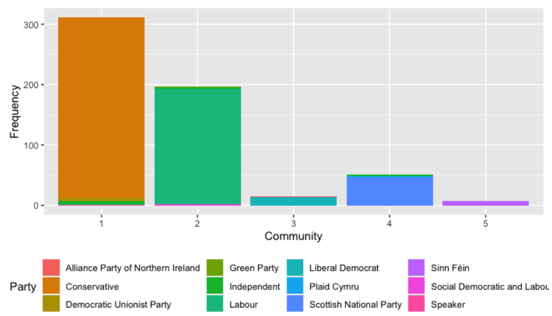

Figure 10: Distribution of parties within predicted communities for the UK network.

We proceed with our descriptive hypothesis testing by plotting the UK's community structure along with party and region, to understand if regional and partisan relationships intertwine in the community assignment via Infomap. Since the US's greater level of geographic aggregation is states, and there are 50 of them, we choose to restrict our visualization exercise to the UK's 13 regions.

In Figure 11, one feature stands out: the biggest parties (Labour and Conservatives) are concentrated in, at most, two regions and two communities. Intuitively, it would not make sense that the two major UK parties are only found in the regions of Eastern, South West and North West - otherwise they would not be the majority. Additionally, it is known that the London region only has one Independent MP, which contradicts its assignment to the second community. This suggests that community structure is mostly partisan, and therefore regional relationships are not a big factor in the community structure found on the MP's Twitter accounts.

In light of connecting our analysis of community structures to our results for node prediction and

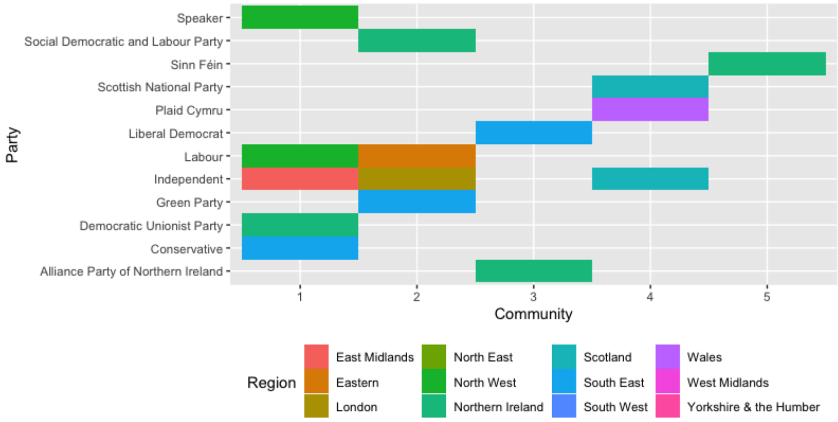

Figure 11: Distribution of parties and regions within predicted communities for the UK network.

therefore further confirming our hypothesis that Twitter interactions are mostly related to partisanship according to the available variables in our dataframe, we plot the community structure along with party and the (predicted and observed) results in the two most important recent voting procedures in both legislatures.

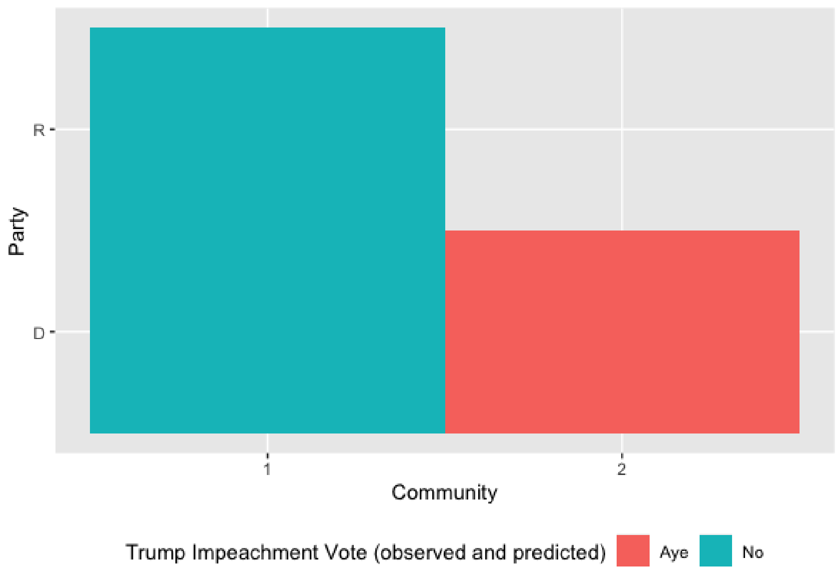

Figure 12: Votes for Trump impeachment within parties and predicted communities for the US network.

We find that, in line with our results in the Missing Attributes section, Trump's first impeachment process would not have passed in the House if it had the current composition: while all of the Republicans, classified into community 1, would vote no, some of the democrats classified in the same community would also have rejected the impeachment process. This suggests that the Democrats assigned membership into the same community as the Republicans are not only interacting more with the "other side" via Twitter, but they are also more aligned with their political opponents in terms of relevant policies.

In the case of the UK, the results are even more surprising, as they suggest that with the current composition of the House of Commons, split into communities according to their virtual interactions, the great majority would have voted "Remain", with only the Democratic Unionist Party and the Independents, both assigned to community 1, voting to leave the EU. This result also suggests closer ideological leaning between MPs who follow each other on twitter, after all their predicted communities point towards a common vote in the polemic Brexit vote.

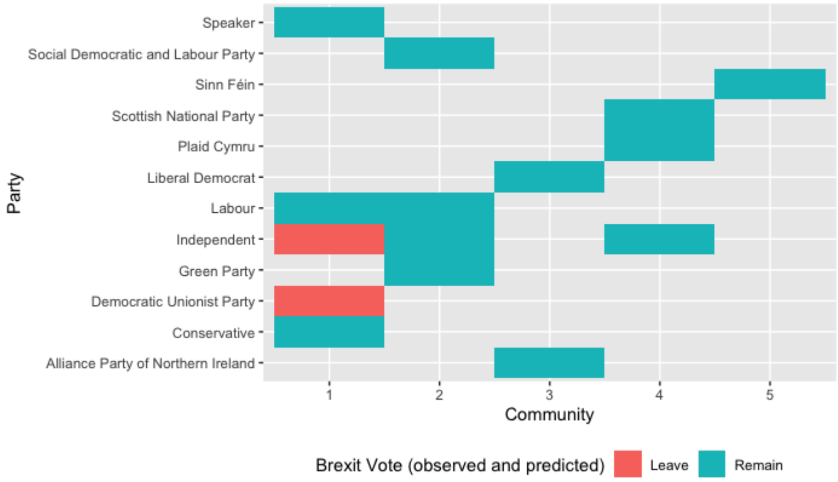

Figure 13: Votes for Brexit within parties and predicted communities for the UK network.

6 Conclusion

In this paper, we aimed to investigate, using network theory tools, what are the main drivers of the interaction between politicians on Twitter. To do that, we collected relevant data on the elected representatives of two of the biggest democracies' elected officials for the legislative house - the House of Commons of the UK, and the US Congress. We analyzed how the networks were shaped by creating a directed graph object for each of the countries, where an edge existed between two nodes whenever a politician followed another on Twitter. Our main goal was to understand how the virtual activity of elected officials interferes with their political activity, and how it is shaped by partisan and regional ties.

We find evidence in contrast to a "diplomatic" hypothesis where all the elected officials would follow each other. Both networks present a below-than-expected density and, therefore, connectivity: 0.31 in the US network and 0.28 in the UK one. These results suggest that there are many potential connections not realized in the network. To confirm our hypothesis about the network structures, we evaluate clustering measures both at the global and local levels and find again a marginally higher degree of connectivity in the US when compared to the UK: 0.72 for the global clustering coefficient in the American network against 0.68 in the British one. At the local level, we find an even smaller difference, suggesting that clustering at the local level is even stronger than by analyzing the entire network. In other words, we observe that nodes are more tightly connected at the local level, and not necessarily to the entire network.

Another interesting finding relates to influence measures: we compute the degree centrality at the node level, as well as the average one for each of the networks. We again observe a small difference between the average values of centrality for both networks, but we use measures of betweenness centrality and eigenvector centrality to verify who are the most important nodes in our networks. Our results are counterintuitive to what is observed anecdotally: the House Speakers are not necessarily the most influential nodes. In the case of the US, it is Raja Krishnamoorthi, a Democrat representing Illinois, who has the biggest influence in terms of betweenness centrality. Additionally, popular elected officials that have a large media presence are not necessarily ranked highly in the measures of influence computed. We can therefore draw conclusions on the similarity of virtual political networks of the UK and the US, despite their different political systems. Though the UK network has more MPs and overall connections, the densities of the two are not dissimilar: both are quite clustered and, thus, not that expansive. The reciprocity in both Houses is around 50%, suggesting a diverse range of relationships within the virtual network.

In light of exploring the community structure even further and using the observed connections to make predictions, we apply a Local Smoothing Algorithm to predict missing attributes in our networks. We use this method to predict Brexit opinion (in the UK) and impeachment votes (for Trump, in the US) for the 43.4% and 35.9% of elected officials who are missing data on that topic but are present on Twitter. While the predictions for the uS were more in line with a bipartisan split, the story was a bit different in the case of the UK. The distribution of Brexit opinions for the two largest parties, Labour and Conservative, led the algorithm to predict that all 68 missing from the Labour party would choose remain, while only 38 Conservative MPs would vote to leave and the 107 Conservatives left would prefer remain. This suggests that, despite a fifty-fifty split in the party for those whose preferences we could observe, MPs would vote according to their neighbors and a significant share would side with the majority of the House, implying a biggest cross-party interaction than in the United States.

Finally, we use community detection methods to evaluate the interaction between nodes, interacting the predicted communities with other relevant characteristics such as party, region and the aforementioned votes. We first use the Degree-Corrected Stochastic Block Model algorithm for predicting communities, and find a high sensitivity of the community assignment to different network specifications. This means that many of the predicted clusters are overfitted, and therefore the results indicate several groups that are not in line with the observed network structure. Thus, we use the Infomap algorithm, and find a better prediction of communities according to attributes, which we validate by plotting the community assignment with the expected attributes that define the interactions on Twitter, such as party, region and votes. We find that community assignment is mostly correlated with partisanship, which is stronger in the case of the US, with almost all the Democrats belonging to one community, and all Republicans belonging to another. This is also the case for the predicted and observed votes, but not the case for region, which confirms our hypothesis that politicians from the same party interact more with each other than they do with members of other parties on Twitter.

References

Bastos, M. T. (2012). Public opinion revisited: The propagation of opinions in digital networks. *Journal of Arab Muslim Media Research.*, 4(2):185–201.

Benkler, Y. (2007). The wealth of networks: How social production transforms markets and freedom.

Bohlin, L., Edler, D., Lancichinetti, A., and Rosvall, M. (2014). *Community Detection and Visualization of Networks with the Map Equation Framework*, pages 3–34.

Bruns, A. and Highfield, T. (2016). *Is Habermas on Twitter? Social Media and the Public Sphere.*

Buccoliero, L., Bellio, E., Crestini, G., and Arkoudas, A. (2020). Twitter and politics: Evidence from the us presidential elections 2016. *Journal of Marketing Communications*, 26(1):88–114.

Graham, T., Broersma, M., Hazelhoff, K., and van 't Haar, G. (2013). Between broadcasting political messages and interfering with voters: The use of twitter during the 2010 uk general election campaign. *Information Communication Society*, 16(5):692–716.

Gruzd, A. (2012). Investigating political polarization on twitter: A canadian perspective. *Internet, Politics, Policy.*

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology Politics*, 13(1):72–91.

Kruikemeier, S. (2014). How political candidates use twitter and the impact on votes. *Computers in Human Behavior*, 34:131–139.

Margaretten, M. and Gaber, I. (2012). The crisis in public communication and the pursuit of authenticity: An analysis of the twitter feeds of scottish mps 2008-2010. *Parliamentary Affairs*, 67(5):328–350.

Mislove, A., Viswanath, B., Gummadi, K. P., and Druschel, P. (2010). You are who you know: inferring user profiles in online social networks. *In Proceedings of the third ACM international conference.*

Pariser, E. (2011). *The Filter Bubble.* The Penguin Press, New York.

Recuero, R., Zago, G., and Soares, F. B. (2017). Mídia social e filtros-bolha nas conversações políticas no twitter. *In Proceedings of XXVI Encontro Anual da Compós. Compós, São Paulo. Brazil.*

Rosvall, M. and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.

Silva, B. C. and Proksch, S.-O. (2017). Political polarization in social media: Analysis of the "twitter political field" in japan. *IEEE International Conference on Big Data (BIGDATA).*

Silva, B. C. and Proksch, S.-O. (2022). Politicians unleashed? political communication on twitter and in parliament in western europe. *Political Science Research and Methods*, 10:776–792.

Small, T. A. (2011). What the hastag? *Information, Communication Society*, 14(6):872–895.

Smith, M., Rainie, L., Himelboim, I., and Shneiderman, B. (2014). Mapping twitter topic networks: From polarized crowds to community clusters.

Soares, F. B., Recuero, R., and Zago, G. (2018). Influencers in polarized political networks on twitter. *Proceedings of the 9th International Conference on Social Media and Society*, page 168–177.

Statista (2022). Number of twitter users worldwide from 2019 to 2024.

Wojcieszak, M. E. and Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication*, page 40–56.

Yoon, H. Y. and Park, H. W. (2014). Strategies affecting twitter-based networking pattern of south korean politicians: social network analysis and exponential random graph model. *Qual Quant*, 24:409–423.