Insights into using temporal coordinated behaviour to explore connections between social media posts and influence

Elisa Sartori [1], Serena Tardelli [2], Maurizio Tesconi[2], Mauro Conti [1,3], Alessandro Galeazzi [1], Stefano Cresci [2], Giovanni Da San Martino [1]

[1]University of Padova, Italy,
[2]Institute of Informatics and Telematics, National Research Council, Italy,
[3]Örebro University, Sweden
Correspondence: elisa.sartori.7@phd.unipd.it

Abstract

Political campaigns increasingly rely on targeted strategies to influence voters on social media. Often, such campaigns have been studied by analysing coordinated behaviour to identify communities of users who exhibit similar patterns. While these analyses are typically conducted on static networks, recent extensions to temporal networks allow tracking users who change communities over time, opening new opportunities to quantitatively study influence in social networks. As a first step toward this goal, we analyse the messages users were exposed to during the UK 2019 election, comparing those received by users who shifted communities with others covering the same topics. Our findings reveal 54 statistically significant linguistic differences and show that a subset of persuasion techniques, including loaded language, exaggeration and minimization, doubt, and flag-waving, are particularly relevant to users' shifts. This work underscores the importance of analysing coordination from a temporal and dynamic perspective to infer the drivers of users' shifts in online debate.

1 Introduction

The unprecedented rise in social media usage creates the opportunity to share ideas and opinions with an ever-broader audience. This is particularly impactful in the political context, as demonstrated by the central role of social media in political campaigns and by the widespread use of targeted digital strategies to influence and coordinate voters (Ate et al., 2023). Several works study this phenomenon from different angles. Social media content has been examined for its use of persuasion techniques (Moral et al., 2023; Alam et al., 2022) and linguistic tricks (Stepaniuk and Jarosz, 2021), or through the network of user interactions (Mastroeni et al., 2023), e.g., relying on retweets or hashtags, to identify coordination among users (Pacheco et al., 2020). The analysis of coordination yields clusters of users that allow to extrapolate information about them: Nizzoli et al. (2021) show that the main groups identified correspond to supporters of political parties and activist groups during the UK 2019 elections.

Recently, Tardelli et al. (2024) extended the static analysis on coordination to the temporal dimension. In their work, they uncover different classes of user behaviour, which they map to archetypes. In particular, one of the archetypes, *Archetype 2*, corresponds to users that change their original community and stay in the destination community for a relatively long time. This temporal behaviour is especially interesting, as "the shifts detected via dynamic analyses of coordination could contribute to identifying successful cases of influence over users or communities in a network" (Tardelli et al., 2024), therefore possibly providing an invaluable tool for quantitative studies of persuasion on social networks.

As an initial step toward this goal, we first verify the quality of the dynamic communities identified in Tardelli et al. (2024) by using different types of user interactions to compute them, i.e. likes instead of retweets. Then we analyse the messages that users who changed community (Archetype 2) were exposed to, in order to detect linguistic cues related to persuasion.

Specifically, we tackle the following research questions:

- RQ1 Would interaction signals different than retweets and hashtags still yield comparable communities?
- RQ2 Are there significant differences between the messages that the users who have changed community (Archetype 2) have been exposed to and other messages shared in the same time period?

The contributions of the paper are the following:

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 24392–24404
November 4-9, 2025

i) we compare the dynamic communities based on retweets with the study of the *like* patterns of the users and show the consistency of the two results; *ii)* in Tardelli et al. (2024) only superspreader users (the 1% of users with the highest number of retweets) have been considered; we extend the computation of communities to all users to allow us to have more complete analyses; *iii)* we compare the content of the posts that users who have changed community have been exposed to with random sets of posts (still on the election topics), showing differences in the use of several linguistic features and an increased presence of persuasion techniques.

2 Related works

Social media networks and user behaviour. In the literature, we find multiple works addressing different types of user behaviour and their relationship to influence, such as building a retweet network to analyse the influence of opinions on wind energy (Mastroeni et al., 2023). In the context of politics, another form of user behaviour, coordination, has gained interest, as it is necessary for large-scale online campaigns. Nizzoli et al. (2021) present a network-based framework that discovers coordination as a significant similarity between users by constructing a user similarity network. However, this method aggregates user activities and does not consider their variations over time. To close this gap, Tardelli et al. (2024) apply a dynamic community detection algorithm to identify groups of users with similar behaviours, and analyse their changes over time. In their analysis they describe two types of users, which they refer to as Archetypes. Specifically, *Archetype 1* or "stationary" users are the ones who do not change community in the period under consideration and *Archetype 2* users are the ones who change community and then remain in the destination community for a long time. Related to our work is the paper from Hristakieva et al. (2022), which combines the analysis of static coordinated communities with the messages shared within them to identify the use of propaganda during the UK 2019 elections. However, besides not taking into account the temporal evolution of communities, it applies a classifier for propagandistic texts, while we focus on the analysis of messages to find linguistic cues related to persuasion.

Influence and social media content. Da San Martino et al. (2019) propose a BERT-based multi-granularity model capable of identifying the presence and location of 18 persuasion techniques, selected from those commonly present in political propaganda (Nakov et al., 2021b,a). The work of Stepaniuk and Jarosz (2021) deals with shorter texts, analysing Facebook posts from Polish travel agencies. They investigate the presence of Persuasive Linguistic Tricks which, however, are textual cues tailored to marketing and are not adaptable to the political context. A previous work from Addawood et al. (2019) identified and measured the use of 49 potential context-independent deceptive language cues in tweets from fraudulent accounts. Their work shows that these types of linguistic features can help discriminate troll accounts from authentic ones and may also be useful when addressing influence.

3 Dataset

As our goal is to investigate the changes in the communities highlighted in the work of Tardelli et al. (2024), we use the same dataset: the Twitter 2019 UK Election dataset, which was first presented in Nizzoli et al. (2021). From the 1M distinct users in the dataset, Tardelli et al. (2024) extract 12K superspreaders, i.e. the 1% of users with the highest number of retweets, and 3M tweets they shared, of which 441K are original content (i.e., not retweets). They construct a multiplex temporal network based on co-retweets. The network is built over overlapping time-windows of 7 days each, with a 1 day shift between them. Then they apply the Leiden community detection algorithm to identify and track the evolution of communities over time and observe the behaviour of superspreaders, i.e. whether they stay in the same community or change it at some point in time.

The main communities found in the dataset are: **LAB1-**labourist party, **LAB2-**labourists with different temporal behaviours than LAB1, **RCH-**labourists spreading the manifesto and pushing others to vote, **B60-**users against the pension age equalization law, **TVT-**a group composed of multiple political parties militating for a tactical vote in favor of labourists, **SNP-**users supporting the Scottish National Party, **SNPO-**opposers to the Scottish National Party, **CON-**conservative party, **ASE-**conservative party engaged in attacking the labour party, and **BRX-**users in favor of Brexit.

4 Dataset Extension

In the dataset, only 211 out of $12K$ superspreaders belong to Archetype 2, not enough to present solid results. To overcome this, we extend the analysis to non-superspreader users (**NonSS**), who had been excluded in the original work. To identify communities, Tardelli et al. (2024) used the Leiden community detection algorithm, which identifies more densely connected groups, as such it cannot be applied to non-superspreaders, of whom only 6% made at least 20 retweets in the entire month of collection (see Table 1). Therefore, in the rest of this section we describe a method to assign NonSS users to communities. Our method leverages on stationary superspreaders (Archetype 1), users that are representative of their community because they have not moved to another one. For each time-window, we represent each user as the vector of retweets made during this period. We then assign NonSS users to the majority community based on their nearest stationary superspreaders, using cosine distance computed from retweet vectors.

#retweets made	#users	percentage
1 retweet	1'167'798	100%
2 retweets	594'786	51%
3 retweets	414'402	35%
5 retweets	264'359	23%
10 retweets	142'602	12%
20 retweets	74'033	6%

Table 1: Retweets made by non superspreaders.

In order to evaluate the quality of the algorithm, we use two methods:

1. we evaluate the accuracy of the assignment using subsets of superspreaders whose community at each time-window is already known;

2. we measure intra-community and inter-community distances as a way to define the severity of assignment errors, leveraging the knowledge that some communities should be more similar (e.g., two left-leaning communities) than others (e.g., a left-leaning community and a right-leaning one).

Figure 1 shows lineplots of the average cosine distance between same, similar and dissimilar communities. The distance is computed with respect to the retweet vectors. Considering vectors composed of all retweets made by the users, the average distances between same, similar and dissimilar communities appeared to fall within a very close range (see "all tweets" in Figure 1), making a clear separation difficult.

By checking the percentage of tweets in common between users of each community (Table 2), we notice that even users belonging to the same community do not share many retweets.

	avg %common	avg % different
intra-community	17.36	82.64
left-left	6.86	93.14
right-right	3.14	96.86
left-right	0.19	99.81

Table 2: Percentage of tweets in common between users of the same (**intra-community**), similar (**left-left**, **right-right**) or different community (**left-right**).

This means that only a subset of retweets is useful to associate users with communities, which, we hypothesize, results in the high average distance between users of the same community. We construct vectors using only tweets retweeted by at least one eighth and up to one third of the stationary users in each community. Indeed, by focusing on subsets of tweets that are retweeted by only a ratio of stationary members (third/.../eighth on the x-axis in Figure 1), the average distances become more widespread.

We also observe an overlap between the distance ranges of similar communities and same communities. This overlap is consistently given by the distances between communities **RCH**, **LAB1** and **LAB2**, which appear to be so close to each other that they result comparable with intra-community distances. As such, any assignment errors made between this three communities are negligible and the maximum of their distances is a good upper bound for the "slight" error category.

Let A be a user belonging to community x but assigned to community y, and be $dist(c_x, c_y)$ the distance between communities x and y, we classify assignment errors as follows:

1. **slight**, errors between very similar communities;

$$dist(c_x, c_y) \leq max(dist(c_{RCH}, c_{LAB1}),\\ dist(c_{RCH}, c_{LAB2}), dist(c_{LAB1}, c_{LAB2}))$$

2. **moderate**, errors between similar communities;

$$max(dist(c_{RCH}, c_{LAB1}), dist(c_{RCH}, c_{LAB2}),\\ dist(c_{LAB1}, c_{LAB2})) < dist(c_x, c_y) < 0.99$$

3. **severe**, errors between dissimilar communities.

   $dist(c_x, c_y) \geq 0.99.$

To measure the accuracy of our algorithm, we use stationary and Archetype 2 superspreaders. Since these classes have very distinct behaviours, they are ideal for evaluation. We assign a community to each stationary user using the rest of the stationary superspreaders, and assign each Archetype 2 user using all stationary superspreaders. We decided to use the subset of tweets retweeted by at least 1/8 of the stationary community members and assign a user to the majority community among the 4 closest stationary superspreaders, as this provided the best trade-off between accuracy and errors, which can be seen in Table 3. In particular, if we consider slight errors as matches, since they occur between communities whose distance is comparable to an intra-community one, we reach an accuracy of 95.33% for stationary users and 85.21% for Archetype 2 users. The results for all subsets can be found in Appendix A.3.

By applying the algorithm and definition of Archetype 2 users to all NonSS users who have at least one retweet for each time-window, we obtain a total of 8562 Archetype 2 users.

5 Analysis

5.1 RQ1: Assesment of dynamic analysis

The dynamic communities in Tardelli et al. (2024) were identified and analysed using retweets and hashtags. Since our goal is to determine the robustness of their findings, we repeat the analysis with respect to user likes, which we retrieved through the Twitter API, an interaction signal previously not used, to see if we obtain consistent results.

Methodology - We only consider superspreaders since the original analyses were performed on that subset of users. First, we show that likes are used differently than retweets. Figure 2 shows a visual comparison of distributions of user likes and retweets. We further conducted a chi-square test, resulting in a p-value of $2.2E-16$ and an effect size using Cohen's w (Cohen, 1988) of 41.58, indicating a statistically significant difference.

To assess whether the dynamic communities found by Tardelli et al. (2024) using retweets remain consistent when analysing user likes, we consider the likes given to official accounts of political parties and their leaders.

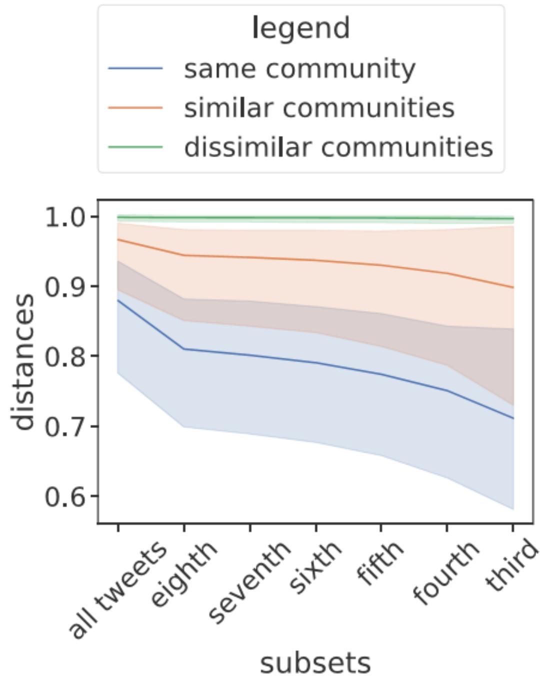

Figure 1: Lineplot of average distances between a **same community** (intra-community distance), **similar communities** (left-left or right-right), and **dissimilar communities** (left-right) with respect to different subsets of tweets. The band in each line diagram represents the range of the distribution: the minimum and maximum distances. *Third* = subset of tweets retweeted by at least 1/3 of communities' stationary members.

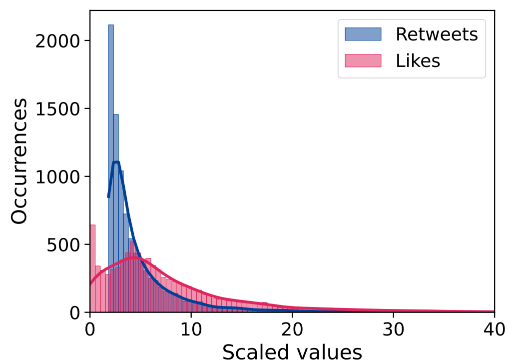

Figure 2: Comparison between users' likes and retweets. The values in the distributions were scaled between 0 and 100. Only the significant parts of the long-tailed distributions are shown.

		all tweets	seventh	eighth	eighth_maj_2	eighth_maj_3	eighth_maj_4	eighth_maj_5
stationary	Match	91.41%	91.54% (+0.13)	91.92% (+0.51)	91.92% (+0.51)	92.43% (+1.02)	92.52% (+1.11)	92.37% (+0.96)
vs	Mismatch	8.59%	8.46%	8.08%	8.08%	7.57%	7.48%	7.63%
stationary	severe errors		0.19%	0.20%	0.20%	0.18%	0.17%	0.18%
2639 users	Match + slight errors		94.34% (+2.93)	94.74% (+3.33)	94.74% (+3.33)	95.22 (+3.33)%	95.33% (+3.92)	95.31% (+3.90)
arch2	Match	52.70%	53.11% (+0.41)	54.11 % (+1.41)	54.11% (+1.41)	52.36% (-0.34)	53.09%(+0.39)	52.90% (+0.2)
vs	Mismatch	47.30%	46.89%	45.89%	45.89%	47.64%	46.91%	47.10%
stationary	severe errors		0.81%	0.80%	0.80%	0.98%	0.92%	1%
211 users	Match + slight errors		83.26% (+30.56)	84.30% (+31.6)	84.30% (+31.6)	84.52% (+31.82)	85.21% (+35.51)	85.42% (+32.72)

Table 3: Accuracy of our algorithm for assigning users to communities considering different subsets of tweets (*seventh*, *eighth*) and number of neighbors(*maj_2*, *maj_3*, *maj_4*, *maj_5*). *Seventh* = subset of tweets retweeted by at least 1/7 of communities' stationary members, *maj_x* = assignment given to majority community according to x closest stationary members. Full results can be seen in Table A1 in the Appendix.

At the time, 8 parties were running for the election [12]: *CON*-Conservative Party (right-leaning), *LAB*-Labour Party (center-left), *SCO*-Scottish National Party (center-left), *DEM*-Liberal Democrats (center), *CYM*-Plaid Cymru (left), *GRE*-Green Party of England and Wales (left), *REF*-Reform UK (Brexit) Party (right) and *CUK*-Change UK (center). Change UK was dissolved in December 2019, leaving us unable to identify the official Twitter account IDs, and was therefore excluded from the analysis. We assign each superspreader to the party with the highest number of liked tweets. We conclude by visualizing the percentages of user overlap between Tardelli's communities (**Algorithm communities**) and the communities created using likes (**Likes communities**), as shown in Figure 3. We first calculate the overlap for each time-window, then we aggregate all time-windows and scale the results. The same analysis performed on non-superspreader users, giving the same results, is reported in Appendix A.2.

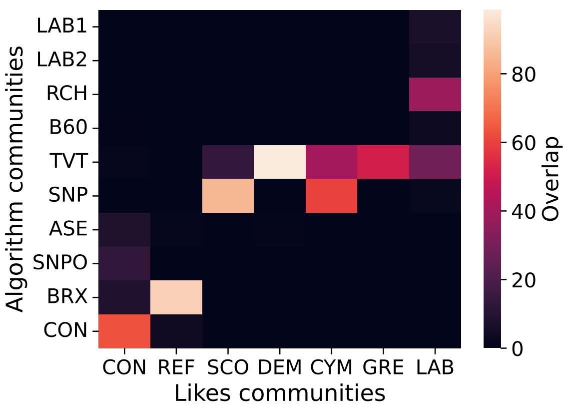

Figure 3: Overlap of superspreaders within communities created using likes to political parties and communities found by Tardelli's algorithm.

As a further consistency check we computed the political polarization of the communities with respect to likes and compare it with Tardelli et al. (2024), where it was computed with respect to hashtags. We calculate our polarization score by once again considering user likes to parties' official accounts. Each account is assigned a score $s \in [0, 1]$ based on the political orientation declared by the party: 1 for right-wing parties, 0.75 for the center-right, 0.5 for the center, 0.25 for center-left and 0 for the left. Finally, the community polarity score is calculated in two steps. We first multiply the number of likes that community members have given to the parties' official accounts by the respective polarity score of those accounts. Then, we add up the values and divide the result by the total number of likes members collectively gave to official accounts. We calculate the community polarity score for each time-window, we then aggregate all time-windows and scale the results so that communities at the extremes of the spectrum are at the extremes of the plot, as done by Tardelli et al. (2024). The results of this process can be seen in Figure 4.

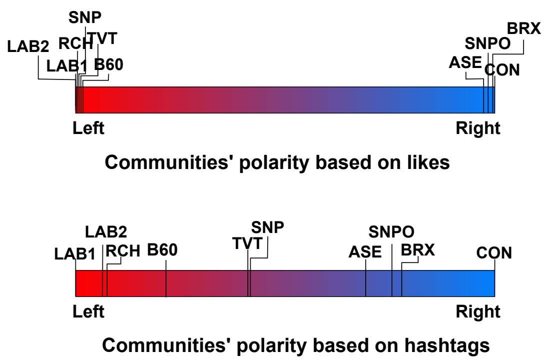

Figure 4: Comparison of communities' polarity as in the original work (using hashtags) and our approach using likes given to official political party accounts.

Findings - Looking at Figure 3, we can observe some consistencies with the results obtained by

[1] https://www.bbc.co.uk/newsround/50459942l

[2] https://en.wikipedia.org/wiki/2019_United_Kingdom_general_election

Tardelli. There is a high overlap in the communities *REF*, Reform UK campaigning for Brexit, and **BRX**, which the authors found was composed by users in favor of Brexit. Similarly, the conservative parties *CON* and **CON** and the communities in favor of the Scottish National Party *SCO* and **SNP** share a high percentage of users. We can also see a certain degree of overlap in the two labourist parties *LAB* and **RCH**. In addition to that, the community **TVT** comprises many left-wing parties. This aligns with the fact that it is a group of multiple political parties militating for a tactical vote favoring labourists. There is only a very slight overlap between the community *LAB* and the labourist communities **LAB1**, **LAB2** and **B60**. This is probably due to the fact that **RCH** and **TVT** are bigger in size; they have 80% more users, and this causes the values of the rest of the communities to be overshadowed. A similar argument applies to the conservative **ASE** community, with **CON** having 75% more users. Lastly, there is virtually no overlap between right-winged communities and left-leaning ones. This in confirmed by Figure 4, showing that the polarities are mostly consistent: left-leaning communities are still left-leaning communities, and the same goes for right-leaning ones. This consistency confirms that the communities found are robust.

5.2 RQ2: Comparison between messages Archetype 2 users were exposed to versus other messages shared

After having expanded the number of Archetype 2 users, we investigate whether the messages they were exposed to present any peculiarities.

Methodology - As previously introduced, Archetype 2 users are individuals who belonged to a specific community that we call the original community and who, at a certain point, shift and remain for a long time in a different community that we call the destination community. We begin by investigating whether the shift can be linked to a change in user behaviour. We compare the number of likes provided by Archetype 2 users to tweets written by members of the original community (**og**) and destination community (**dest**). We consider the time-window prior to the shift (**tw_before**), and the time-window after (**tw_after**). We measure the changes in likes given before and after the shift ($likes(tw\_after) - likes(tw\_before)$). The results are reported in Table 4, showing that for up to 65% of users there is a change in behaviour after the shift. The results are quite heterogeneous, but when aggregated we see that users tend to like the destination community more after the shift (35.89%, compared to 24.47% who like it less) and like the original community less (24.65%, compared to 18.44% who like it more). One reason why no specific trend emerges is that the majority (97.53%) are non-superspreaders, who were less active during the period and produced fewer and noisier data.

	all Arch. 2	high-conf. Arch. 2
Total (>= 1 like to a community)	5918	2348
likes og and dest do not change	2032 (34.33%)	769 (32.75%)
likes og and dest both change	2236 (37.78%)	995 (42.38%)
* -likes og, + likes dest	1127	546
* +likes og, - likes dest	722	304
* -likes og, - likes dest	166	60
* +likes og, + likes dest	221	85
same likes og, dest changes	1336 (22.58%)	437 (18.61%)
* +likes dest	776	262
* -likes dest	560	175
same likes dest, og changes	314 (5.31%)	147 (6.26%)
* +likes og	148	67
* -likes og	166	80

Table 4: Comparison of differences in community likes given by Archetype 2 users in the time-windows before (**tw_before**) and after (**tw_after**) the shift. We consider all Archetype 2 users and the subset of those who had high confidence ($> 50\%$) in the community assignment before the shift. Confidence is calculated as the percentage of neighbors belonging to the assigned community.

To find potential signals that distinguish the content perceived by Archetype 2 users compared to the rest, we define five sets of tweets: **A2**, the tweets that Archetype 2 users liked in the time-window just before the shift, and four disjoint random sets as control groups (**Rand1**, **Rand**, **Rand3** and **Rand4**) with the same number of tweets in **A2** (18'098) and that do not contain any post liked by Archetype 2 users. Following Addawood et al. (2019), we measure the occurrences of linguistic features in our sets of tweets. One of the tools used to compute the features, LIWC (Tausczik and Pennebaker, 2010), was updated in 2022. While maintaining the old version of the dictionaries, we also include the updated sense terms (Attention, Motion, Space, Visual, Auditory, Feeling), their aggregated feature Perception, and four new categories: Clout (language of leadership and status), Authentic (perceived honesty and genuineness), Analytic (metric of logical and formal thinking) and Tone (degree of emotional tone). Furthermore, we add other textual features and metadata that were not considered in the original work. As features we add extra punctuation classes (all punctuation, periods, exclamation points, commas, and a class for other punctuation marks) and emojis. As metadata, we include the number of likes and replies to a tweet. We end up with a total of 79 features. To verify the importance of differences among the features we perform a chi-square test. Table 5 shows the number of significant features for each comparison.

		Rand1	Rand2	Rand3	Rand4
p-value<0.001	small ES	0	0	0	0
	medium ES	0	0	0	0
	large ES	54	54	54	54

Table 5: Number of statistically different features between **A2** and each of the random groups.

The 54 statistically significant features are the same for all comparisons. We proceed by comparing their average to assess how they change. For each feature we define an importance score, calculated as $|min(A_2 - Rand_x)| - 2 * max(|Rand_x - Rand_y|)$. We set a threshold of 0.01 for the importance score, considering as meaningful those features for which the minimum difference between A2 and the Random sets is at least twice the maximum difference observed among the Random sets. We consider values lower than 0.01 to be negligible. Of the 54 common features, 40 appear more in content liked by Archetype 2 users, of which 31 are meaningful: tweet engagement (likes, retweets, follows), author outreach (following and listed count), information given and expressivity (length of tweet, number of words, words per sentence, articles, adjectives, verbs, adverbs, function words, conjunctions), quotations, commas, logic (Analytic), emotional language (Tone), leadership/status (Clout), genuineness (Authentic), group references (we), sense terms (all sense terms, see), relativity (space, time, motion) and focus on the present. Fifteen appear less, of which 12 are meaningful: author productivity (tweet count), general punctuation(all, exclamation point, other less common punctuation marks), words with more than six letters, hastags, numbers, emojies and exclusionary markers (negation, exclusion words). Two are mixed or very close in values and not meaningful. Table 7 shows the differences that emerge for all 54 statistically significant features we identified to set apart content proposed to Archetype 2 users.

Finally, we investigate the presence and use of persuasion techniques using Tanbih API for propaganda techniques detection [3]. The model is trained to detect 7 techniques (Loaded Language, Name Calling, Doubt, Flag Waving, Exaggeration or minimisation, Repetition, Flag Waving and Causal Oversimplification) plus 12 additional less common techniques that are grouped as Other.

We conduct a chi-square test to assess whether persuasion techniques are used differently across the tweets in **A2** and the four control sets randomly selected: **Rand1**, **Rand**, **Rand3** and **Rand4**. The results are reported in Table 6, indicating both statistical and practical significance.

	Rand1	Rand2	Rand3	Rand4
p-value	1.8E-17	1.0E-11	6.4E-22	3.0E-16
effect-size	0.49	0.35	0.61	0.46

Table 6: Results of chi square test on use of persuasion techniques between **A2** and each of the random groups. Cohen's w was used to determine the effect-size.

Findings - We find 54 statistically significant features that set apart content proposed to Archetype 2 users. There are 16 features in common with the 19 most important in predicting disingenuous accounts identified by Addawood et al. (2019), although with varying degrees of importance. These are: Hashtags, Number of Retweets for a Tweet, Nouns, Tweet length, Authors tweet count, Author followers count, Words per sentence, Words with more than 6 letters, Self references, Hedges, Author following, Causation, Sense Terms, All punctuation, Function words and Verbs. Furthermore, as seen in Table 6, persuasion techniques are present and used differently. In particular, loaded language, exaggeration and minimisation, doubt and flag-waving occur much more in tweets to which Archetype 2 was exposed (see Appendix A.5). It is important to note that statistical significance does not imply causality; there could be multiple factors explaining why a user switched communities, factors that can also occur outside the digital space.

6 Conclusion

The temporal analysis of coordinated behaviour highlights users that changed community over time. This may open up new possibilities to quantitatively study influence in social networks. By analysing the patterns of likes usage, we provided additional evidence supporting the communities identified through temporal analysis. In addition, we analysed the messages that users who changed community have been exposed to, comparing them with

[3] https://apihub.tanbih.org/docs

Feature	A2	Rand1	Rand2	Rand3	Rand4	min(A2 – Randx)	max(|Randx - Randy|)	importance scores	type diff
Tweet_likes	605.729	52.184	78.667	44.707	45.02	527.062	33.96	459.142	bigger
Tweet_retweets	207.842	18.45	22.968	16.154	17.057	184.874	6.814	171.246	bigger
Tweet_replies	58.291	4.634	4.885	3.955	4.242	53.406	0.93	51.546	bigger
Tweet_number_char	202.808	175.024	173.708	173.675	174.714	27.783	1.35	25.083	bigger
Author_followers_count	21.196	1.585	1.459	1.238	1.718	19.478	0.48	18.518	bigger
Quotations	22.068	2.705	2.097	1.411	2.153	19.363	1.294	16.775	bigger
Analytic	65.053	57.522	57.88	57.518	57.87	7.173	0.362	6.449	bigger
Information_quantity_number_words	32.88	27.194	27.016	26.972	27.155	5.686	0.222	5.242	bigger
Tone	36.039	30.365	30.376	30.388	29.656	5.651	0.733	4.185	bigger
Function_words	41.766	38.117	37.981	38.12	38.061	3.646	0.139	3.368	bigger
Clout	57.53	53.507	53.007	52.942	53.316	4.023	0.566	2.891	bigger
Authentic	29.122	26.18	26.112	26.339	26.58	2.541	0.468	1.605	bigger
Words_per_sentence	13.946	12.506	12.454	12.394	12.589	1.357	0.195	0.967	bigger
Sense_terms_perception_2022	7.235	6.275	6.217	6.188	6.226	0.96	0.087	0.786	bigger
Articles	5.925	4.98	5.05	5.023	5.001	0.874	0.07	0.734	bigger
Relativity_space	5.018	4.266	4.253	4.262	4.268	0.75	0.015	0.72	bigger
Relativity_time	4.003	3.496	3.53	3.489	3.519	0.472	0.042	0.388	bigger
Information_quantity_adjectives	5.571	5.173	5.106	5.155	5.112	0.398	0.067	0.264	bigger
Group_reference_we	1.661	1.225	1.281	1.218	1.265	0.38	0.064	0.252	bigger
Discouse_markers_conj	3.699	3.445	3.441	3.446	3.435	0.253	0.011	0.231	bigger
Information_complexity_commas	2.248	1.993	1.96	1.965	1.952	0.255	0.042	0.171	bigger
Information_quantity_verbs	5.593	5.262	5.215	5.243	5.286	0.306	0.071	0.164	bigger
Information_quantity_adverbs	3.233	3.078	3.034	3.052	3.057	0.156	0.044	0.068	bigger
Present_focus	4.902	4.623	4.525	4.63	4.57	0.272	0.105	0.062	bigger
Author_listed_count	0.06	0.007	0.007	0.006	0.007	0.053	0.001	0.051	bigger
Sense_terms_see_2015	0.8	0.72	0.707	0.716	0.721	0.079	0.015	0.049	bigger
All_sense_terms_2015	1.69	1.57	1.549	1.541	1.534	0.119	0.037	0.045	bigger
Relativity_motion	1.134	0.99	0.956	0.95	0.997	0.137	0.047	0.043	bigger
Sense_terms_visual_2022	0.69	0.648	0.632	0.628	0.636	0.042	0.019	0.004	bigger
Modifier_words	0.005	0.004	0.004	0.004	0.004	0.001	0	0.001	bigger
Morality_authority_virtue	0.012	0.01	0.01	0.01	0.01	0.001	0.001	-0.001	bigger
Author_following_count	0.194	0.152	0.161	0.144	0.142	0.033	0.019	-0.005	bigger
Group_reference_they	0.698	0.653	0.657	0.67	0.682	0.017	0.028	-0.039	bigger
Causation	1.2	1.171	1.126	1.132	1.133	0.03	0.045	-0.06	bigger
Quotations_single_quotes	1.349	1.324	1.266	1.318	1.29	0.024	0.058	-0.092	bigger
Sense_terms_hear	0.584	0.579	0.556	0.545	0.518	0.005	0.061	-0.117	bigger
Information_complexity_periods	6.086	6.083	6.055	5.935	5.893	0.002	0.191	-0.38	bigger
Author_tweet_count	80884.808	93970.931	92497.16	92780.149	93227.451	-13086.123	1473.771	10138.581	smaller
Information_complexity_other_punctuation	14.848	20.525	20.617	20.603	20.356	-5.768	0.261	5.246	smaller
Information_complexity_all_punctuation	27.498	33.677	33.517	33.588	33.176	-6.178	0.5	5.178	smaller
Words_>_six_letters	25.069	28.53	28.446	28.65	28.426	-3.58	0.224	3.132	smaller
Hashtags	6.278	8.601	8.579	8.609	8.463	-2.331	0.147	2.037	smaller
Use_of_numbers	8.717	10.619	10.636	10.68	10.613	-1.963	0.067	1.829	smaller
Information_complexity_exclamation_marks	0.879	1.414	1.382	1.433	1.391	-0.553	0.051	0.451	smaller
Emoji	2.728	4.649	3.85	4.09	3.975	-1.921	0.799	0.323	smaller
Information_complexity_question_marks	0.531	0.772	0.731	0.757	0.751	-0.241	0.041	0.159	smaller
Information_quantity_nouns	7.284	7.587	7.508	7.58	7.493	-0.304	0.094	0.116	smaller
Discourse_markers_negation	1.407	1.516	1.572	1.577	1.564	-0.17	0.062	0.046	smaller
Exclusion_words	1.537	1.594	1.619	1.586	1.605	-0.082	0.033	0.016	smaller
Emotions_neg	0.499	0.513	0.521	0.539	0.548	-0.049	0.036	-0.023	smaller
Emotions_pos	0.669	0.702	0.685	0.746	0.677	-0.077	0.069	-0.061	smaller
Hedges	0.024	0.025	0.025	0.024	0.025	-0.002	0.001	0	smaller
Self_reference	1.065	1.069	1.037	1.047	1.06	-0.004	0.032	-0.06	mixed
Morality_ingroup_virtue	0.007	0.008	0.007	0.007	0.007	-0.001	0.001	-0.001	mixed

Table 7: Comparison of averages of statistically and practically significant linguistic features and metadata. Included is the minimum difference among **A2** and the Random sets, and the maximum difference among the Random sets. Colors distinguish between types of differences and meaningful versus not meaningful features (importance score > 0.01).

random messages on the same topic. We found 54 statistically significant different linguistic features, as well as different use of some persuasion techniques, namely loaded language, exaggeration and minimisation, doubt, and flag-waving.

However, there are many factors that can push a user to switch communities, and these can happen across multiple platforms or even outside the online sphere. As future work, we plan to set up an experiment with humans in a controlled environment to test whether messages with the characteristics we identified tend to be more persuasive than, for example, messages having other persuasion techniques or other linguistic features.

7 Ethics Policy

Although our work is done to study the effect of coordinated behaviour in influencing online users, the features that we found to be more effective could be exploited with intent to harm. However, for example, knowing that flag waving seem to be an effective persuasion strategy does not provide messages that are effective in every scenario.

8 Limitations

Our work reveals some possible linguistic features that could be used alongside other NLP techniques to improve tools that work on targeted digital strategies, but we recognize several limitations. Although we attempted to limit random errors by using four control groups, our results may not be generalizable or limited to this dataset. Moreover, repeating these analyses on other datasets is necessary to consolidate our findings.

Acknowledgments

This work is partially funded by the European Union under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of March 15, 2022 of Italian Ministry of University and Research – NextGenerationEU; Code PE00000014, Concession Decree No. 1556 of October 11, 2022 CUP D43C22003050001, Progetto "SEcurity and RIghts in the CyberSpace (SERICS) - Spoke 2 Misinformation and Fakes - DEcision supporT systEm foR cybeR intelligENCE (Deterrence).

This work is also partially funded by the Qatar National Research Fund, part of Qatar Research Development and Innovation Council (QRDI) by grant NPRP14C0916-210015.

This work is also partly supported by SoBigData.it which receives funding from European Union – NextGenerationEU – National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) – Project: “SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics” – Prot. IR0000013 – Avviso n. 3264 del 28/12/2021.

References

Aseel Addawood, Adam Badawy, Kristina Lerman, and Emilio Ferrara. 2019. Linguistic cues to deception: Identifying political trolls on social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):15–25.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Asan Ate, Joseph Chiadika, Scotland Ekene, and Nwadiwe E. 2023. Use of social media and digital strategies in political campaigns. 8:1–13.

Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2 edition. Lawrence Erlbaum Associates.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. Association for Computational Linguistics.

Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, and Preslav Nakov. 2022. The spread of propaganda by coordinated communities on social media. In *14th ACM Web Science Conference 2022*, WebSci '22, page 191–201, New York, NY, USA. Association for Computing Machinery.

Loretta Mastroeni, Maurizio Naldi, and Pierluigi Vellucci. 2023. Wind energy: Influencing the dynamics of the public opinion formation through the retweet network. *Technological Forecasting and Social Change*, 194:122748.

Pablo Moral, Guillermo Marco, Julio Gonzalo, Jorge Carrillo de Albornoz, and Iván Gonzalo-Verdugo. 2023. Overview of dipromats 2023: automatic detection and characterization of propaganda techniques in messages from diplomats and authorities of world powers. *Procesamiento del Lenguaje Natural*, 71(0):397–407.

Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino, and Yifan Zhang. 2021a. COVID-19 in Bulgarian social media: Factuality, harmfulness, propaganda, and framing. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 997–1009, Held Online. INCOMA Ltd.

Preslav Nakov, Firoj Alam, Shaden Shaar, Giovanni Da San Martino, and Yifan Zhang. 2021b. A second pandemic? analysis of fake news about COVID-19 vaccines in Qatar. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1010–1021, Held Online. INCOMA Ltd.

Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated behavior on social media in 2019 uk general election. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):443–454.

Diogo Pacheco, Alessandro Flammini, and Filippo Menczer. 2020. Unveiling coordinated groups behind white helmets disinformation. In *Companion Proceedings of the Web Conference 2020*, WWW '20, page 611–616, New York, NY, USA. Association for Computing Machinery.

K. Stepaniuk and K. Jarosz. 2021. Persuasive linguistic tricks in social media marketing communication-the memetic approach. *PloS one*, 16(7).

Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti, Preslav Nakov, Giovanni Da San Martino, and Stefano Cresci. 2024. Temporal dynamics of coordinated online behavior: Stability, archetypes, and influence. *Proceedings of the National Academy of Sciences*, 121(20):e2307038121.

Yla Tausczik and James Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29:24–54.

A Appendix

A.1 Checking the possibility of echo chambers

Our findings would have been very limited if we had found ourselves in the case where Archetype 2 users were the only ones exposed to messages from other communities, while the rest of the users lived in an echo chamber and were only exposed to intra-community messages. To verify that, we look at the distribution of likes among stationary community members. We have to put this limitation because, if we also consider users who shift, since time-windows have overlapping days, we would not be able to know which community to assign it to among those to which they belonged. For each user, we check which community the author of the liked tweets belongs to. Finally, we aggregate the results for each community.

As we can see in Figure 5, users are not in an echo chamber. There are some communities (i.e., **RCH**, **CON**, **ASE**) where a large part of the likes are given to members of the same community, but in general, tweets from at least one other community are liked.

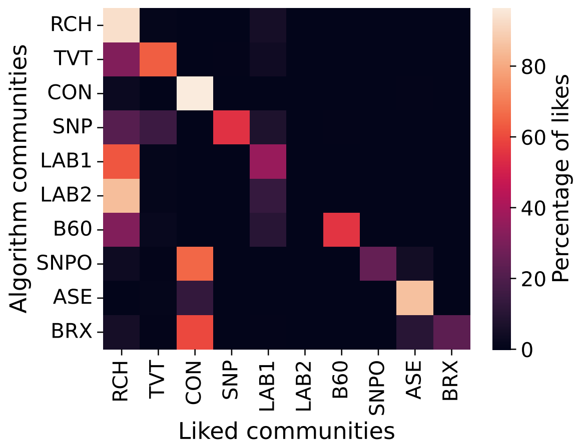

Figure 5: Likes given by stationary members to communities' posts.

A.2 Checking the consistency of assignments for NonSS

In our work, we did a series of analyses within Section 5.1 to ground the results obtained by Tardelli et al. (2024) using users' likes. However, when we extended the dataset to include non-superspreaders, we used a different algorithm; therefore, we should check whether the results obtained using NonSS remain consistent. We replicate the procedure used to create Figure 3 using all NonSS users resulting in Figure 6. The distribution among communities is very similar between the two figures, which shows that the results are consistent also for NonSS users.

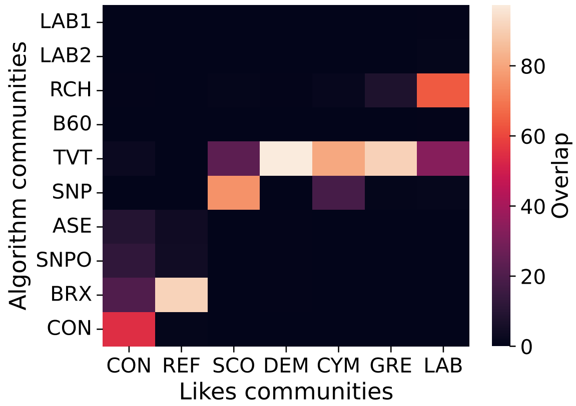

Figure 6: Overlap in NonSS users within communities created using likes to political parties and communities found by the algorithm.

A.3 Full results of assignment algorithm

The complete results of our algorithm for assigning NonSS users to communities are shown in Table A1.

A.4 Average distance between communities across subsets

Table A2 shows the average cosine distances between communities for all subsets of tweets considered in our work.

A.5 Use of persuasion techniques among A2 and the four Random sets

Figure 7 shows the difference in persuasion techniques occurrences among the tweets in **A2** and the four control sets **Rand1**, **Rand**, **Rand3** and **Rand4** described in Section 5.2. We can see a much higher occurrence in **A2** of the *Loaded_Language*, *Exaggeration-Minimization* and *Other*, less common techniques, alongside a slightly higher use of the techniques *Doubt* and *Flag-Waving*.

		all tweets	third	fourth	fifth	sixth	seventh	eighth	eighth_maj_2	eighth_maj_3	eighth_maj_4	eighth_maj_5
stationary	Match	91.41%	81.26%	86.95%	89.75%	90.99%	91.54%	91.92%	91.92%	92.43%	92.52%	92.37%
vs	Mismatch	8.59%	18.74%	13.04%	10.25%	9.01%	8.46%	8.08%	8.08%	7.57%	7.48%	7.63%
stationary	slight (%)		3.83%	2.94%	2.83%	2.77%	2.80%	2.82%	2.82%	2.79%	2.82%	2.94%
	medium (%)		14.43%	10.10%	7.18%	6.00%	5.47%	5.06%	5.06%	4.60%	4.49%	4.51%
2639 users	severe (%)		0.48%	0.26%	0.24%	0.23%	0.19%	0.20%	0.20%	0.18%	0.17%	0.18%
	Match + slight		85.09%	89.90%	92.58%	93.76%	94.34%	94.74%	94.74%	95.22%	95.33%	95.31%
	Mismatch - slight		14.91%	10.10%	7.42%	6.24%	5.66%	5.26%	5.26%	4.78%	4.67%	4.69%
arch2	Match	52.70%	43.97%	49.80%	51.44%	53.03%	53.11%	54.11%	54.11%	52.36%	53.09%	52.90%
vs	Mismatch	47.30%	56.03%	50.20%	48.56%	46.97%	46.89%	45.89%	45.89%	47.64%	46.91%	47.10%
stationary	slight (%)		29.14%	28.33%	29.02%	29.41%	30.15%	30.19%	30.19%	32.16%	32.12%	32.51%
	medium (%)		25.86%	20.84%	18.58%	16.70%	15.93%	14.91%	14.91%	14.50%	13.87%	13.58%
211 users	severe (%)		1.02%	1.02%	0.96%	0.86%	0.81%	0.80%	0.80%	0.98%	0.92%	1%
	Match + slight		73.11%	78.14%	80.46%	82.44%	83.26%	84.30%	84.30%	84.52%	85.21%	85.42%
	Mismatch- slight		26.89%	21,86%	19.54%	17.56%	16.74%	15.70%	15.70%	15.48%	14.79%	14.58%

Table A1: Accuracy of our algorithm for assigning users to communities considering different subsets of tweets (*third*, *fourth*, *fifth*, *sixth*, *seventh*, *eighth*) and number of neighbors(*maj_2*, *maj_3*, *maj_4*, *maj_5*). *Third* = subset of tweets retweeted by at least 1/3 of communities' stationary members, *maj_x* = assignment given to majority community according to x closest stationary members.

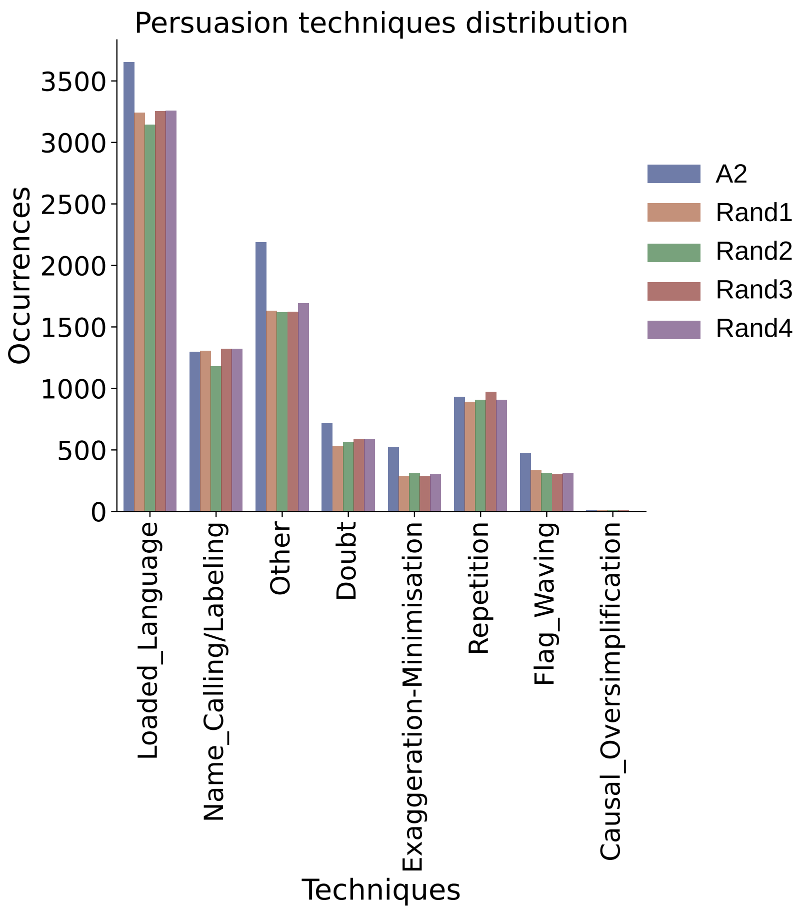

Figure 7: Persuasion techniques occurrences among sets **A2**, **Rand1**, **Rand**, **Rand3** and **Rand4**.

Com 1	Com 2	all tweets	third	fourth	fifth	sixth	seventh	eighth
RCH	RCH	0.827	0.664	0.713	0.739	0.756	0.765	0.772
RCH	TVT	0.973	0.899	0.923	0.935	0.942	0.946	0.95
RCH	CON	0.995	0.992	0.992	0.993	0.993	0.994	0.994
RCH	SNP	0.986	0.953	0.965	0.97	0.973	0.976	0.977
RCH	LAB1	0.905	0.732	0.789	0.816	0.836	0.845	0.853
RCH	LAB2	0.897	0.757	0.803	0.826	0.841	0.849	0.855
RCH	B60	0.957	0.861	0.896	0.916	0.925	0.93	0.933
RCH	BRX	0.999	0.997	0.998	0.998	0.999	0.999	0.999
RCH	ASE	0.999	0.997	0.998	0.998	0.998	0.998	0.998
RCH	SNPO	0.999	0.996	0.997	0.998	0.998	0.998	0.998
TVT	TVT	0.934	0.837	0.835	0.837	0.845	0.85	0.856
TVT	CON	0.999	0.997	0.997	0.998	0.998	0.998	0.998
TVT	SNP	0.976	0.894	0.908	0.925	0.933	0.939	0.944
TVT	LAB1	0.977	0.888	0.917	0.931	0.94	0.945	0.949
TVT	LAB2	0.974	0.899	0.919	0.932	0.94	0.945	0.948
TVT	B60	0.983	0.923	0.938	0.951	0.957	0.961	0.963
TVT	BRX	0.999	0.998	0.998	0.998	0.998	0.998	0.999
TVT	ASE	0.998	0.996	0.996	0.996	0.996	0.997	0.997
TVT	SNPO	0.999	0.997	0.997	0.997	0.997	0.998	0.998
CON	CON	0.778	0.583	0.628	0.66	0.679	0.691	0.701
CON	SNP	0.999	0.998	0.998	0.999	0.999	0.999	0.999
CON	LAB1	0.998	0.994	0.995	0.995	0.996	0.996	0.996
CON	LAB2	0.997	0.994	0.995	0.996	0.996	0.996	0.996
CON	B60	0.999	0.996	0.997	0.998	0.998	0.998	0.998
CON	BRX	0.986	0.975	0.976	0.977	0.978	0.978	0.979
CON	ASE	0.98	0.961	0.964	0.965	0.967	0.968	0.969
CON	SNPO	0.958	0.894	0.912	0.922	0.928	0.933	0.935
SNP	SNP	0.908	0.697	0.749	0.777	0.796	0.81	0.821
SNP	LAB1	0.985	0.935	0.951	0.959	0.965	0.968	0.97
SNP	LAB2	0.983	0.941	0.954	0.961	0.966	0.969	0.971
SNP	B60	0.988	0.947	0.96	0.969	0.973	0.975	0.977
SNP	BRX	0.999	0.998	0.998	0.998	0.999	0.999	0.999
SNP	ASE	0.999	0.998	0.998	0.998	0.999	0.999	0.999
SNP	SNPO	0.999	0.997	0.998	0.998	0.998	0.998	0.998
LAB1	LAB1	0.875	0.655	0.708	0.739	0.763	0.776	0.787
LAB1	LAB2	0.917	0.754	0.805	0.83	0.848	0.858	0.865
LAB1	B60	0.955	0.841	0.877	0.9	0.911	0.918	0.922
LAB1	BRX	0.999	0.997	0.998	0.998	0.999	0.999	0.999
LAB1	ASE	0.999	0.998	0.998	0.998	0.998	0.999	0.999
LAB1	SNPO	0.999	0.996	0.997	0.998	0.998	0.998	0.998
LAB2	LAB2	0.891	0.749	0.791	0.811	0.826	0.834	0.84
LAB2	B60	0.965	0.877	0.908	0.926	0.935	0.94	0.942
LAB2	BRX	0.999	0.997	0.998	0.999	0.999	0.999	0.999
LAB2	ASE	0.999	0.998	0.998	0.998	0.999	0.999	0.999
LAB2	SNPO	0.999	0.996	0.997	0.998	0.998	0.998	0.998
B60	B60	0.918	0.821	0.841	0.859	0.869	0.877	0.88
B60	BRX	0.999	0.998	0.999	0.999	0.999	0.999	0.999
B60	ASE	1	0.998	0.999	0.999	0.999	0.999	0.999
B60	SNPO	0.999	0.997	0.998	0.999	0.999	0.999	0.999
BRX	BRX	0.89	0.639	0.706	0.743	0.768	0.781	0.799
BRX	ASE	0.986	0.984	0.979	0.977	0.977	0.977	0.978
BRX	SNPO	0.985	0.978	0.974	0.973	0.973	0.974	0.975
ASE	ASE	0.868	0.683	0.735	0.759	0.775	0.79	0.799
ASE	SNPO	0.985	0.973	0.973	0.973	0.973	0.974	0.975
SNPO	SNPO	0.909	0.786	0.801	0.817	0.831	0.842	0.849

Table A2: Average distance between communities using different subsets of tweets. Distances are computed as the average cosine distance among communities' stationary superspreaders members, represented as the vectors of retweets made that are present in the subset. *Third* = subset of tweets retweeted by at least 1/3 of communities' stationary members. Legend: same community, similar communities, dissimilar communities.