4 Supplementary Information

4.1 Training

To accelerate the training process, we implemented an undersampling of negative records, i.e. tweets that were viewed but not engaged with by a specific participant. We ensure the inclusion of negative records from the same session as positive ones. Additionally, we aimed to balance the representation of tweets from friends and non-friends, as well as from authors of engaged tweets. To enhance the diversity of our training dataset, we incorporated all impressions collected during Twitter searches, as these searches are expected to provide a more varied set of tweets compared to home timelines.

After an extensive exploratory data analysis, we designed a parsimonious set of features and performed a systematic hyperparameter tuning procedure. The hyperparameter tuning was conducted using equal time budgets and the Optuna "stepwise algorithm"[53], which efficiently explores the hyperparameter space. In particular, we explored the possibility of incorporating a wider history by considering the distance in the follower space between the authors of the tweet to be recommended and the last five authors that the users liked or retweeted. We also explored the integration of text embeddings, such as the cosine similarity between the tweet's text embedding and previously liked or retweeted tweets. However, we found that these additional features did not significantly improve the predictive power of the models compared to the computational cost involved. Future work focused on explainability rather than solely maximizing accuracy may consider incorporating these features for the additional insights they may provide. Finally, we settled for the set of features as a trade-off between parsimony and predictive power disaplyed in Table 1. We provide in the Figure 5, the importance of each features in the models predicting the like and the retweet, in terms of number of times the features are used in the LightGBM models.

Table 1. Set of features

Feature Category	Features names and description
Impression Related	'nb_min_since_publication': Number of minutes elapsed since the tweet was published.
Tweets Related	'tw_nb_characters': Number of characters in the tweet. 'tw_nb_words': Number of words in the tweet. 'tw_mean_length_words': Mean length of words in the tweet. 'tw_nb_hashtag': Number of hashtags in the tweet. 'tw_nb_urls': Number of URLs included in the tweet. 'tw_nb_mentions': Number of user mentions in the tweet.
Author Related	'author_created_days_ago': Number of days since the author's Twitter account was created. 'author_created_years_ago': Number of years since the author's Twitter account was created. 'author_followers_count': Number of followers the author has. 'author_friends_count': Number of accounts the author follows. 'author_listed_count': Number of public lists the author is a part of. 'author_statuses_count': Number of tweets the author has posted. 'author_followers_count_rate': Number of followers divided by the number of days since account creation. 'author_friends_count_rate': Number of friends divided by the number of days since account creation. 'author_listed_count_rate': Number of list the authors is a part of divided by the number of days since account creation. 'author_statuses_count_rate': Number of tweet posted by the author divided by the number of days since account creation. 'author_default_profile_image': Binary indicator representing whether the author has a default profile image. 'author_verified': Binary indicator representing whether the author's Twitter account is verified.
Relation to authors	'subjectFollowsAuthor': Binary indicator representing whether the subject follows the author. 'authorSubjectJaccard': Jaccard similarity coefficient between the author and subject's Twitter friends. 'authorSubjectOverlapCoef': Overlap coefficient between the author and subject's Twitter friends.
Relation with past engagement	'author_pagerank_ratio_previously_rt': Ratio of PageRanks between the author of message and the last retweeted author 'author_pagerank_ratio_previously_like':Ratio of PageRanks between the author of message and the last liked author 'author_reduced_l2_previously_like': L2 norm between the author of message and the last liked author (in 8D follow space) 'author_reduced_cosine_sim_previously_like': Cosine similarity between the author of message and the last liked author 'author_reduced_l2_previously_retweet': L2 norm between the author of message and the last liked author 'author_reduced_cosine_sim_previously_retweet': Cosine similarity between the author of message and the last retweeted author

4.2 Login hour

The analysis of Twitter usage patterns has been performed leveraging our data donation browser add-on. Our simulation incorporates the specific distribution for each day, see Figure 6. While we acknowledge that Twitter usage extends beyond browser platforms and includes mobile devices, leveraging this distribution provides the most precise and comprehensive representation available to us.

4.3 Political leaning assignation

As explained in the main text, the political orientations were estimated using the Politoscope database, which includes 711 million French political tweets since 2016. To enhance the reliability of the analysis, we only retained edges with a minimum weight of 2. In other words, an edge was included between two accounts only if at least two political-related retweets had occurred between them in 2022.

Firstly, we set the opinion values of the far-left and far-right leaders to ± 0.75, chosen arbitrarily. Next, the opinion of the leader with a centrist leaning is determined by averaging the opinions of the two extreme leaders weighted by the angular similarities between the nodes' embeddings; determined via node2vec[27]. Interestingly, this calculation yields an opinion value close to zero (-0.02). For each Twitter account, we calculate the angular similarity between the account's embedding and the embeddings of the three leaders. The political leaning of the account is then determined as the average opinion of the two closest leaders, weighted by their angular similarities. In cases where the two closest leaders are the extreme ones, we account for the periodicity of the opinion space, ensuring that the assigned opinion spans the entire range from -1 to +1. Also, we only assign a political leaning if the angular similarity with the closest leader is at least 10% higher than the similarity with the farther away leader. Accounts without a clear political leaning, make up less than 10% of the accounts in our database.

We conducted experiments to ensure the stability of the resulting opinion scale when different anchors were chosen, including both the selected leaders and the arbitrary assigned opinions. Moreover, we observed a correspondence between the opinion scale and a cluster analysis of the retweet graph[36]. Additionally, the political groups declared by French members of Parliament align with the opinion scale and the assessment made by Political expert[37], as shown in Figures 11 & 12. To visually represent the political landscape, Figure 10 presents a spatialized representation of the retweet graph, where nodes are color-coded according to their assigned numerical opinion, facilitating a clear interpretation of the political landscape.

4.4 Noisy predictions

To explore a more diverse range of recommender systems, we introduced noise into the predictions generated by our predictive models. Specifically, the recommender system would present the user with the top L tweets based on their engagement score, which corresponds to the probability of receiving a "like". Gaussian noise was then added to this engagement score. We focused on "likes" as they represent the most common form of engagement, allowing us to analyze the evolution of the metrics introduced based on average precision (evaluated on the test dataset), rather than the level of noise.

As illustrated in Figures 14a, 14b, and 14c, as the average precision increases, the distortion metrics also increase.

4.5 Graph of follow

In order to gain further insight into the structure captured by node2vec embeddings, we conducted an analysis involving half a million randomly selected accounts from the graph. We computed the pairwise Euclidean distance within three different dimensional spaces: the original 64-dimensional space produced by node2vec, as well as the reduced 8-dimensional and 2-dimensional spaces obtained through PaCMAP[28] reduction. The results, depicted in Figure 15, reveal that the pairwise distance of embeddings for accounts sharing a connection (i.e. at least one account follows the other) is significantly lower compared to pairs of accounts without such a connection. This behavior holds true across all three investigated dimensions.

Furthermore, we calculated the Jaccard index for each pair of accounts, measuring the similarity of the accounts they follow. Through the computation of the Spearman's rank correlation coefficient, we observed that the distance between two accounts in the embedding space is negatively correlated with the Jaccard index. This pattern persisted in 64 original node2vec space, with a Spearman's $\rho = -0.79$, as well as in the reduced dimensions with values of $\rho = -0.70$ (in 2D) and $\rho = -0.72$ (in 3D) (all $p-$values $< 10^{-16}$). These findings confirm that node2vec embeddings effectively capture the similarity between account neighborhoods.

The observed neighborhood similarity can be interpreted as homophily of interests, where accounts that produce similar content form communities in the graph of follow[2]. Figure 16 illustrates the latent space of followers in two dimensions, clustered using HDBSCAN[28], with the most popular accounts within the largest clusters labeled. Notably, we observe various interest-based communities related to sports, US politics, cryptocurrencies, gaming, etc. Also, Twitter accounts within each cluster tend to share the same language. The predominant language in account descriptions represents over 77.3% of the accounts description within a cluster, the second most popular language accounts for 17.9%; over the whole dataset the two most popular language accounts for 47.6% and 37.4% respectively of the accounts descriptions.

Lastly, we computed the cosine similarity between tweet text embeddings (computed by TwHIN-BERT[55], a language model for multilingual tweet representations) authored either by accounts within or outside a cluster. As depicted in Figure 17, the tweets authored by accounts within the same cluster are more similar to each other than to those authored by accounts outside the cluster.

Overall, the embedding of the graph of follow effectively captures the (dis)similarities between accounts neighborhoods, embedding closely accounts that follow similar accounts. To some extent, these homophily of neighborhood can be linked to homophily of interest. As presented in[2], the user-user graph of follow is the foundation of Twitter SimClusters community-based recommendation algorithms.

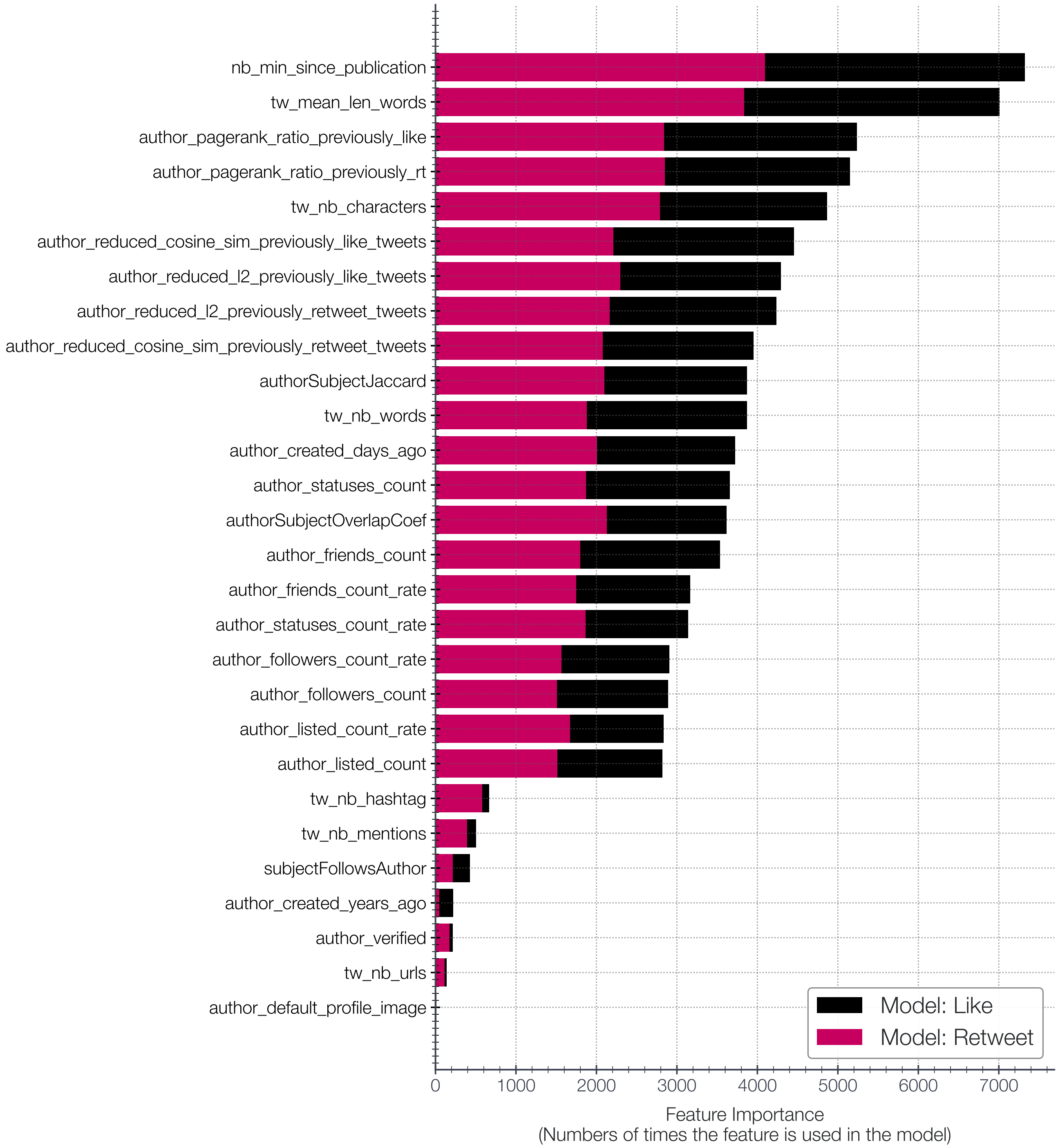


Figure 5. Features importance

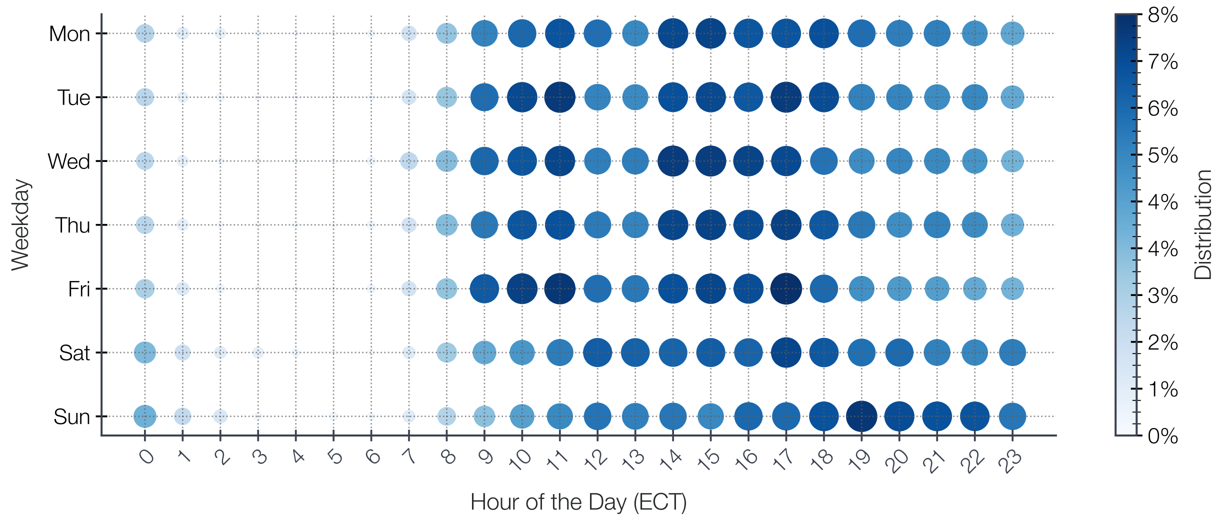
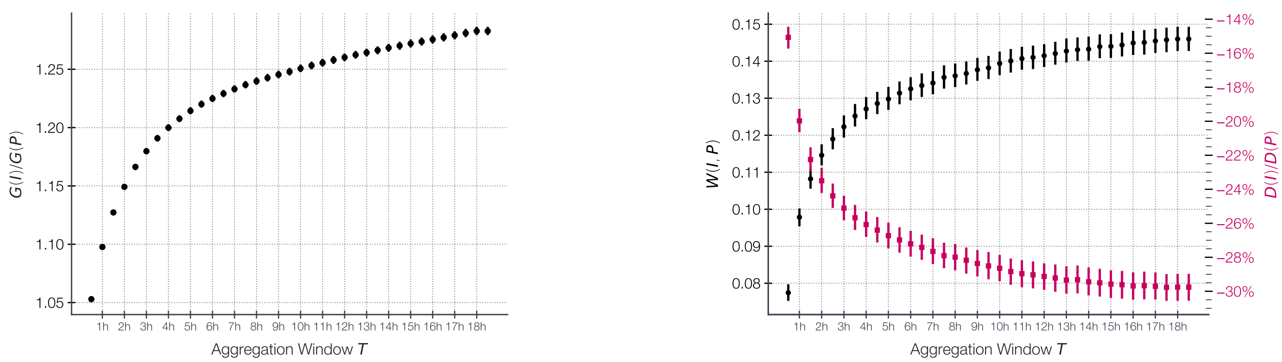


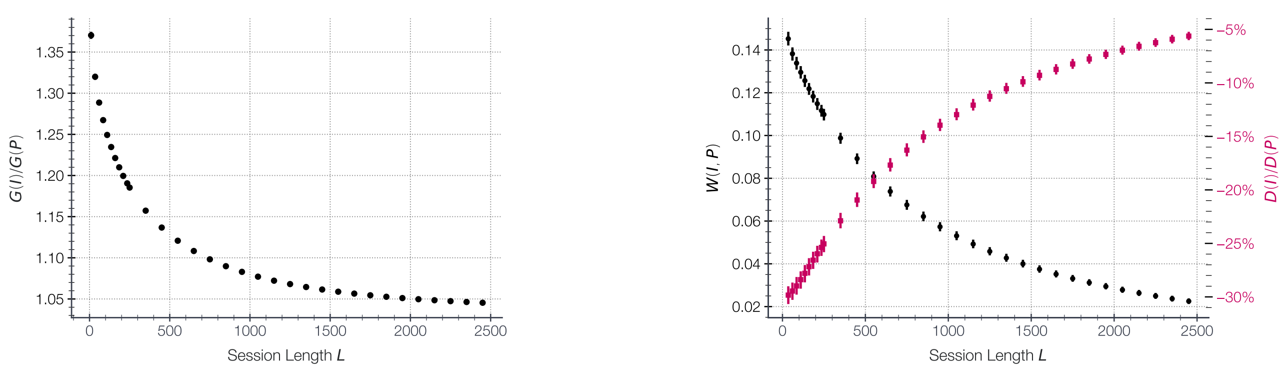
Figure 6. Distribution of hour of the day at which users log-in Twitter, as captured by our data donation desktop browser add-on.



(a) Ratio $G(I)/G(P)$

(b) $W(I,P)$ and $D(I)/D(P)$

Figure 7. Metrics as a function of the aggregation window T, (error bars represent 95% confidence intervals, determined via bootstrapping over users)



(a) Ratio $G(I)/G(P)$

(b) $W(I,P)$ and $D(I)/D(P)$

Figure 8. Metrics as a function of the session length L, (error bars represent 95% confidence intervals, determined via bootstrapping over users).

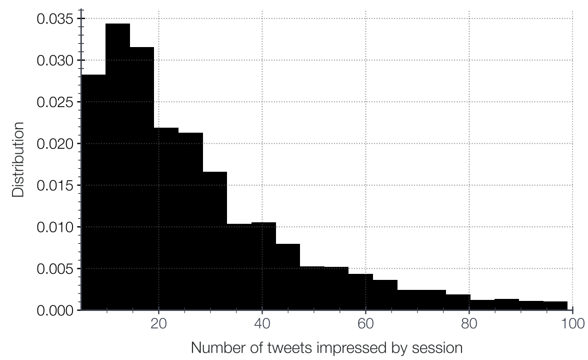


Figure 9. Distribution of session length (number of impressed tweets) as measured by our browser add-on. We truncated the distribution at 100 tweets, 2.6% of the session are in the truncated tails.

Figure 10. Graph of retweets associated to political messages published in 2022. Spatialized via ForceAtlas2[54]. Nodes are colored by the assigned numerical opinion.

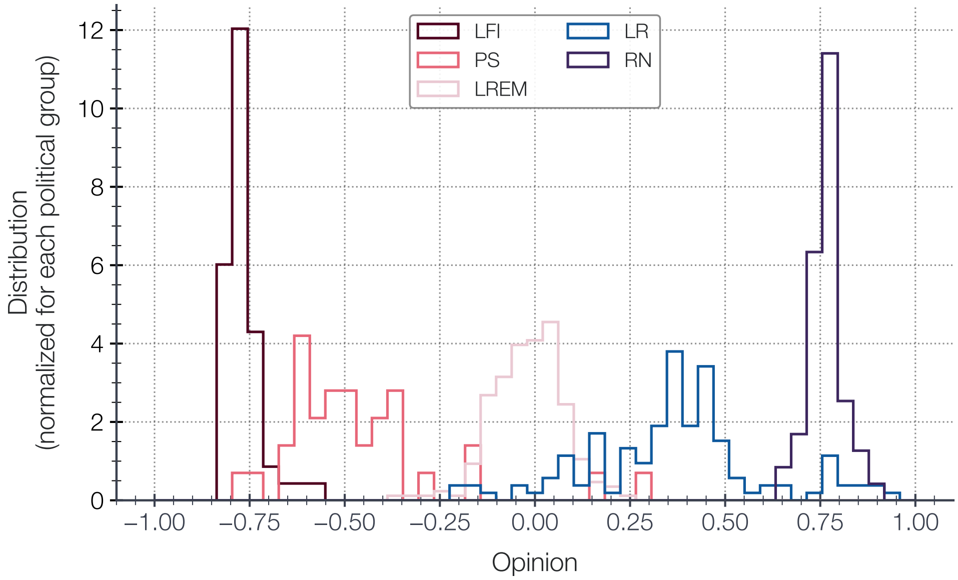


Figure 11. Distribution of the assigned numerical opinion for the member of French Parliament depending of their declared political group. Main french political groups from left to right: "LFI", "PS","LREM", "LR", "RN"

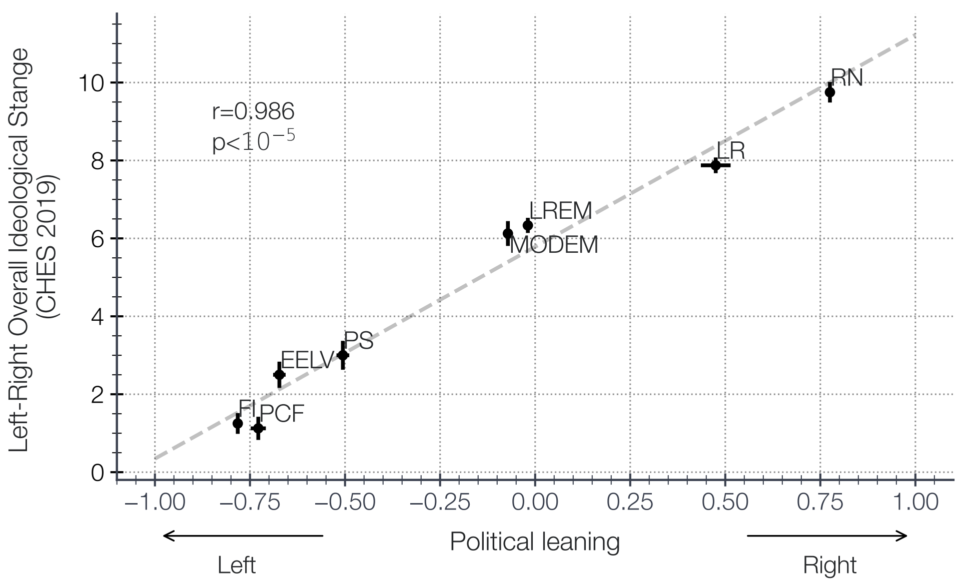


Figure 12. Position on the opinion scale French political group, averaged over their MPs, compared to the estimation made by political experts within the 2019 Chapel Hill Expert Survey[37]; error bars represent bootstrap standard errors over the MPs.

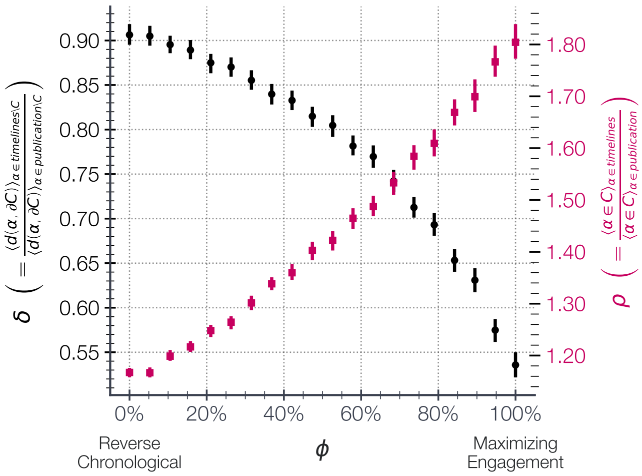
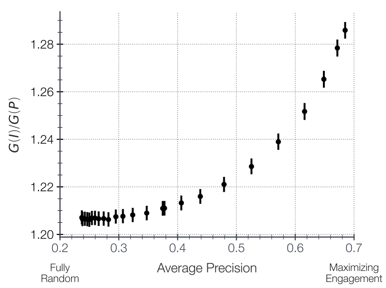
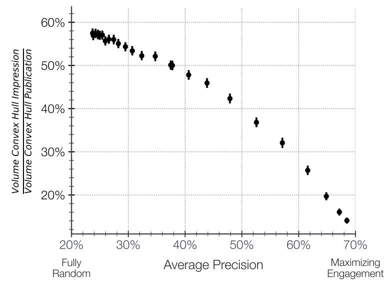


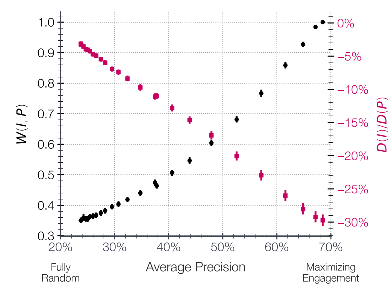
Figure 13. Latent unexpectedness δ (in black dots) and ρ (in red squares) with respect to ϕ, in 3D, (error bars represent 95% confidence intervals, determined via bootstrapping over users)



(a) Ratio $G(I)/G(P)$



(b) Convex hull volumes ratio



(c) Distortion of the political landscape

Figure 14. Metrics as a function of the average precision of the "like" predictive model, (error bars represent 95% confidence intervals, determined via bootstrapping over users)

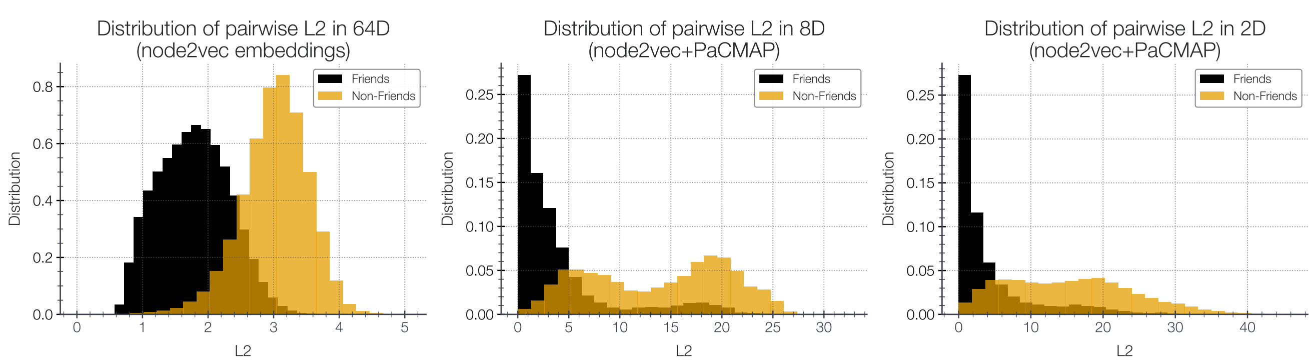


Figure 15. Distribution of the euclidean distance between random pairs of accounts, segmenting pairs in which at least one account follow the other.

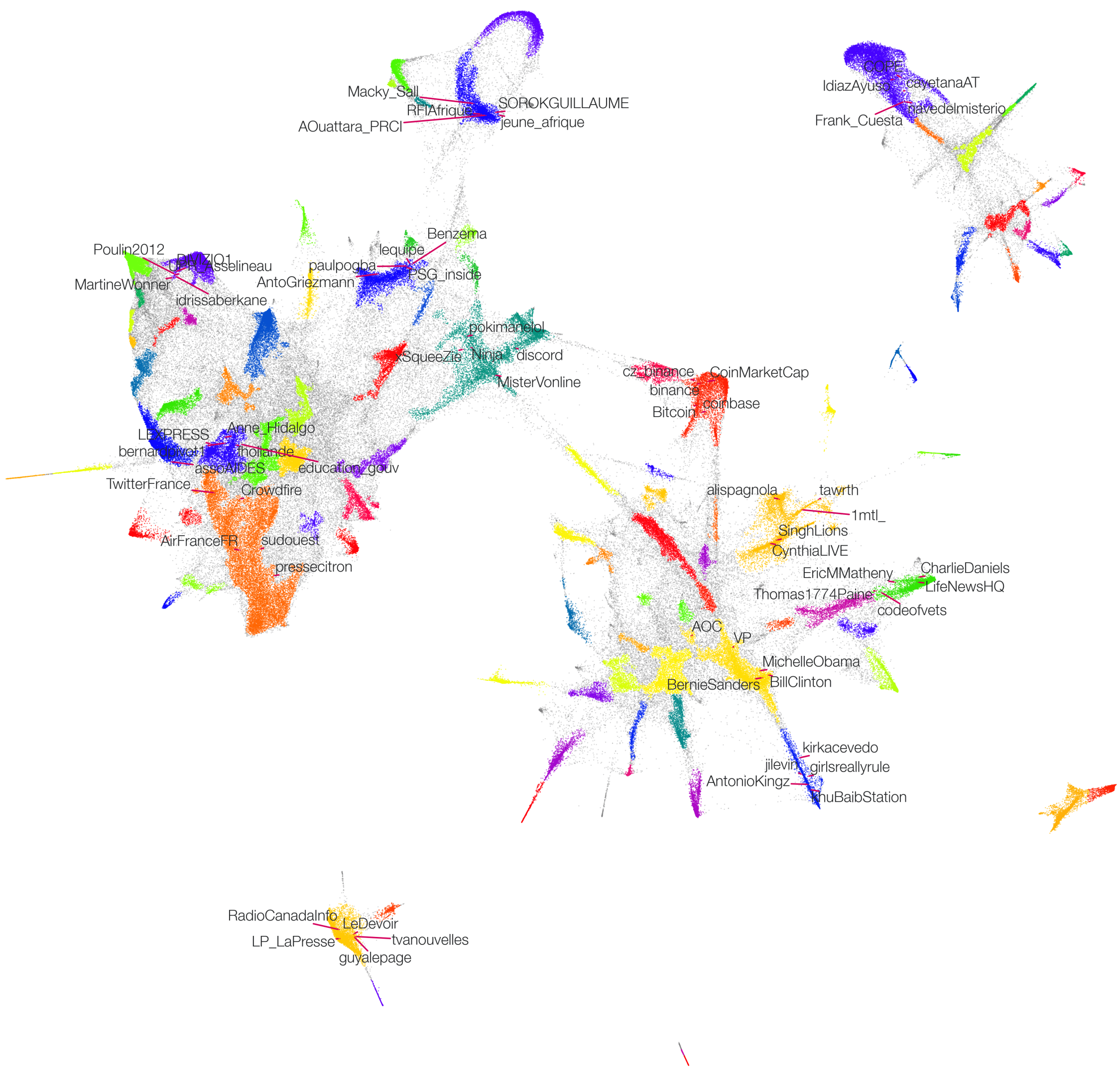


Figure 16. Latent follow space, clustered using HDBSCAN. In grey are displayed outliers, belonging to no identified clusters. For each of the 16 larger cluster we label the most popular accounts.

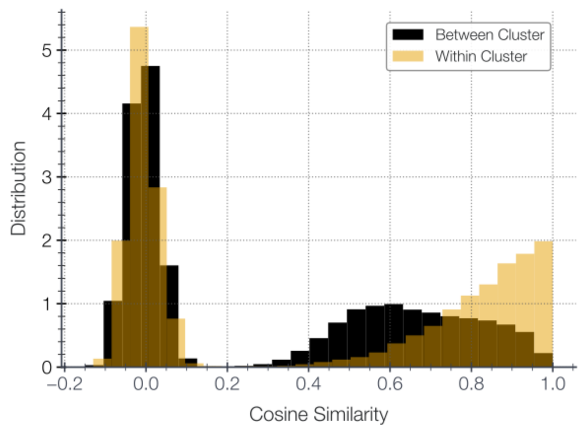


Figure 17. Distribution of the cosine similiarty between the TwHIN-BERT embeddings of tweets published by members of either the same cluster or different clusters.