PNAS

Supporting Information for

Temporal Dynamics of Coordinated Online Behavior: Stability, Archetypes, and Influence

Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti,Preslav Nakov, Giovanni Da San Martino, Stefano Cresci

Corresponding author: Stefano Cresci.
E-mail: stefano.cresci@iit.cnr.it

This PDF file includes:

Supporting Information Text

Data

UK 2019 General Election. The dataset comprises tweets collected between November 12 and December 12, 2019 – that is, the month before the UK 2019 election day. Specifically, all tweets containing at least one of the hashtags shown in Table S1 were collected. In addition, all tweets produced by the two main parties and their leaders, as well as all interactions (i.e., retweets and replies) they received, were also collected. Table S1 reports the number of tweets collected for each hashtag and account.

USA 2020 Presidential Election. The tweets included in this dataset were collected between October 4 to November 3, 2020, the month before the election. In particular, we gathered all tweets that included at least one of the hashtags listed in Table S2. Additionally, we gathered all tweets originating from the two major political parties and their respective leaders, as well as all interactions such as retweets and replies directed towards them. Table S2 provides a breakdown of the tweet counts associated with each hashtag and account.

Honduras 2019 Information Operation. We leveraged data from Twitter's Moderation Research Consortium (TMRC).* The TMRC has been sharing information about accounts banned for being involved in state-sponsored Information Operations (IOs) since 2018, an effort that has led to the disclosure of numerous datasets containing information on thousands of banned accounts, covering various languages and regions around the globe. These datasets have been widely regarded as authoritative ground-truth of malicious users involved in misinformation campaigns (1–3). We focused on a specific IO allegedly promoted by the government of Honduras between 2019 and 2020. According to Twitter, this relatively unexplored operation involved thousands of accounts working to falsely increase the visibility and popularity of Honduras President Juan Orlando Hern'andez on Twitter. The inorganic campaign involved 3,104 inauthentic accounts who shared numerous Spanish and English tweets, and were found to be retweeting the President's tweets in large numbers from the same IP address range in Honduras, creating a false impression of grassroots support. In particular, we selected the activities of malicious accounts performed during one month prior their banning, from 11 November, 2019 to 11 December, 2019. Accounts that did not tweet in such a time span were not included in our datasets. To effectively analyze IOs, it is essential to compare malicious users with genuine ones (4), an approach that has been effectively used in recent studies (1, 4, 5). Here, we followed the same approach of these state-of-the-art works to collect a comparable set of genuine accounts for the Honduras 2019 IO, together with their activity, so as to "contrast" or "rebalance" the set of inauthentic accounts provided by Twitter. The strategy to collect a set of genuine accounts that is comparable to the inauthentic ones, yet unrelated to the IO, is based on collecting non-banned accounts that discussed the same topics of the IO at around the same time. We therefore extracted all the hashtags used by the inauthentic accounts, we ranked them based on their frequency, and we used the most frequent ones to query Twitter's Academic Search API. In Table S3, we list the hashtags and report the tweet counts associated with each hashtag in the dataset. This process allowed us to collect all non-banned (i.e., genuine) accounts and their tweets shared on the same dates of the IO and with the same hashtags as the inauthentic accounts. The final Honduras 2019 Information Operation dataset is obtained by merging data about the inauthentic accounts with that about the genuine ones. The final dataset includes 251,191 tweets shared by 75,845 distinct users. As it is typically the case when studying online harms and manipulations (6), the dataset is unbalanced. For example, out of all the superspreaders that were active in the last month before the ban, 35.4% were inauthentic accounts and 64.6% were genuine ones, which makes the task of detecting the inauthentic accounts, or the IO, particularly challenging. This imbalance is similar to that obtained in related works for other IOs (4, 5).

Results

Dynamic UK 2019 and USA 2020 communities. We characterize the top-10 dynamic communities by analyzing the temporal trends of the hashtags used by their members, which allows identifying the key issues and the predominant themes of each CC. Figures S1a and S1b show an excerpt of this analysis by highlighting the most frequent hashtag used by each CC at each point in time. In addition, Table S4 shows their size in terms of users.

RQ2: Temporal dynamics of user behavior. The analysis of Figure 7 only considers the number of memberships and shifts between CCs. However, not all shifts are the same, as moving between two opposite communities (e.g., at the extremes of the political spectrum) entails a much bigger change – a farther leap – than moving between two similar ones. To account for this facet we assign a weight $w_{k,j}$ to all shifts $s_{k \to j}$ between any origin community C_k and any destination community C_j, based on the (dis)similarity between C_k and C_j. We compute the similarity between two CCs as the cosine similarity of the TF weighted vectors of the hashtags used by the communities. Then, we weight shifts proportionally to the dissimilarity of the involved CCs: $w_{k,j} = 1 - \text{sim}(k, j)$. Finally, we analyze the patterns of user shifts between CCs. Figure S2 shows the similarity matrices obtained by computing the pairwise similarities between each CC. The figure highlights three clusters of similar communities for UK 2019 and one big cluster for USA 2020, each marked with red borders. A first cluster of highly similar communities comprises `LAB1`, `LAB2`, and `RCH`. A second cluster extends the previous one with `B60` and `TVT`. Finally, `CON`, `BRX`, `SNPO`, and `ASE` form the last UK 2019 cluster, while `SNP` stands as largely dissimilar to all other communities. The big cluster identified for USA 2020 includes all CCs, except for `DEM`, `BFR`, and `IRN`. This result is well aligned with the placement of

*https://transparency.twitter.com/en/reports/information-operations.html

the communities within the US political spectrum, as shown in Figure 2b. Overall, also the UK 2019 clusters closely resemble the UK political landscape in 2019 (7) and the position of the CCs in the political spectrum of Figure 2a.

Figure S3 represents all weighted user shifts between CCs as a weighted directed radial node-link diagram with edge bundling (i.e., a chord diagram). In figure, only *net shifts* between CCs are shown. Specifically, an edge $C_k \rightarrow C_j$ exists only if there is a positive net user flow $F_{k \rightarrow j}$ from C_k to C_j: $F_{k \rightarrow j} = \sum s_{k \rightarrow j} - \sum s_{j \rightarrow k} > 0$. Then, edge thickness is proportional to $w_{k,j} \times F_{k \rightarrow j}$. Figure S3 provides interesting insights into the patterns of user shifts between CCs. In particular, it shows that the majority of shifts in both datasets occurred between similar communities, such as between the communities on the same side of the political spectrum. This result is particularly relevant for the UK 2019 dataset, which feature CCs spanning the whole political spectrum.Nonetheless, Figure S3 also surfaces a net flow of users that moved *across* the political spectrum, such as those moving from `CON` to `RCH` in UK 2019 and those moving from `DEM` to `REP` in USA 2020. This highlights that major ideological shifts do occur, albeit infrequently with respect to all other shifts. Another interesting observation derived from Figure S3a is that the vast majority of shifts for UK 2019 occurred towards the left of the political spectrum. This means that, overall, the users involved in the online electoral debate ideologically moved towards the left as the debate unfolded. We formalize this observation by leveraging the polarity score p associated to each UK 2019 CC, with $-1 \leq p < 0$ for left-leaning communities and $0 < p \leq +1$ for right-leaning ones, shown in Figure 2a. In detail, we compute the difference in polarity entailed by each shift $s_{k \rightarrow j}$ as: $\Delta p_{k \rightarrow j} = p_j - p_k$. Thus, a negative Δp represents a change in polarization towards the left. Then, we derive the mean and total changes in political polarization resulting from all the shifts shown in Figure S3a. On average, each user shift contributed to a change in polarization $\Delta p = -0.37$, which led to a total change $= -12.91$. These results testify the strong pull exerted by the left-leaning CCs during the UK 2019 online debate.

RQ3: Archetypes and drivers of user behavior. Here we investigate the possible drivers for the different temporal behaviors of the three archetypes of users that we identified.

Archetype 1: Stationary. We model each stationary user with the TF ordered ranking of the hashtags they used and each CC with the TF ordered ranking of the hashtags used by its members. We recall that since our study is focused on *superspreaders* – users characterized by very large numbers of retweets (8) – excluding retweets from this analysis would result in dropping the vast majority of our dataset. For this reason, we did not exclude retweets in this and in subsequent hashtags-based analyses. We then compute the similarity between stationary users and all CCs as the Rank-Biased Overlap (RBO) of the respective rankings (9). RBO is a probabilistic measure of similarity based on the overlap between two rankings, which can handle tied ranks and rankings of different lengths, as in our case. Furthermore, it has a bias component that favors (i.e., weights more) overlaps between top items in the rankings (10). Once computed the RBO scores, we compare the CC to which each stationary user belongs with that to which it is mostly similar, as shown in Figure S4. Results show that 94% of all UK 2019 stationary users and 97% of all USA 2020 stationary users are mostly similar to the CC to which they belong, as highlighted in figure by the great prevalence of users along the main diagonal.

Archetype 2: Influenced. Topic-based similarities are measured with the RBO scores obtained by comparing the TF-IDF embeddings of hashtag rankings of influenced users to those of C_k and C_j. In addition, we operationalize the distance in the multiplex temporal network between influenced users and their origin and destination communities by computing their closeness centrality with respect to C_k and C_j. Before carrying out both analyses, we temporally align all influenced users so that their shifts all occur at time t_i. Then, we evaluate topic-based similarities and closeness centralities before and after t_i. The temporal trends shown in Figure S5 reveal that influenced users indeed exhibited an increasing similarity – be it topic- or network-based – to their destination community as time went by. Meanwhile, they also became increasingly dissimilar to their origin community. This finding holds under all the viewpoints considered in our analyses. Specifically, this phenomenon occurred both within the UK 2019 and the USA 2020 dataset. Then, the behavior is observable both before the time t_i of the shift as well as in its aftermath, denoting a clear behavioral trajectory. Furthermore, we note that the opposite similarity trends observed for the origin *vs.* destination communities cross just before t_i, which might explain the subsequent shift. The above results are consistent for both topic-based similarity and closeness centrality. For topic-based similarity, the result holds both in terms of TF-IDF embeddings (Figures S5a and S5c) and hashtag occurrences (Figure S6a and Figure S6b). However, closeness centrality provides a much stronger signal than topic-based similarity, as reflected by the marked differences between the trends shown in Figures S5b and S5d with respect to those in Figures S5a and S5c, which also results in the differences in Figures S5b and S5d being statistically significant at each considered time step, according to a Kruskal-Wallis test.

Archetype 3: Volatile. To further investigate the behavior of volatile users, we analyze the distance spanned by their shifts. Similarly to our analysis in RQ2, we measure the distance spanned by a shift $s_{k \rightarrow j}$ based on the similarity between the communities C_k and C_j, so that shifts between similar CCs span a shorter distance than those between dissimilar ones. We thus operationalize shift distances with the weights $w_{k,j}$ that we computed based on the similarities between the CCs reported in Figure S2. For the UK 2019 dataset, Figure S7a shows the distribution of the shift distances for volatile users, influenced users, and for those users that do not match the definition of any archetype. As shown, the vast majority of shifts by volatile users are relatively short. Qualitatively, the distribution of shift distances for volatile users resembles that of the other generic users, suggesting that a simple analysis of shift distances is insufficient to explain the difference between volatile and others. Nonetheless, a Kruskal-Wallis test supports the difference between volatile and influenced ($p < 0.05$) or other users ($p < 0.01$). Instead, much more evident is the difference between the distribution of shift distances for influenced users with respect to all other types of users, with the former being a bimodal distribution. This last observation is particularly interesting for the

study of online influence as it suggests that a subset of users permanently joined a community that was politically far from their origin community. Figure S7b shows the distribution of the shift distances for three groups of users in the USA 2020 dataset. Similar to the UK 2019 one, most of the shift distances for volatile users are relatively short. Instead, the observed shift distances for influenced users are notably higher than those in UK 2019. This could be attributed to the more extreme political polarization in the country. In fact, the broader and more extreme political polarization in the USA, as indicated by the greater spread of community clusters across the political spectrum (Figure 1 of the main manuscript), could contribute to larger ideological shifts among influenced users. This wider ideological range suggests that the political events or discourses in the USA during 2020 may have been particularly divisive, leading to more pronounced shifts in online behavior among users.

Dynamic communities and real-world events. Our previous analyses on both the UK 2019 and the USA 2020 scenarios highlighted the presence of multiple unstable CCs who experienced major fluctuations in their size and membership. Here we conclude our analyses with a semi-automatic investigation aimed at providing real-world context for a subset of notable cases. We first identify the days in which the CCs experienced a marked change in size. To do so we compute the fraction of users who joined and left each CC at each time step. We then identify check-worthy time steps as those where the size of a CC differs by more than five standard deviations with respect to the average size of the same CC in the two previous time steps. This simple anomaly detection approach yields pairs of CCs–time steps, which we subsequently manually investigate. An excerpt of our results are presented in Figure S8 and briefly discussed as follows.

On November 28, 2019 the `LAB1` community experienced a sudden reduction to the number of joining users. As shown in Figure S8, this behavior deviated significantly from both its previous and subsequent trends of joining users. Interestingly, we found that at that time a prominent Labour candidate was removed from office amidst allegations of anti-Semitism (inset *a*) (11). The event also impacted the `LAB2` community, albeit to a lower extent. Nonetheless, it temporally corresponded to the time step from which `LAB2` began shrinking in size, as also shown in Figure 4a. The progressive shrinking of `LAB2` went on until December 07, 2019 when the community lost a significant share of its members following the BBC television debate between the two leaders Jeremy Corbyn and Boris Johnson (inset *b*) (12). Similarly, between December 02 and 03, 2019 the `SNP` community experienced a marked reduction in its user base. Notably, the change quickly followed a major television debate featuring Scottish party leaders arguing about the possibility of a second independence referendum (inset *c*) (13). Finally, `CON` exhibited a spiky increment of $\sim 10\%$ of its size around November 24, 2019, which precisely coincided with the official unveiling of the Conservative manifesto (inset *d*) (14). In the subsequent weeks the situation quickly changed, and the `CON` community progressively reduced its size, as also shown in Figure 4a. A significant decline occurred on December 08, 2019, in conjunction with the final television debate (inset *e*) (15). Interestingly, Figure S8 shows a similar pattern for the `SNPO` community, which shared the same side of the political spectrum and multiple themes with `CON`, such as their support for Brexit.

In summary, this analysis revealed that the temporal dynamics of coordinated communities often align with significant real-world events. This offers valuable insights into the motivations that drive user engagement with the different communities, including the factors ultimately influencing their shifts between two community. In addition to providing context for some of our results, this manual validation also supports the reliability of our findings.

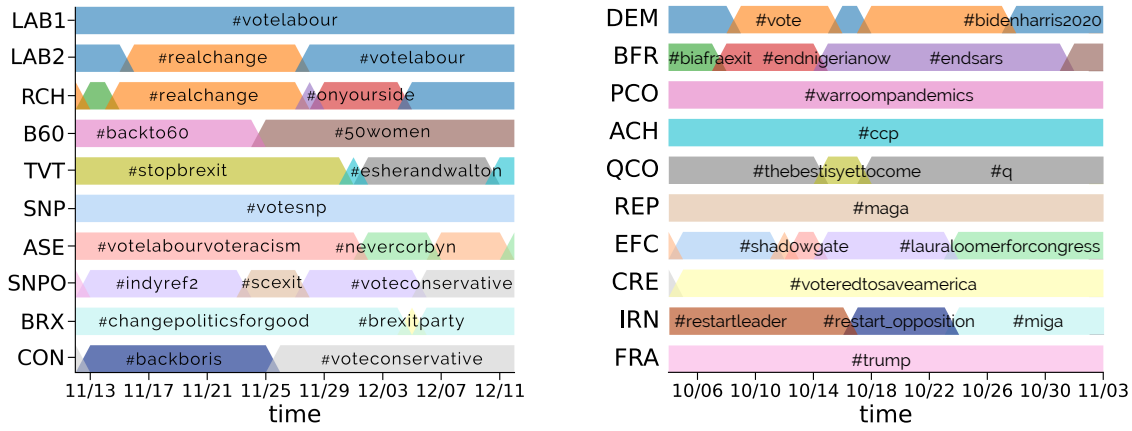
Validation. We validate the efficacy of our methodology on the Honduras 2019 dateset. As mentioned, the dataset is composed of the inauthentic accounts involved in the Honduras IO, and of a comparable set of genuine accounts. This dataset represents a favorable and reliable benchmark for multiple reasons. First, it serves for validating the efficacy of our method at capturing the behavioral patterns of different accounts (i.e., inauthentic and genuine forms of coordination). Second, it informs the optimization of our method's parameters. Third, it illustrates the practical usefulness of the dynamic analysis of coordinated behavior for a relevant computational social science task.

To validate our method we perform a grid search analysis where we re-run the method with different combinations of the three main parameters: the time window length d, the time window step δ, and the resolution γ of the community detection algorithm. At each run – that is, for each configuration of parameters – we quantify the capacity of our framework at capturing meaningful behavioral patterns. In detail, we measure the goodness of the separation between the inauthentic and the genuine accounts, in terms of the well-known F1 score (16), for each layer of the multiplex temporal network. Then, we aggregate the results across all layers, so as to obtain an average result for each multiplex temporal network. Detailed information on how we computed the F1 scores for each community, layer, and for the multiple networks are given in the *Materials and Methods* section of the main manuscript.

Figure 9a of the main manuscript reports the results of this analysis. As shown, the lowest F1 $= 0.75$ is obtained when using time windows of length $d = 2$ days, with a step $\delta = 1$ day. This result likely corresponds to a situation where the time windows are simply too small to collect meaningful statistics at each time step and, therefore, obtain meaningful high-level analyses. Apart from this outlier configuration, the rest of the grid search results reveal an interesting trend where the largest F1 scores are obtained for relatively small values of the two parameters. In fact, the best F1 $= 0.91$ is obtained with time window length $d = 3$ days and step $\delta = 2$ days. F1 scores gradually decrease when increasing the length and step of the time windows, and markedly lower F1 scores are obtained for time window length $d \geq 11$ days and step $\delta \geq 8$ days. This result tells us that very long time windows aggregate too much information, erasing the different behavioral patterns in the data. To this end, we remark that the longer the time window, the more the dynamic analysis collapses (i.e., becomes similar) to a static one, negating the advantages of the former. These results are also in line with those of many related studies, which suggest adopting relatively short time windows for highlighting inauthentic and inorganic behaviors, which are often characterized by

tightly synchronized interactions (17, 18). Instead, slightly longer time windows should be preferred when studying organic and genuine behaviors, which are typically loosely synchronized (19).

Figure S9 shows the sensitivity of our method to the resolution parameter γ of the Leiden community detection algorithm. For each resolution value that we tested in our grid search, we summarized the distribution of the corresponding F1 values resulting from all tested combinations of time window length and step. In particular, each point in figure corresponds to the mean F1 of the distribution and the red-colored error bars indicate the standard deviation. As shown, the resolution has a considerable impact on the results of our framework. Large resolution values ($\gamma \geq 0.5$) result in the formation of many small communities, which negatively impacts the resulting F1 scores. In fact, having many small clusters thus reduces the *Recall* metric, since the inauthentic accounts are split across multiple communities and no single community contains a large share of all inauthentic accounts. Instead, small resolution values ($\gamma \leq 0.2$) result in the formation of a few large communities. Here the challenge lies in maximizing the *Precision* metric – that is, having large yet homogeneous communities that either contain only inauthentic or genuine accounts. The large F1 scores measured in our grid search for $\gamma \leq 0.2$ indicate that our framework produced largely homogeneous communities, as also shown in the example presented in Figure 9b of the main manuscript. Differently from the resolution, the small error bars in Figure S9 indicate that, in proportion, the time window length and step parameters of our method have a relatively small impact on the outputs of the framework. Conversely, our results are more sensible to the resolution parameter γ of the underlying community detection algorithm. It is important to clarify that while this parameter is not introduced by our method, it is a well-known factor common to many community detection algorithms. The sensitivity of certain community detection algorithms to the choice of the resolution parameter is a widely recognized phenomenon in the literature (20, 21). Indeed, there is no single optimal value for this parameter, as its configuration depends on various factors such as the characteristics of the analyzed network and the objectives of the analysis. Consequently, the selection of appropriate values for the resolution parameter is an aspect of the analysis process that is left to the discretion of the analyst. This uncertainty has prompted studies aimed at providing guidance to analysts in choosing appropriate values for this parameter (22).

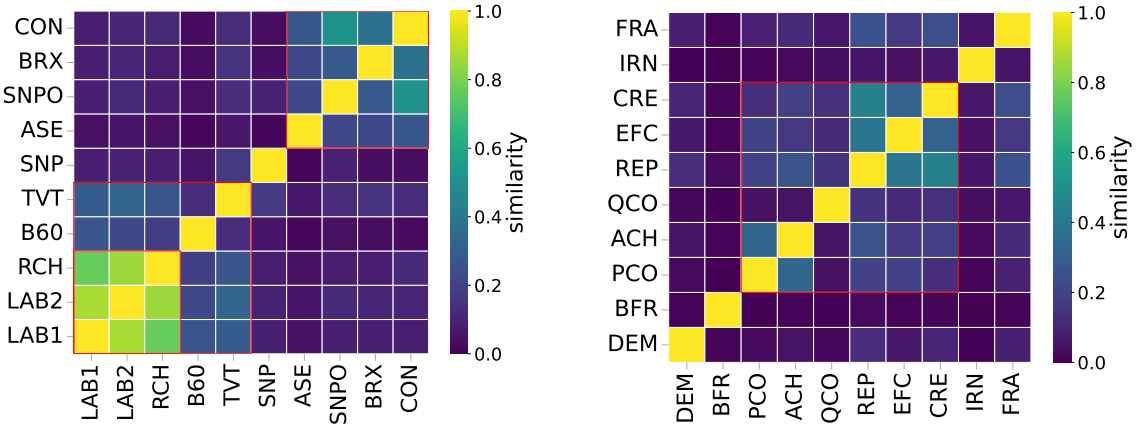


(a) UK 2019.

(b) USA 2020.

Fig. S1. Main theme discussed by each CC through time. The dynamic analysis allows differentiating CCs that appear as overall similar, but that feature different temporal behaviors (e.g., `LAB1`, `LAB2`, and `RCH` for the UK 2019 dataset).

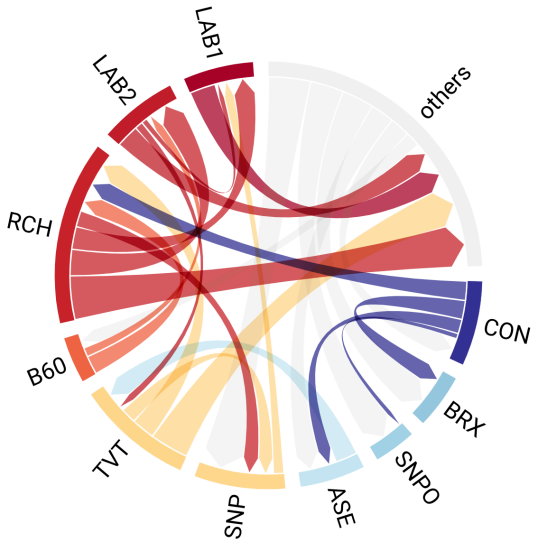
Fig. S2. Topic-based similarity between CCs. Three clusters of similar CCs, two left-leaning and one right-leaning, are identified for UK 2019. One big cluster of right-leaning communities is identified for USA 2020. Clusters are highlighted with red borders.



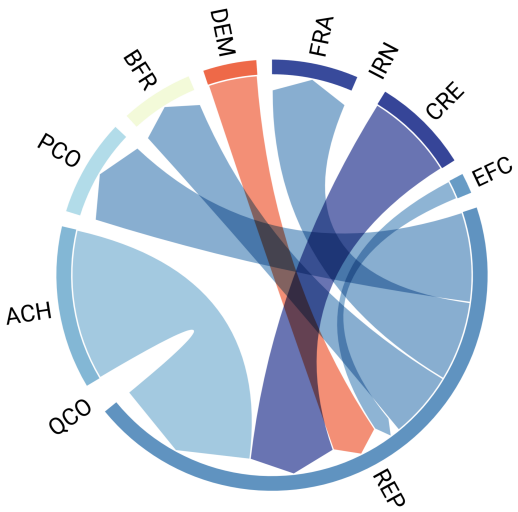
(a) UK 2019.

(b) USA 2020.

Fig. S3. Net weighted user shifts between CCs. Edges are colored based on the political leaning of the origin community. The majority of shifts occurred between politically-aligned communities, although some users also moved across the political spectrum (e.g., from `CON` to `RCH` during UK 2019 and from `DEM` to `REP` during USA 2020).



(a) UK 2019.



(b) USA 2020.

Table S1. UK 2019 dataset statistics: tweets collected for each hashtag (#) and account (@), and their political leaning.

leaning	hashtag/account	tweets
labour	@jeremycorbyn	2,422,162
	@UKLabour	668,264
	#VoteLabour	1,051,204
	#VoteLabour2019	272,975
	#ForTheMany	42,486
	#ForTheManyNotTheFew	42,403
	#ChangeIsComing	15,903
	#RealChange	314,279
neutral	#GE2019	3,006,685
	#GeneralElection19	335,202
	#GeneralElection2019	855,162
conservative	@BorisJohnson	973,546
	@Conservatives	592,068
	#VoteConservative	338,689
	#VoteConservative2019	35,272
	#BackBoris	186,794
	#GetBrexitDone	204,368

Table S2. USA 2020 dataset statistics: tweets collected for each hashtag (#) and account (@), and their political leaning.

leaning	hashtag/account	tweets
democrats	@JoeBiden	12,499,125
	@DrBiden	288,255
	@KamalaHarris	2,971,072
	@SenKamalaHarris	260,468
	@TheDemocrats	324,688
	#JoeBiden	743,605
	#biden	1,106,452
	#VoteBlue	153,856
	#VoteBlueToSaveAmerica	157,107
	#Biden2020	141,351
	#BidenHarris2020	979,039
	#joebiden2020	19,789
	#NeverTrump	32,740
neutral	#usa2020elections	4
	#usa2020	4,263
	#Election2020	2,560,391
	#ElectionDay	584,095
	#Debates2020	703,307
	#Vote	2,694,442
	#VoteEarly	349,301
	#Ivoted	45,957
republicans	@POTUS	3,670,176
	@realDonaldTrump	43,131,119
	@Mike_Pence	1,022,082
	@VP	853,809
	@MELANIATRUMP	45,442
	@FLOTUS	1,260,890
	@GOP	1,727,397
	#DonaldTrump	118,199
	#trump	1,976,369
	#VoteRedToSaveAmerica	380,114
	#VoteRed	245,643
	#trump2020	3,032,268
	#trumppence2020	512,048
	#donaldtrump2020	3,808
	#MAGA	4,863,113
	#KAG	859,569
	#NeverBiden	44,045
	#WakeUpAmerica	122,378

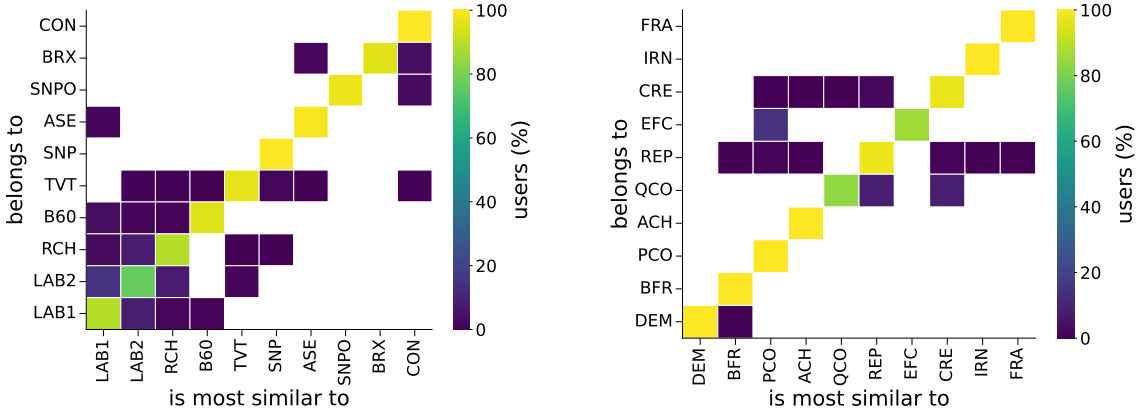
Table S3. Honduras 2019 dataset statistics: tweet counts associated with the most frequent hashtags in the dataset

hashtag	tweets
#AlivioDeDeuda	1,024
#NavidadCatracha	2,407
#EEUU	151,105
#VidaMejor	1,274
#ParqueVidaMejor	590
#VivaHonduras	22
#FiestasPatrias2019	17
#FeriadoMor‘azanico	6
#HondurasEnLaONU	3

Table S4. Size of the dynamic coordinated communities (CCs). CCs are ordered by political polarization, from left to right. Core size represents the number of users that belong to the same CC for all the time.

UK 2019 communities	*core size*	*size at* t_0	*top hashtag*
LAB1	107	253	#votelabour
LAB2	76	152	#votelabour
RCH	720	1,537	#realchange
B60	96	120	#backto60
TVT	933	1,376	#stopbrexit
SNP	327	380	#votesnp
ASE	66	100	#votelabourvoteracism
SNPO	67	87	#snpout
BRX	63	102	#changepoliticsforgood
CON	184	407	#voteconservative
USA 2020 communities			
DEM	139	1,219	#bidenharris2020
BFR	27	119	#biafraexit
PCO	23	181	#warroompandemic
QCO	12	52	#q
ACH	14	98	#ccp
EFC	7	67	#justsaying
REP	1,810	4,595	#maga
FRA	10	182	#trump
CRE	918	1,418	#trump2020
IRN	28	163	#restartleader

Fig. S4. Topic-based similarity between stationary users and all communities. The heatmap maps stationary users based on the CC to which they belong (*y* axis) and the CC to which they are most similar to (*x* axis). The prevalence of users on the main diagonal shows that the vast majority of stationary users is most similar to the CC to which they already belong, explaining why they never shift.



(a) UK 2019.

(b) USA 2020.

Fig. S5. UK 2019: *a*, *b*. USA 2020: *c*, *d*. Temporal trends of topic-based similarity (*a*, *c*) and closeness centrality (*b*, *d*) between influenced users and their origin and destination CC, around the time t_i of the shift. Both trends anticipate the shift, although only differences in closeness centrality (*b*, *d*) are statistically significant for each time step. Statistical significance levels: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.

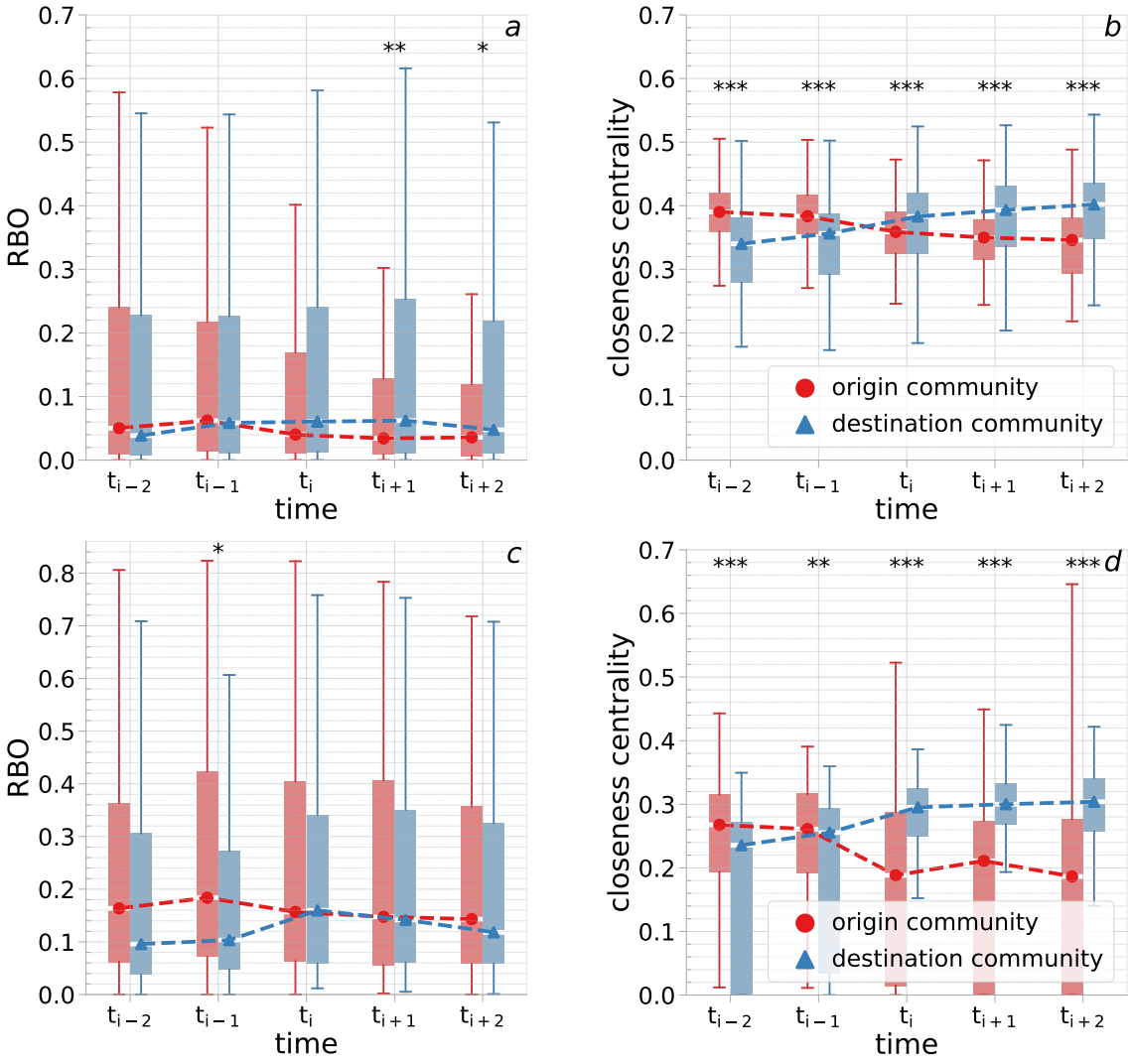
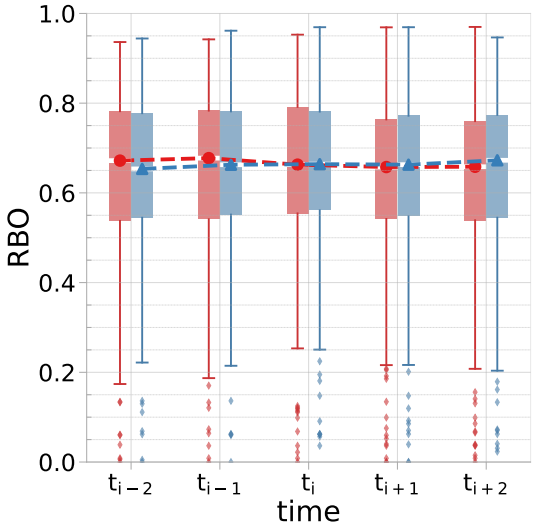
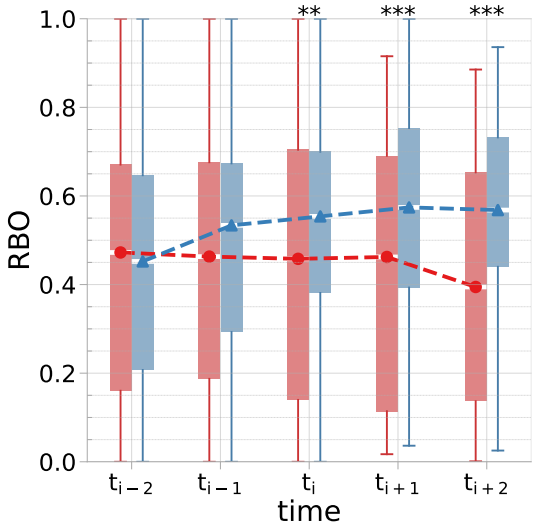


Fig. S6. Temporal trends of topic-based similarity using RBO of hashtag occurrences between influenced users and their origin and destination CC, around the time t of the shift.

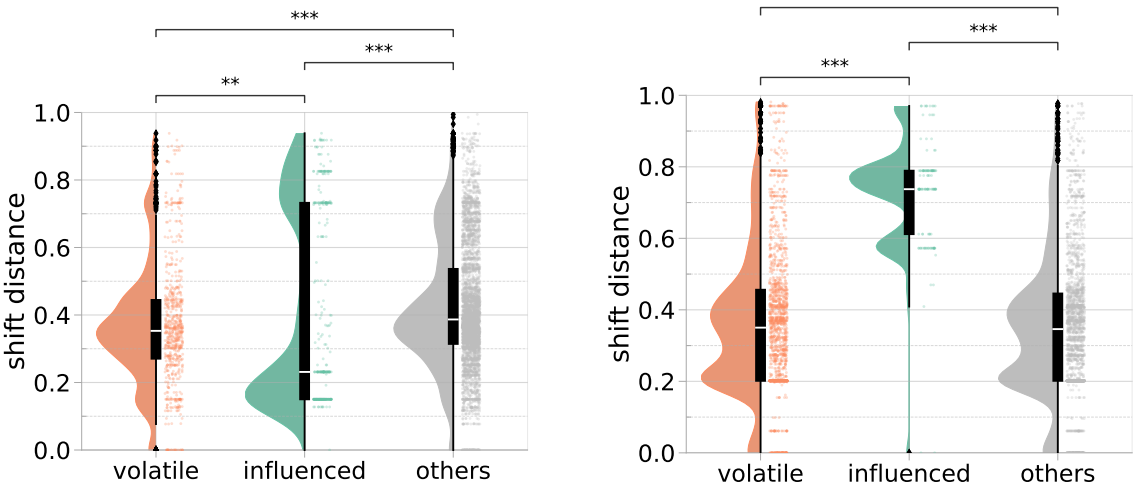


(a) UK 2019.



(b) USA 2020.

Fig. S7. Distribution of shift distances for volatile, influenced, and other users. While all distributions are significantly different, influenced users stand out as their bimodal distribution suggests that some users joined politically far communities. Statistical significance levels: ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$.



(a) UK 2019.

(b) USA 2020.

Fig. S8. Fraction of users who joined (green-colored) and left (red-colored) a subset of UK 2019 communities at each time step. Time steps when the communities experienced marked changes in size are highlighted in yellow. We manually investigated each of those events, providing real-world context for the changes as shown in the insets.

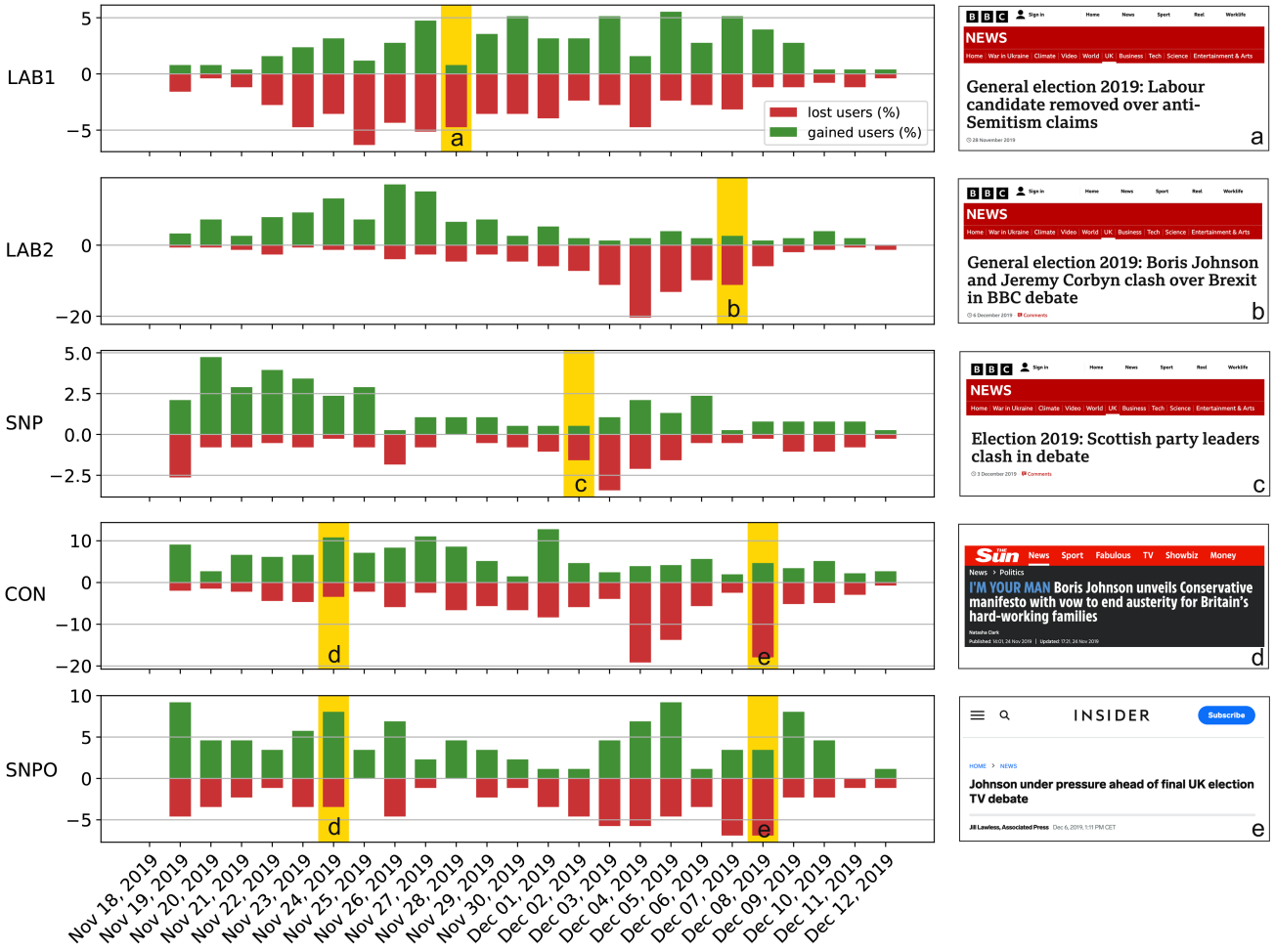
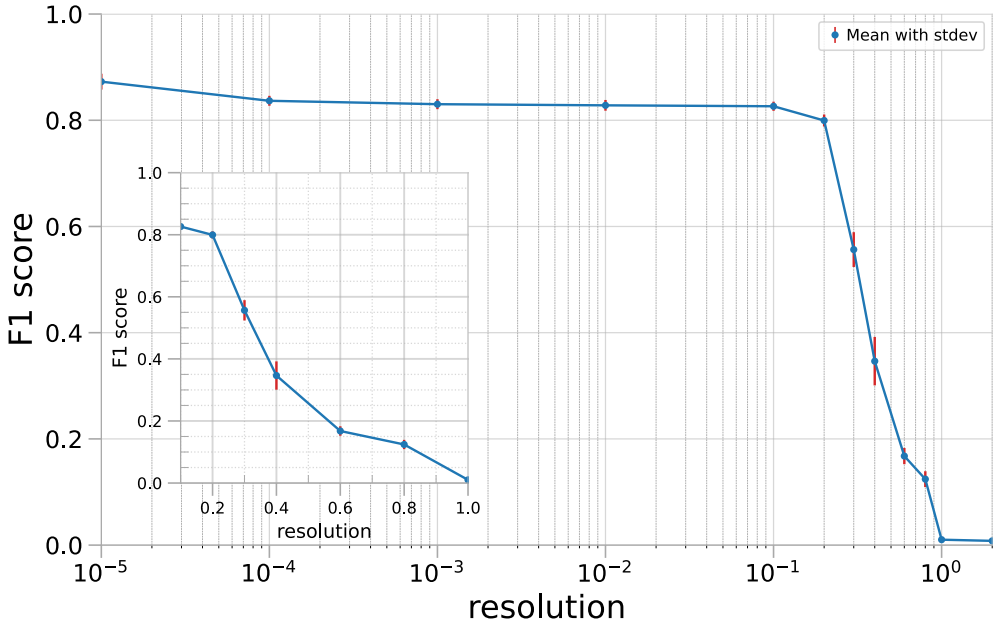


Fig. S9. Sensitivity of our method to the resolution parameter γ of the Leiden community detection algorithm. For each resolution value tested in our grid search, we report the mean and standard deviation of the distribution of F1 scores obtained with all combinations of time window lengths and steps. The inset focuses on the range of resolution values that correspond to the large differences in F1 scores.



References

1. M Alizadeh, JN Shapiro, C Buntain, JA Tucker, Content-based features predict social media influence operations. *Sci. advances* **6**, eabb5824 (2020).
2. X Wang, J Li, E Srivatsavaya, S Rajtmajer, Evidence of inter-state coordination amongst state-backed information operations. *Sci. reports* **13**, 7716 (2023).
3. Q Kong, P Calderon, R Ram, O Boichak, MA Rizoiu, Interval-censored transformer hawkes: Detecting information operations using the reaction of social systems in *Proceedings of the ACM Web Conference 2023*. pp. 1813–1821 (2023).
4. D Schoch, FB Keller, S Stier, J Yang, Coordination patterns reveal online political astroturfing across the world. *Sci. Reports* **12**, 4572 (2022).
5. AC Nwala, A Flammini, F Menczer, A language framework for modeling social media account behavior. *EPJ Data Sci.* **12**, 33 (2023).
6. R Robertson, Uncommon yet consequential online harms. *J. Online Trust. Saf.* **1** (2022).
7. D Jackson, E Thorsen, D Lilleker, N Weidhase, UK election analysis 2019: Media, voters and the campaign, (Bournemouth University Centre for Comparative Politics and Media Research), Technical report (2019).
8. S Pei, L Muchnik, JS Andrade, Jr, Z Zheng, HA Makse, Searching for superspreaders of information in real-world social media. *Sci. Reports* **4**, 5547 (2014).
9. W Webber, A Moffat, J Zobel, A similarity measure for indefinite rankings. *ACM Transactions on Inf. Syst. (TOIS)* **28**, 1–38 (2010).
10. A Trujillo, S Cresci, Make Reddit Great Again: Assessing community effects of moderation interventions on r/The_Donald in *Proceedings of the 25th ACM Conference On Computer-Supported Cooperative Work And Social Computing (CSCW'22)*. (ACM), pp. 1–28 (2022).
11. General election 2019: Labour candidate removed over anti-Semitism claims (BBC News) (2019) https://www.bbc.com/news/election-2019-50585278, accessed 19 September 2023.
12. General election 2019: Boris Johnson and Jeremy Corbyn clash over Brexit in BBC debate (BBC News) (2019) https://www.bbc.co.uk/news/election-2019-50681321, accessed 19 September 2023.
13. Election 2019: Scottish party leaders clash in debate (BBC News) (2019) https://www.bbc.com/news/election-2019-50637417, accessed 19 September 2023.
14. N Clark, I'M YOUR MAN Boris Johnson unveils Conservative manifesto with vow to end austerity for Britain's hard-working families (The Sun) (2019) https://www.thesun.co.uk/news/10412356/boris-johnson-conservative-manifesto-launch/, accessed 19 September 2023.
15. J Lawless, Johnson under pressure ahead of final UK election TV debate (Insider) (2019) https://www.insider.com/johnson-under-pressure-ahead-of-final-uk-election-tv-debate-2019-12, accessed 19 September 2023.
16. D Powers, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *J. Mach. Learn. Technol.* **2**, 37–63 (2011).
17. D Pacheco, et al., Uncovering coordinated networks on social media: Methods and case studies in *The 15th International AAAI Conference on Web and Social Media (ICWSM'21)*. pp. 455–466 (2021).
18. T Magelinski, L Ng, K Carley, A synchronized action framework for detection of coordination on social media. *J. Online Trust. Saf.* **1** (2022).
19. L Nizzoli, S Tardelli, M Avvenuti, S Cresci, M Tesconi, Coordinated behavior on social media in 2019 UK General Election in *The 15th International AAAI Conference on Web and Social Media (ICWSM'21)*. pp. 443–454 (2021).
20. VL Dao, C Bothorel, P Lenca, Community structure: A comparative evaluation of community detection methods. *Netw. Sci.* **8**, 1–41 (2020).
21. S Fortunato, M Barthelemy, Resolution limit in community detection. *Proc. Natl. Acad. Sci.* **104**, 36–41 (2007).
22. N Veldt, D Gleich, A Wirth, Learning resolution parameters for graph clustering in *The 2023 ACM Web Conference (WWW'19)*. pp. 1909–1919 (2019).