Ishaan Jhaveri
Instructor: Jon Kleinberg
CS 6850, Spring 2018: Structure of Information Networks
16 May, 2018

Differences in network properties of "Early Adopters" of the Leave and Remain Campaigns in the Brexit referendum of 2016

Introduction

On June 23, 2016 the citizens of the United Kingdom of Great Britain and Northern Ireland (UK) voted by referendum for the UK to leave the European Union (EU). The vote was very close, and the issue very contentious. Analysis published after the result was announced[1] suggested that the side campaigning for "Remain" (i.e. advocating that the UK should remain in the EU) didn't campaign as hard as the side campaigning for "Leave" (i.e. advocating that the UK should leave the EU). In this work I analyze interactions between Facebook Users advocating for one of the sides of the argument (So called "leave early adopters" or "remain early adopters") and other Facebook Users to see whether early adopters have any meaningful influence on other users' opinion on the Brexit debate. I perform this analysis on various Facebook networks, some in which leave early adopters are more active, others in which remain early adopters are more active and still others where both kinds of adopters are active. I then analyze various aspects of the structure of all these networks to see if, at least from a network theoretic standpoint, there is some suggestion that the leave campaigners were indeed more effective at campaigning on social media.

Data

In this section I will first explain some vital aspects of the dataset, then explain how the data is used to construct interaction networks and finally explain how the dataset is used to identify early and late adopters.

Political Parties' Posts

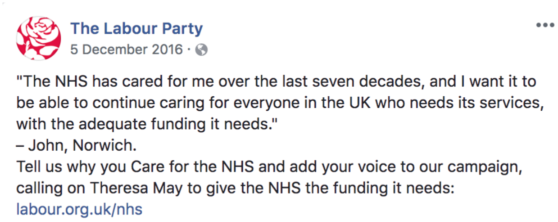
There are two types of data for each political party, "posts" and "comments". For example consider the liberal Labour Party. Figures 1a and 1b show an example post and some comments on this post from the their official Facebook page:[2]

We represent posts as:

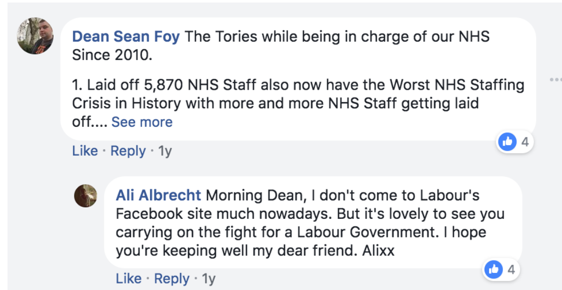
```
{'_id': ObjectId('5848dc89e153d4d48be7f018'),
 'link_name': "Join Labour's campaign to care for the NHS >>",
 'num_comments': 55,
 'num_reactions': 226,
 'num_shares': 85,
 'page_id': 'labourparty',
 'permalink_url': 'https://www.facebook.com/labourparty/posts/10154026776332411',
 'reaction_ANGRY': 3,
 'reaction_LIKE': 211,
 'reaction_LOVE': 8,
```

[1] Rafael Behr. *How remain failed: the inside story of a doomed campaign — Rafael Behr*. 2016. URL: https://www.theguardian.com/politics/2016/jul/05/how-remain-failed-inside-story-doomed-campaign.

[2] *The Labour Party*. URL: https://www.facebook.com/labourparty.



(a) Example post from Labour Party



(b) Example comments on a post from the Labour Party

Figure 1

```
'reaction_SAD': 2,
'reactions_list': {'1005046109613228': 'LIKE',
   ...
 '1307502705946681': 'SAD',
   ...
 '924679667661179': 'LIKE'},
'status_id': '25749647410_10154026776332411',
'status_link': 'http://labour.org.uk/nhs',
'status_message': '"The NHS has cared for me over the last seven decades, and I want it to be able to
     continue caring for everyone in the UK who needs its services, with the adequate funding it needs."\
     n   John, Norwich.\nTell us why you Care for the NHS and add your voice to our campaign, calling on
     Theresa May to give the NHS the funding it needs: labour.org.uk/nhs',
'status_published': '2016-12-05 10:12:00',
'status_type': 'shared_story',
'timestamp': datetime.datetime(2016, 12, 5, 10, 12)}
```

The `reactions_list` variable here is of interest. It gives the unique IDs of all the Facebook users that have “reacted to” this post. Reactions include “Like”, “Love”, “Angry”, “Sad” and “Wow”. For simplicity we will assume all reactions mean the same thing and just say a user liked a post or a comment even if they actually used one of the other reactions.

And we represent all the comments on a given post, like the one above as:

```
{'_id': ObjectId('5849448fdf6f4577cf09d964'),
 'author_id': '1821238044779060',
 'comment_author': 'DELIBERATELY REMOVED',
 'comment_id': '10154026776332411_10154034664507411',
 'comment_message': "Morning Dean, I don't come to Labour's Facebook site much "
                    "nowadays. But it's lovely to see you carrying on the "
                    "fight for a Labour Government. I hope you're keeping well "
                    'my dear friend. Alixx',
 'like_count': 2,
 'likes': ['10210476718460086', '10210257587784676'],
 'page_id': 'labourparty',
 'parent_id': '10154026776332411_10154034641577411',
 'position': 2,
 'status_id': '25749647410_10154026776332411',
 'timestamp': datetime.datetime(2016, 12, 5, 11, 10, 48)}
```

The `author_id`, `num_replies/parent_id` and `position` variables here are of interest. `author_id` is the unique ID of a given Facebook User as discussed above. In this case it is the ID of the user that authored the comment. `num_replies` specifies the number of users who have commented in response directly to this comment. For comments in a reply thread as in the second comment above `num_replies` is replaced with `parent_id` which gives the original comment at the head of the reply thread that this comment occurs in, and `position` which gives the position of this comment in the reply thread.

Interaction Networks

The above data is used to construct two types of interaction networks, the "Reply" network and the "Like" network:

(a) Nodes and Edges in Reply Network (b) Nodes and Edges in Like Network

Figure 2

The "Reply" Network (Figure 2a): The nodes are all users who commented on any post posted by the official Facebook group of a given political party. There is an edge between a node C and a node D if C replied to D's comment. This means C's comment has to be the first comment in the reply thread of D's comment, or if D's comment is itself in a reply thread, C's comment has to be the next comment in this thread. In the networks we use here, C and D's comments have to be adjacent as described, but a version of the network could be created where there is an edge between C and D if their comments were within some window size w of each other. Also for simplicity, the edges are undirected. A version of the network could be created where the edges are directed and C has an edge to D (C is responding to D, and therefore C is potenitally influenced by D) but D doesn't have an edge to C (unless of course D responds to C at some other point).

The "Like" Network (Figure 2b): The nodes are all users who commented on any post posted by the official Facebook group of a given political party and all users who liked any comment on any post posted by the official Facebook group of a given political party. There is an edge between a node A and a node B if A liked B's comment. For simplicity, the edges are undirected. A version of the network could be created where the edges are directed and B has an edge to A (A liked B's comment, and therefore A is potentially influenced by B) but A doesn't have an edge to B (unless of course B liked some other comment made by A).

Identifying Early and Late Adopters

The question of whether the United Kingdom should leave the European Union had been debated nationally with different levels of attention since 1977[3] but talk of the referendum of June 2016 only started gaining traction in 2013, when under pressure from some members of his party and the conservative UKIP party David Cameron (leader of the center-right leaning party, Conservatives) announced that if the Conservatives won the general election in 2015 he would hold a referendum to decide whether the UK should leave the EU.[4] The Conservatives did win in 2015, and on the 22nd of February, 2016 Cameron (now Prime Minister) announced that the

[3] *Brexit*. 2018. URL: `https://en.wikipedia.org/wiki/Brexit#Opinion_polls_19772015`.
[4] *Brexit*. 2018. URL: `https://en.wikipedia.org/wiki/Brexit#Negotiations_for_EU_reform`.

referendum would take place on the 23rd of June, 2016.[5] Therefore, between February 22 and June 23 both the Leave and Remain campaigns received increased media coverage.[6] Given this increase in media attention after the official announcement, I define "early adopters" as Facebook users who engaged with either the Leave or the Remain campaign before the official announcement date. "Late adopters" are Facebook users who engaged with either the Leave or the Remain campaign between the announcement date and the date of the vote itself.

Engagement with the Leave campaign is defined as liking a post on the Leave campaign's official Facebook group "Vote Leave". So "leave early adopters" are defined to be those Facebook users that liked any post posted by the Vote Leave group before February 22, 2016.

Engagement with the Remain campaign is defined as liking a post on the Leave campaign's official Facebook group "Open Britain" (formerly named to "Stronger in Europe"). So "remain early adopters" are defined to be those Facebook users that liked any post posted by the Open Britain group before February 22, 2016.

Note the data doesn't contain the dates of the likes, so its possible that someone who went back and liked an old post from before February 22 is incorrectly treated as an early adopter, but because of the frequency of posts that are available to engage with, we can assume that most people who engage with a group mostly engage with posts posted during the time they are engaging, and not before.

The Actual Created Networks

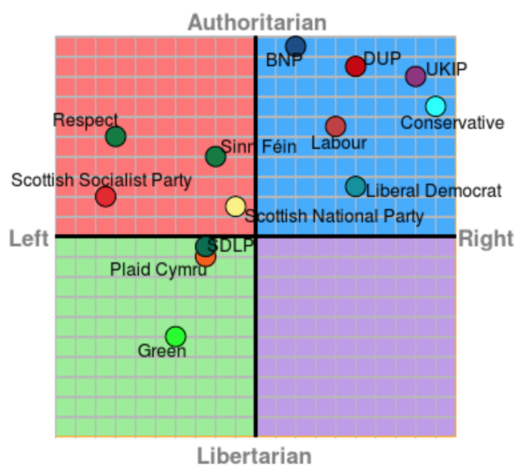


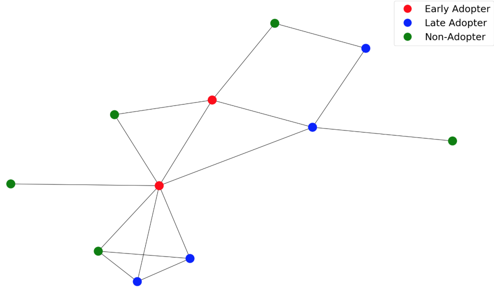
Figure 3: UK Political Spectrum[7]

Using the above data I created 24 different networks. 4 types:

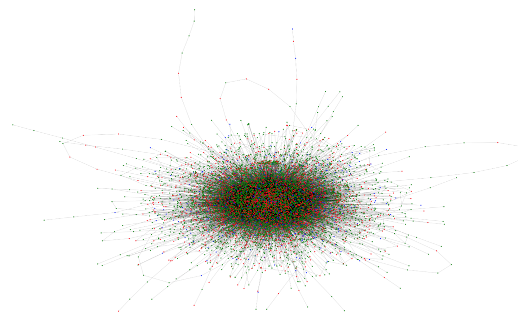
- **Reply Leave** - The Reply Network with Leave Early Adopters and Leave Late Adopters identified

[5] Hansard, Westminster, and House of Commons. *House of Commons Hansard Debates for 22 Feb 2016 (pt 0001)*. URL: `https://publications.parliament.uk/pa/cm201516/cmhansrd/cm160222/debtext/160222-0001.htm#16022210000001`.

[6] ideas@global-initiative.com Global Initiative. *UK press coverage of the EU Referendum*. URL: `http://reutersinstitute.politics.ox.ac.uk/our-research/uk-press-coverage-eu-referendum`.



(a) A simplified example of one of the networks



(b) The UKIP Reply Leave Network

Figure 4

- **Like Leave** - The Like Network with Leave Early Adopters and Leave Late Adopters identified
- **Reply Remain** - The Reply Network with Remain Early Adopters and Remain Late Adopters identified
- **Like Remain** - The Like Network with Remain Early Adopters and Remain Late Adopters identified

for each of the following 6 political parties from the left to the far right. Each of the networks looks at the specific interactions of that group between a particular date and the date of the referendum result announcement. This particular date varies by network, and is generally chosen so that the network has close to 10000 nodes. There are two outliers which are discussed below:

1. **Scottish National Party**: Depicted in Figure 3 as a yellow dot to the middle left, SNP leans left and was generally pro-Remain leading up to the referendum.[8]
   (a) The SNP Reply Networks' start date (6 months before the referendum result announcement) was chosen so as to constrain it to 7498 nodes but I removed all but the main connected component such that it contains 7151 nodes.
   (b) The SNP Reply Leave Network has 302 leave early adopters and 846 leave late adopters.
   (c) The SNP Reply Remain Network has 183 remain early adopters and 574 remain late adopters.
   (d) The SNP Like Networks' start date (1 month before the referendum result announcement) was chosen so as to constrain it to 9320 nodes but I removed all but the main connected component such that it contains 9141 nodes.
   (e) The SNP Like Leave Network has 397 leave early adopters and 1412 leave late adopters.
   (f) The SNP Like Remain Network has 210 remain early adopters and 740 remain late adopters.

[8]Guest Submission. *Where the UK Political Parties Stand on Brexit SJS*. 2017. URL: `http://www.socialjusticesolutions.org/2017/06/05/uk-political-parties-stand-brexit/`.

2. **Labour**: Depicted in Figure 3 as a maroon dot in the middle of the upper right quadrant, Labour leans relatively left and was generally pro-Remain leading up to the referendum.[9]
   (a) The Labour Reply Networks' start date (1 year before the referendum result announcement) was chosen so as to constrain it to 8990 nodes but I removed all but the main connected component such that it contains 8534 nodes.
   (b) The Labour Reply Leave Network has 643 leave early adopters and 1702 leave late adopters.
   (c) The Labour Reply Remain Network has 741 remain early adopters and 2237 remain late adopters.
   (d) The Labour Like Networks' start date (1 week before the referendum result announcement) was chosen so as to constrain it to 14975 nodes but I removed all but the main connected component such that it contains 14650 nodes.
   (e) The Labour Like Leave Network has 866 leave early adopters and 2636 leave late adopters.
   (f) The Labour Like Remain Network has 948 remain early adopters and 3937 remain late adopters.

3. **Liberal Democrats**: Depicted in Figure 3 as a turquoise dot in the lower part of the upper right quadrant Liberal Democrats is relatively centrist and was generally pro-Remain leading up to the referendum.[10]
   (a) The Liberal Democrats Reply Networks' start date (1 year before the referendum result announcement) was chosen so as to constrain it to 2357 nodes but I removed all but the main connected component such that it contains 2189 nodes. Even though this is an outlier, I didn't want to constrain the network to more than a year of interactions as that felt too long. So this one network is an outlier with about 1/4 as many nodes as the average of the others.
   (b) The Liberal Democrats Reply Leave Network has 120 leave early adopters and 274 leave late adopters.
   (c) The Liberal Democrats Reply Remain Network has 380 remain early adopters and 632 remain late adopters.
   (d) The Liberal Democrats Like Networks' start date (1 year before the referendum result announcement) was chosen so as to constrain it to 7855 nodes but I removed all but the main connected component such that it contains 7591 nodes.
   (e) The Liberal Democrats Like Leave Network has 407 leave early adopters and 920 leave late adopters.
   (f) The Liberal Democrats Like Remain Network has 1016 remain early adopters and 2090 remain late adopters.

4. **Conservatives**: Depicted in Figure 3 as a bright blue dot in the right part of the upper right quadrant Conservatives leans center-right and was mixed with a slight preference for pro-Remain leading up to the referendum.[11]

[9] Ibid.
[10] Ibid.
[11] Ibid.

(a) The Conservatives Reply Networks' start date (6 months before the referendum result announcement) was chosen so as to constrain it to 5106 nodes but I removed all but the main connected component such that it contains 4865 nodes.

(b) The Conservatives Reply Leave Network has 733 leave early adopters and 1358 leave late adopters.

(c) The Conservatives Reply Remain Network has 228 remain early adopters and 708 remain late adopters.

(d) The Conservatives Like Networks' start date (3 months before the referendum result announcement) was chosen so as to constrain it to 9763 nodes but I removed all but the main connected component such that it contains 9552 nodes.

(e) The Conservatives Like Leave Network has 1288 leave early adopters and 2762 leave late adopters.

(f) The Conservatives Like Remain Network has 340 remain early adopters and 1335 remain late adopters.

5. **UKIP**: Depicted in Figure 3 as a purple dot on the top right, UKIP leans right and was strongly pro-Leave leading up to the referendum.[12]

(a) The UKIP Reply Networks' start date (1 year before the referendum result announcement) was chosen so as to constrain it to 10519 nodes but I removed all but the main connected component such that it contains 9885 nodes.

(b) The UKIP Reply Leave Network has 3120 leave early adopters and 5727 leave late adopters.

(c) The UKIP Reply Remain Network has 233 remain early adopters and 911 remain late adopters.

(d) The UKIP Like Networks' start date (1 week before the referendum result announcement) was chosen so as to constrain it to 7808 nodes but I removed all but the main connected component such that it contains 7588 nodes.

(e) The UKIP Like Leave Network has 2186 leave early adopters and 4300 leave late adopters.

(f) The UKIP Like Remain Network has 158 remain early adopters and 740 remain late adopters.

6. **Britain First**: Not depicted in Figure 3 but founded by former BNP members (blue dot in the middle to the top) Britain First leans far right and was strongly pro-Leave leading up to the referendum.[13]

(a) The Britain First Reply Networks' start date (2 weeks before the referendum result announcement) was chosen so as to constrain it to 9897 nodes but I removed all but the main connected component such that it contains 7952 nodes.

(b) The Britain First Reply Leave Network has 569 leave early adopters and 1347 leave late adopters.

[12]Ibid.
[13]Ibid.

(c) The Britain First Reply Remain Network has 56 remain early adopters and 272 remain late adopters.

(d) The Britain First Like Networks' start date (1 week before the referendum result announcement) was chosen so as to constrain it to 55167 nodes but I removed all but the main connected component such that it contains 52652 nodes. Even though this is an outlier, I didn't want to constrain the network to less than a week of interactions as that felt too short. So this one network is an outlier with about 5 times as many nodes as the average of the others.

(e) The Britain First Like Leave Network has 3079 leave early adopters and 7571 leave late adopters.

(f) The Britain First Like Remain Network has 417 remain early adopters and 2430 remain late adopters.

Related Work

The question I am interested in investigating is whether the Leave campaign indeed showed any signs of campaigning more effectively than the Remain campaign. The above social interaction networks allow me to investigate, to some extent, how each side used social media to forward its campaign. Since the network data I have constructed contains early and late adopters, to investigate the question would be to try and measure what bearing the early adopters have on the late adopters, and whether this differs between leave early adopters and remain early adopters. The inspiration to model this experiment with a particular focus on early adopters comes from work by Coleman, Katz and Menzel (1977) that focuses on the importance of influential early adopters.[14]

So I set about trying to find a way to measure the influence these early adopters wield over their respective networks. To measure influence I first looked to Aral and Walker's 2012[15] paper on identifying influential and susceptible members of a social network. They model nodes by their qualitative properties and trying to identify which qualitative properties translate to high influence. Since I don't have access to enough qualitative properties of the nodes in my networks, I decided to look instead at work that tries to understand influence more in terms of nodes' network properties.

To this end I found work by Katz (1953) that attempts to measure a nodes influence by modeling various network properties of the node.[16] The properties of early adopters that I investigate in the ensuing sections borrow from this work.

Experiments

Testing for the presence of heterophily

In order for the early adopters to have any scope of exerting influence over other nodes in the network, we must first establish that there is some heterophily in the network, i.e. that there

[14] James Coleman, Elihu Katz, and Herbert Menzel. "The Diffusion of an Innovation among Physicians". In: *Social Networks* (1977), 107124. DOI: 10.1016/b978-0-12-442450-0.50014-6.

[15] Sinan Aral and Dylan Walker. "Identifying influential and susceptible members of social networks." In: *Science* 337 6092 (2012), pp. 337–41.

[16] Leo Katz. "A new status index derived from sociometric analysis". In: *Psychometrika* 18.1 (1953), 3943. DOI: 10.1007/bf02289026.

exist edges between early adopters and non-early adopters. If there are no such edges, i.e. if early adopters demonstrate perfect homophily, then there can be no scope for them to exert influence over other nodes.

To calculate this we use Newman's assortativity coefficient:[17]

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} = \frac{Trace(\mathbf{e}) - ||\mathbf{e}^2||}{1 - ||\mathbf{e}^2||}$$

Where e_{ij} is the fraction of edges in the network that connect a node of type i (here type i connotes early adopter node) to a node of type j (here type j connotes non-early adopter node). a_i and b_i are the fraction of each type of end of an edge that is attached to vertices of type i. Since our networks contain only undirected edges, $a_i = b_i$. The calculation gets reduced to finding the trace of $\mathbf{e}$, a matrix whose elements are e_{ij} and the sum of the squares of all elements of $\mathbf{e}$.

Using this measure, if $r = 1$ then there is perfect homophily in the network (i.e. no heterophily) and if this were to be the case, the early adopters would not ever interact with non-early adopters. if $r \neq 1$ there is some heterophily and we can proceed with other measures of influence of early adopter nodes.

I compute r for each of the networks and average it. The average comes out to 0.068 for the leave networks and 0.029 for the remain networks. Note these values are $\neq 1$ and close to 0, and $r = 0$ means there is no homophily in the network. This small value of r is encouraging, it means that early adopters interact generously with non-early adopters and it means we can proceed to see how we can measure their influence on other nodes.

Conceptualizing a test for the importance of network structure and the network properties of early adopters

Consider the graph shown in Figure 4a. Before the blue nodes (late adopters) became blue, they were green (non adopters). If the red nodes (early adopters) influenced some green nodes to become blue, then there must be some properties of these nodes that depend on red nodes that made them become blue and not stay green.

I phrase this experiment in terms of red, blue and green nodes instead of early, late and non adopters respectively because it is easier to think in terms of these colors and refer to Figure 4a as one is following the experiment.

To test this we can train a classifier as follows:

- Assume all non-red nodes are green at first.
- Extract features for all green nodes that express the relationship of the given green node to the red nodes around it.
- Then assign the given node a label of 0 if it stayed green and never became blue and 1 if it became blue.
- Train a classifier with these features and labels.

[17] M. E. J. Newman. "Mixing patterns in networks". In: *Physical Review E* 67.2 (2003). DOI: 10.1103/physreve.67.026126.

Now we can test this classifier on the green nodes of another network. If it correctly identifies which of the green nodes became blue there is a high likelihood that the red nodes had a bearing on the green nodes becoming blue (since the features all relate to red node properties).

If this is the case we can look further at which features in particular had the greatest importance to the classifier, and conclude that these features of red nodes encode their influence in networks like these.

Performing the test

First we have to choose what features to extract.

Feature Extraction The features extracted for each green node are inspired partly by Katz's work[18] and partly by my own intuition about what features might encode how a given green node might be influenced by close-by red nodes.

1. *Number of neighbors (aka node's degree centrality).* The general connectedness of a green node with the rest of the network is likely to impact how red nodes might influence it.

2. *Number of neighbors that are red.* The red nodes that immediately neighbor a green node are very likely to influence whether this node eventually becomes blue.

3. *Fraction of neighbors that are red.* If this node is only interacting with red nodes, it might be more likely to become blue.

4. *Number of neighbors at distance 2.* Similar reasoning to feature 1.

5. *Number of neighbors at distance 2 that are red.* Similar reasoning to feature 2.

6. *Fraction of neighbors at distance 2 that are red.* Similar reasoning to feature 3.

7. *Number of triangles with 2 early adopters this node is involved in.* Another way of expressing the quality that features 3 and 6 encode.

8. *the local clustering co-efficient of this node.* The local clustering coefficient of node i, C_i is calculated as follows[19]
$$C_i = \frac{2t_i}{k_i(k_i 1)}$$
Where t_i is the number of triangles around node i and k_i is the number of neighbors of node i. It measures the tendency of the node to interact with its neighbors' neighbors. More gregarious Facebook users will have a higher local clustering coefficient and they are also probably the users that early adopters will have easier access to and thus potentially greater influence over than nodes with lower clustering coefficients.

9. *the average of the local clustering coefficients of all neighbors that are red.* As discussed above, more gregarious Facebook users will have a higher local clustering coefficient. This holds true for early adopters too. If a node is surrounded by generally more early adopters that have high local clustering coefficients, the chance that it will have an influential interaction with them or their friends might be higher.

[18] Katz, "A new status index derived from sociometric analysis", op. cit.

[19] Jari Saramki et al. "Generalizations of the clustering coefficient to weighted complex networks". In: *Physical Review E* 75.2 (2007). DOI: `10.1103/physreve.75.027105`.

10. *the average of the local clustering coefficients of all neighbors at distance 2 that are red.* Similar intuition as the above feature.

11. *the average of the degree centralities of all neighbors that are red.* A nearby early adopter with a high degree centrality is more likely to be a more gregarious Facebook user, and therefore potentially a more influential early adopter.

12. *the average of the degree centralities of all neighbors at distance 2 that are red.* Similar intuition as the above feature.

13. *the eigenvector centrality of this node.* The eigenvector centrality for node v, x_v is calculated as follows[20]

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

    Where λ is a constant (it comes from one of the eigenvalues of the adjacency matrix for node v), $M(v)$ is the set of all neighbors of node v, $G := (V, E)$ is the graph and $a_{v,t} = 1$ if there is an edge between node v and node t. This equation can be rearranged to

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

    which results in the eigenvector centrality when λ is set to the greatest eigenvalue of v's adjacency matrix.

    Eigenvector centrality characterizes a nodes centrality in terms of its neighbors' centralities. So instead of measuring walks of length 1 from the node (as in degree centrality), it measures all walks from all neighbors of the node. Eigenvector centrality of nodes buried deep within well-connected components of graphs are high. In the Facebook case a node with a high eigenvector centrality interacts with many very well connected Facebook users, and therefore might have a higher chance of interacting with and potentially being influenced by a red node in its vicinity.

14. *the average of the eigenvector centralities of all neighbors that are red.* As discussed above, Facebook users with a higher eigenvector centrality generally have access to a large number of nodes. This holds true for early adopters too. If a node is surrounded by generally more early adopters that have high eigenvector centralities, the chance that it will have an influential interaction with them or their friends might be higher.

15. *the average of the eigenvector centralities of all neighbors at distance 2 that are red.* Same intuition as above feature.

Next we have to choose which machine learning algorthim to use for classification.

Choosing the classifier

I extracted the above features for the UKIP Reply Leave network and partitioned the data into an 80-20 train-test split. I then experimented with Logistic Regression, Linear SVC, Random Forest Classification and Multi-Layer Perceptron Classification with various hyperparamters and found Random Forest Classification with the gini criterion as the best algorithm.

[20] *Eigenvector centrality.* 2018. URL: `https://en.wikipedia.org/wiki/Eigenvector_centrality#Using_the_adjacency_matrix_to_find_eigenvector_centrality`.

Training and testing the classifier
I partition the networks into three classes:

1. Left-leaning which consists of the Labour networks and the SNP networks
2. Center-leaning which consists of the Conservatives networks and the Liberal Democrat networks
3. Right-leaning which consists of the UKIP networks and the Britain First networks

I then perform the following 12 classification tasks:

1. Train on UKIP Reply Leave, Test on Britain First Reply Leave
2. Train on Britain First Reply Remain, Test on UKIP Reply Remain
3. Train on UKIP Like Remain, Test on Britain First Like Remain
4. Train on Britain First Like Leave, Test on UKIP Like Leave
5. Train on Conservatives Reply Leave, Test on Lib Dems Reply Leave
6. Train on Lib Dems Reply Remain, Test on Conservatives Reply Remain
7. Train on Conservatives Like Remain, Test on Lib Dems Like Remain
8. Train on Lib Dems Like Leave, Test on Conservatives Like Leave
9. Train on Labour Reply Leave, Test on SNP Reply Leave
10. Train on SNP Reply Remain, Test on Labour Reply Remain
11. Train on Labour Like Remain, Test on SNP Like Remain
12. Train on SNP Like Leave, Test on Labour Like Leave

Establishing a baseline
The baseline is to just flip a coin with

$$\text{bias} = \frac{\#\text{blue nodes in network}}{\#\text{red nodes in network}}$$

for each green node in the test network to determine whether it will be classified blue or green.

Table 1: Results for classification tasks

#	Training Set	Test Set	Accuracy	Recall	Precision
1	UKIP Reply Leave	Britain First Reply Leave	0.70	0.24	0.11
	Baseline		0.57	0.42	0.11
2	Britain First Reply Remain	UKIP Reply Remain	0.92	0.00	0.09
	Baseline		0.89	0.03	0.08
3	UKIP Like Remain	Britain First Like Remain	0.94	0.093	0.14
	Baseline		0.88	0.09	0.04
4	Britain First Like Leave	UKIP Like Leave	0.57	0.31	0.48
	Baseline		0.57	0.09	0.41
5	Conservatives Reply Leave	Lib Dems Reply Leave	0.71	0.24	0.08
	Baseline		0.78	0.15	0.08
6	Lib Dems Reply Remain	Conservatives Reply Remain	0.79	0.11	0.11
	Baseline		0.76	0.20	0.13
7	Conservatives Like Remain	Lib Dems Like Remain	0.79	0.07	0.27
	Baseline		0.73	0.11	0.19
8	Lib Dems Like Leave	Conservatives Like Leave	0.75	0.18	0.27
	Baseline		0.76	0.07	0.18
9	Labour Reply Leave	SNP Reply Leave	0.90	0.03	0.10
	Baseline		0.81	0.18	0.11
10	SNP Reply Remain	Labour Reply Remain	0.79	0.01	0.14
	Baseline		0.76	0.07	0.21
11	Labour Like Remain	SNP Like Remain	0.91	0.06	0.12
	Baseline		0.74	0.27	0.08
12	SNP Like Leave	Labour Like Leave	0.83	0.25	0.32
	Baseline		0.77	0.12	0.12

Test results

In Table 1 the baseline scores for each test are in the row following the row where that tests scores are reported. Accuracy, Precision and Recall scores are between 0 and 1 and calculated using the standard formulae.[21]

Though a T-test would be needed to determine the significance of any of these results it seems that generally where accuracy is good precision and recall are low. However in 4 cases the classifier beats the baseline on all three metrics. Though they do not beat the baseline by a convincing amount, we can still consult the classifier for which features were the most important in determining these results.

Consistently in these cases, eigenvector centrality and eigenvector centralities of nearby red nodes were reported by the Random Forest Classifier as the most important features in determining its classification.

Though the experiment hasn't shown us convincingly that these red node features are good determiners of how influential a node is, they do still point to eigenvector centrality being an important network property for influential nodes.

With this, we can look at the eigenvector centralities of the early adopters in various contexts:

[21] *API Reference*. URL: `http://scikit-learn.org/stable/modules/classes.html`.

1. Average eigenvector centrality of leave early adopters in all networks: 0.0061
2. Average eigenvector centrality of leave early adopters in left-leaning networks: 0.0070
3. Average eigenvector centrality of leave early adopters in center-leaning networks: 0.0068
4. Average eigenvector centrality of leave early adopters in right-leaning networks: 0.0044
5. Average eigenvector centrality of remain early adopters in all networks: 0.0051
6. Average eigenvector centrality of remain early adopters in left-leaning networks: 0.0033
7. Average eigenvector centrality of remain early adopters in center-leaning networks: 0.0080
8. Average eigenvector centrality of remain early adopters in right-leaning networks: 0.0037

One interesting result from the above is that leave early adopters have on average twice the eigenvector centrality in left-leaning networks than remain early adopters do. This might suggest that Leave campaigners went out of their way to campaign on groups that didn't favor Leave to begin with. Diving further into this question could be an interesting area of future analysis.

Evaluation and Conclusion

Though we arrived at the importance of eigenvector centrality through a relatively on the whole unconvincing experiment, we were still able to pinpoint some differences in how Leave and Remain campaigners conducted themselves on social media.

Here are some drawbacks of the experiments:

1. Even in cases where the classification task does better than baseline, suggesting that the eigenvector centrality of the red nodes surrounding a green node might have a bearing on whether it turns blue, we still haven't shown causality, merely correlation.
2. The time period that the Like and Reply networks exist over is slightly problematic. In some cases the interactions are modeled over a year which implies an interaction between an early adopter (up to 6 months before they became an early adopter) and a late adopter (up to a year before they became a late adopter) is worth looking at. This may or may not be the case, but deserves greater discussion.
3. A major flaw in this experiment is the complete lack of control of spontaneous late adoption, i.e. late adoption that has nothing to do with a person's interactions with early adopters on any network in any time period.
4. Many of the extracted features are not independent. This is not a huge issue but worth mentioning.
5. Finally it is not very scientific to use some data to find a property that is correlated with influence and then measure the distribution of that property in that same data. It would have been better to show this property on some data and measure the distribution of the property on other data.

Because of these drawbacks, we cannot really extract a powerful qualitative conclusion from this result but it is still noteworthy that we were able to find a property (eigenvector centrality) that under certain circumstances has correlation with an early adopter's influence in the network and that this property was differently distributed among leave and remain early adopters.

Future Work

Though this experiment was unconvincing, it does suggest that there might be some network properties of early adopters of either the Leave or Remain point of view that had an influence on other Facebook users (and voters) in the UK. A more substantial experiment would be to add a temporal aspect to the network such that adopters are not classed as "early" or "late" as in the above work but are identified as a function of when they became adopters. If this kind of adoption can be shown to follow a submodular function one could test whether adoption of an opinion on Brexit follows the Independent Cascade Model.

Acknowledgements

I would like to thank Thomas Davidson for sharing this data with me, Soroush Alamdari and Rediet Abebe for helping me with some technical aspects of the project and Professor Jon Kleinberg for encouraging me to search for a project in this realm, that I am deeply interested in.

References

API Reference. URL: http://scikit-learn.org/stable/modules/classes.html.

Aral, Sinan and Dylan Walker. "Identifying influential and susceptible members of social networks." In: *Science* 337 6092 (2012), pp. 337–41.

Behr, Rafael. *How remain failed: the inside story of a doomed campaign — Rafael Behr*. 2016. URL: https://www.theguardian.com/politics/2016/jul/05/how-remain-failed-inside-story-doomed-campaign.

Brexit. 2018. URL: https://en.wikipedia.org/wiki/Brexit#Opinion_polls_19772015.

Brexit. 2018. URL: https://en.wikipedia.org/wiki/Brexit#Negotiations_for_EU_reform.

Coleman, James, Elihu Katz, and Herbert Menzel. "The Diffusion of an Innovation among Physicians". In: *Social Networks* (1977), 107124. DOI: 10.1016/b978-0-12-442450-0.50014-6.

Compass, The Political. *UK Parties 2015 General Election*. URL: https://www.politicalcompass.org/uk2015.

Eigenvector centrality. 2018. URL: https://en.wikipedia.org/wiki/Eigenvector_centrality#Using_the_adjacency_matrix_to_find_eigenvector_centrality.

Hansard, Westminster, and House of Commons. *House of Commons Hansard Debates for 22 Feb 2016 (pt 0001)*. URL: https://publications.parliament.uk/pa/cm201516/cmhansrd/cm160222/debtext/160222-0001.htm#16022210000001.

Initiative, ideas@global-initiative.com Global. *UK press coverage of the EU Referendum*. URL: http://reutersinstitute.politics.ox.ac.uk/our-research/uk-press-coverage-eu-referendum.

Katz, Leo. "A new status index derived from sociometric analysis". In: *Psychometrika* 18.1 (1953), 3943. DOI: 10.1007/bf02289026.

Newman, M. E. J. “Mixing patterns in networks”. In: *Physical Review E* 67.2 (2003). DOI: 10.1103/physreve.67.026126.
Saramki, Jari et al. “Generalizations of the clustering coefficient to weighted complex networks”. In: *Physical Review E* 75.2 (2007). DOI: 10.1103/physreve.75.027105.
Submission, Guest. *Where the UK Political Parties Stand on Brexit SJS*. 2017. URL: http://www.socialjusticesolutions.org/2017/06/05/uk-political-parties-stand-brexit/.
The Labour Party. URL: https://www.facebook.com/labourparty.