



OPEN Leveraging content producer networks and user perception to detect online discursive communities

Stefano Guarino^{1,8}✉, Ayoub Mounim², Guido Caldarelli^{3,4,5} & Fabio Saracco^{1,6,7,8}

Online discussions are often characterized by strong behavioral asymmetries: a relatively small fraction of users actively produces content, while the majority primarily consumes and redistributes it. Here we propose a community-detection framework for online social networks that exploits this asymmetry by first identifying and clustering a set of leading users, and then extending the resulting labels to the broader user base. We introduce two complementary strategies to cluster leaders, one based on their mutual interactions and the other on audience overlap, both relying on entropy-based filtering to separate signal from noise. We evaluate the framework on three major Italian political debates on Twitter/X, using public figures—identified through the pre-2022 verification system—as leaders, and known affiliations of political actors as ground truth labels. Compared with standard baselines, the proposed approach yields more coherent and interpretable communities aligned with political structures, with the two variants respectively recovering parties and coalitions. Activity-based criteria for selecting leaders produce qualitatively similar but consistently weaker results, particularly at the coalition level. Overall, our findings show that creating statistically validated networks of publicly recognized figures, whose off-platform roles constrain and stabilize their online behavior, provides a strong basis to identify discursive communities on social media. Although developed for Twitter/X, the approach is conceptually general, as it leverages structural asymmetries common to many online platforms.

Online social platforms have become crucial arenas for public discourse, with communities forming organically around shared narratives, values, and ideological affinities^{1–3}. While these discursive communities can offer solidarity and foster public engagement^{4–6}, they also risk becoming echo chambers—insulated environments that reinforce existing beliefs and suppress dissenting perspectives^{7–13}. Understanding the structural dynamics that govern these communities is therefore essential to evaluating how information circulates and how opinions crystallize in digital spaces.

Traditional approaches to community detection in social networks have evolved significantly, incorporating network topology, account metadata, and behavioral signals. For instance, some methods enhance classical algorithms by integrating account characteristics¹⁴, while others acknowledge that platforms like Twitter/X function less as egalitarian many-to-many networks, and more as hierarchies dominated by a small number of highly visible content creators^{15–17}. Within this structure, influence is disproportionately concentrated, making the identification of these “elite” accounts central to the analysis of online discourse^{18–20}.

Building on this body of work, the first contribution of this article is the definition of a general framework for detecting online discursive communities, formalizing the idea that online debates are driven by a relatively small set of users who actively produce content. Because content production entails greater exposure and commitment—especially in political discussions—these users tend to adopt more consistent and publicly interpretable positions. Once content producers are identified, the affiliation of less active users can be inferred from their interaction patterns, even when individual behavior is sparse.

The proposed framework operationalizes this rationale in a relatively simple way. We partition the user set U into two disjoint subsets, $U = T \cup \perp$, cluster the content creators (the set T) by applying a standard community-

¹National Research Council of Italy, Institute for Applied Mathematics, Rome, Italy. ²Data Lab, LUISS “Guido Carli” University, Rome, Italy. ³National Research Council of Italy, Institute of Complex Systems, Rome, Italy. ⁴Ca’ Foscari University of Venice, Venice, Italy. ⁵London Institute for Mathematical Sciences, London, UK. ⁶“Enrico Fermi” Research Center, Rome, Italy. ⁷IMT School for Advanced Studies Lucca, Lucca, Italy. ⁸Stefano Guarino and Fabio Saracco contributed equally to this work. ✉email: stefano.guarino@cnr.it

detection algorithm to an appropriately defined network $G = (\mathbb{T}, E)$, and then extend the resulting labels to users in $\perp$ based on their interactions with $\mathbb{T}$.

A key modeling choice, however, concerns the construction of the edge set E , which defines how relationships among content creators are represented. We consider two alternatives: one based on statistically significant interactions directly observed among users in $\mathbb{T}$, and one based on a validated projection of the bipartite network linking content creators in $\mathbb{T}$ to their audiences in $\perp$. These two variants define *MonoDC_T* and *Bipartite-network-induced Discursive Communities* (BiDC_T), respectively. Both methods rely on maximum-entropy null models²¹ to statistically validate inferred connections and to filter out noise—a well-known challenge in the analysis of online social networks^{18,22}. Statistical filtering plays a limited role in MonoDC_T, but is essential in BiDC_T to remove spurious correlations induced by heterogeneous audience activity.

Beyond the construction of E , the framework admits different criteria to identify the set $\mathbb{T}$ of content creators. From this perspective, the second contribution of this article is less methodological and more substantive: we show that, although social-media relevance can be acquired in an anonymous and platform-driven manner, ideological communities in political debates remain predominantly anchored to publicly recognizable political figures.

Accordingly, we focus on Twitter's verified accounts ($\mathbb{T} = V$), which, at the time covered by our datasets, identified public figures whose activity on the platform was closely tied to their off-platform roles. All datasets analyzed precede the introduction of paid verification in November 2022, and therefore correspond to this original, institutionally grounded meaning of verification²³.

We compare this choice with two activity-based criteria derived from the data. We consider *influential* users ($\mathbb{T} = I$) as defined by⁴, based on follower structure and mention activity, and *high-h-index* users ($\mathbb{T} = H$), obtained by adapting the Hirsch index²⁴ to retweet dynamics. These data-driven selections serve as baselines to evaluate the added value of anchoring the analysis on publicly recognized figures.

We analyze three datasets covering major political events in Italy in 2022, where discursive communities are expected to mirror political alignments. When evaluated against the known affiliations of clearly identifiable political actors, MonoDC_V captures party-level structures, whereas BiDC_V is more effective at identifying coalitional configurations. Standard community-detection algorithms applied to the complete retweet network yield noisier and less interpretable partitions, while activity-based criteria perform better than these baselines but remain less reliable than anchoring the analysis on verified accounts, especially at the coalition level.

We argue that this is not an *effect* of verification, but, again, a consequence of the fact that, historically, only publicly recognizable figures were “verifiable”, a requisite that in turn constrained and stabilized their online behavior. In this sense, our results show that, even in environments where visibility can be gained in largely anonymous and platform-driven ways, ideological communities in political debates remain primarily structured around actors endowed with public recognizability beyond the platform. By foregrounding this mechanism, we propose and analyze a framework that, while developed for Twitter/X, is conceptually general, as it leverages asymmetries between content producers and audiences that characterize many contemporary online platforms.

Finally, we show that a carefully selected subset of users and interactions captures a large fraction of the information needed to reconstruct political communities, offering practical insights for designing efficient data-collection strategies under increasingly restrictive platform access conditions.

Related work

A substantial body of research has investigated the structure of the online discourse on Twitter and other social media platforms by analyzing retweet networks, information cascades, and statistically validated interaction patterns (e.g.,^{18,25–27}). These studies have consistently shown, in particular, that political debates on social media exhibit strong polarization and modular structure, often aligned with ideological or partisan divides.

A central methodological challenge in this literature concerns the high level of noise and heterogeneity in user activity. To address this issue, several works have adopted statistical filtering techniques based on null models to extract significant interaction patterns from bipartite or projected networks. In particular, maximum-entropy null models have been widely used to preserve local constraints—such as node degrees—while filtering out correlations that can be explained by random activity alone^{21,28}. These approaches have been shown to uncover meaningful mesoscale structures in retweet, mention, and user-content networks, enabling the identification of latent ideological alignment from interaction data^{13,18–20,29}.

Within this line of work, the Italian political context has received special attention. Becatti et al.¹⁸ have shown that entropy-based filtering improves the interpretability and stability of political communities compared to unfiltered network representations (the method proposed therein is called BiDC_V in the present manuscript). Related studies have further demonstrated that statistically validated user-hashtag bipartite projections can reveal well-defined discursive communities during national electoral campaigns, even in the absence of explicit ideological labels²⁰.

A parallel strand of research has focused on the role of *influential* or *central* users in shaping the online discourse. Early work has defined influence primarily in terms of activity-based or structural measures, such as the number of followers, retweets, mentions, or centrality in interaction networks^{15–17}. Other studies have explored the diffusion power of opinion leaders and elites in online debates, often highlighting the disproportionate impact of a small fraction of highly visible users³⁰.

Despite these advances, two important gaps persist in the literature. First, most community-detection approaches implicitly treat users symmetrically, without explicitly accounting for the strong asymmetry between content producers and content consumers that characterizes participation on social media—in general, and especially for political debates. Second, benchmarks for evaluating communities are often based on internal network metrics, such as modularity or stability, rather than on externally validated ground truth labels. As

a result, it remains unclear (i) whether institutional signals of authority—such as pre-2022 verified status on Twitter—provide particularly informative anchors for community detection, and (ii) how alternative criteria for identifying leading users affect the resulting mappings when embedded into statistically validated pipelines. The present work addresses these gaps by proposing a leader-based framework that explicitly separates content creators from their audiences and systematically compares different strategies for selecting politically relevant producers of content.

Methods

We investigate how online discursive communities form around publicly relevant users on Twitter/X by analyzing large-scale interaction data. While the framework we propose is general, we focus empirically on three major political events in Italy during 2022, a domain in which online debates tend to be especially structured and externally grounded. Our approach builds on the observation that social-media debates are typically characterized by strong asymmetries in user behavior: a relatively small subset of users actively produces salient content, whereas the majority primarily consumes and redistributes it. This asymmetry is particularly pronounced in political discussions, where public figures and institutional actors play a central role. To exploit this structure, we adopt a leader-based framework in which community detection is first performed on a selected set of content creators and subsequently extended to the broader user base.

Detection of political discursive communities

To identify meaningful structures in online political debates, it is necessary to distinguish statistically significant interaction patterns from correlations induced by heterogeneous user activity and content popularity. To this end, our approach relies on maximum-entropy null models as a principled statistical baseline for validating inferred network connections before applying community-detection algorithms.

Within this framework, an ensemble of random networks is defined by maximizing the Shannon entropy subject to a set of local constraints that reproduce selected properties of the observed data, such as degree sequences or activity levels^{21,31,32}. Empirical links are evaluated against their expected occurrence under the null model, and only statistically significant deviations are retained. Model parameters are estimated by maximizing the likelihood of the observed network³³, and statistical significance is assessed using false discovery rate correction³⁴.

This maximum-entropy approach is grounded in information theory and statistical mechanics and has proven effective across a wide range of networked systems^{13,18–20,22,29,35–40}. In the following, we introduce two complementary methods that instantiate this framework in different ways, depending on how interactions among selected content creators are inferred. Further methodological details are reported in Section 1 of Supplementary Information.

In the analyses presented here, we focus on verified accounts as a concrete instantiation of content creators, as—prior to the introduction of paid verification—they largely corresponded to publicly recognized figures. However, the methods described below are general and can be applied to any set $\mathcal{T}$ of users identified as debate leaders.

Construction of the MonoDC_V network

The first method, MonoDC_V (Monopartite-network-induced Discursive Communities; Fig. 1), focuses on retweet interactions among verified users. We start by constructing a bipartite directed network between users and tweets, where directed edges represent authorship and retweet actions. At this stage, we consider all verified users together with all users who retweeted at least one verified account in the dataset under study. This structure is then projected onto a weighted directed user-user network, in which a link from user i to user j encodes the number of times i retweeted content authored by j .

To separate meaningful interaction patterns from noise—an intrinsic feature of online social activity—we apply a maximum-entropy null model tailored to bipartite directed networks. The model uses the observed posting and retweeting activity of users as constraints while randomizing all other aspects, providing a principled

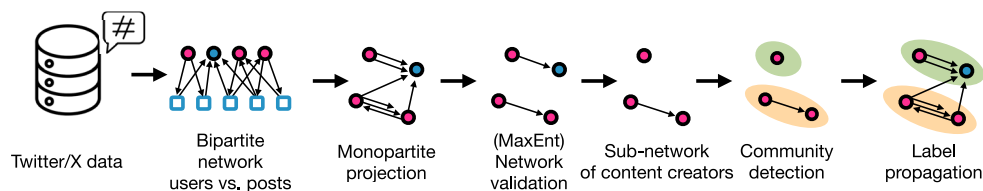


Fig. 1. MonoDC's pipeline. Starting from the data from Twitter/X, we build a bipartite directed network between accounts and posts; a directed arrow from a user (a circle) to a post (a square) represents authorship, while in the opposite direction represents a retweet. Then, the content of the bipartite network is projected into a monopartite directed network of users. Such a network is further statistically validated using the expectations of a maximum entropy null model for bipartite directed networks. Then, the sub-network relative to the selected content creators (magenta circles) is analysed, and communities are detected using standard algorithms. Finally, the so-obtained labels are propagated to standard users (blue circles) through a label propagation algorithm. The icon representing Twitter/X data is obtained by superimposing the icons “Data Base #7994662” by Farrih Icon and “Hashtag #4315467” by Alex Burte, both from Noun Projects.

statistical baseline. For each pair of users i, j , we compute the probability of the observed retweeting weight under the null model, and retain the edge only if its p-value is statistically significant (more details are present in Section 1 of Supplementary Information).

Community detection is then performed on the subnetwork of the validated network induced by verified users. The presented results have been obtained with a modularity-based algorithm, but we verified that alternative algorithms yield essentially the same results. Finally, the detected community labels are propagated to non-verified users using a standard label-propagation scheme⁴¹.

Construction of the BiDC_V network

The second method, BiDC_V (Bipartite-network-induced Discursive Communities; Fig. 2¹⁸), infers relationships among verified users indirectly, based on the similarity of their audiences. Here, we construct a bipartite undirected network connecting verified users to all users who retweeted them at least once during the event.

This bipartite structure preserves the heterogeneity between content creators and their audiences. To extract meaningful similarities between verified users, we apply a maximum-entropy null model for bipartite undirected networks, which accounts for differences in audience sizes and engagement patterns. Statistically significant similarities are retained, yielding a validated projection onto a network of verified users.

Community detection is performed on this projected network, and the resulting labels are propagated to non-verified users according to the composition of their retweet activity. As in MonoDC_V, the propagation step allows us to assign community memberships to the entire user set starting from a limited number of leader nodes.

Evaluation

We evaluate the quality of the detected communities using the V-measure (VM_{β})⁴², an information-theoretic metric that quantifies the agreement between two partitions based on their mutual information. The V-measure is well suited to our setting, as it remains interpretable across different levels of partition granularity and explicitly balances *homogeneity*—how well each cluster contains only members of a single class—and *completeness*—whether all members of a given class are assigned to the same cluster—through the parameter β . For $\beta = 1$, VM_1 reduces to the Variation of Information⁴³, normalized by the total entropy of the partitions. By varying β over $[0, +\infty)$, we account for both finer-grained and more aggregated political structures, such as political movements/trends within parties or political coalitions. More details can be found in Section 2 of Supplementary Information.

As a baseline, we also apply standard community-detection algorithms directly to the complete retweet network. This network is defined as a directed, weighted graph including all active users, where edges represent the number of retweets between pairs of users. Unlike MonoDC_V and BiDC_V, this representation is not statistically filtered and serves to isolate the contribution of entropy-based validation and leader selection.

Although MonoDC_V and BiDC_V ultimately rely on Louvain for final partitioning, the comparison with Louvain applied to the complete retweet network serves a different purpose. In our frameworks, Louvain operates *after* statistically validated filtering, and *with* an informative initialization provided by verified seeds. Running Louvain directly on the raw network therefore represents a baseline that isolates the contribution of our preprocessing and seed selection. This allows us to quantify the extent to which discursive communities emerge from the raw interaction structure alone versus from statistically validated and leader-anchored representations.

Data

We analyze three datasets covering the Italian presidential election, a government crisis, and the 2022 general elections. Each dataset consists of millions of tweets and hundreds of thousands of users, collected via Twitter's Academic Application Programming Interface (API) using event-specific keywords; more details about the queries used for the data collection can be found in Section 3 of Supplementary Information. Table 1 summarizes the main characteristics of the datasets.

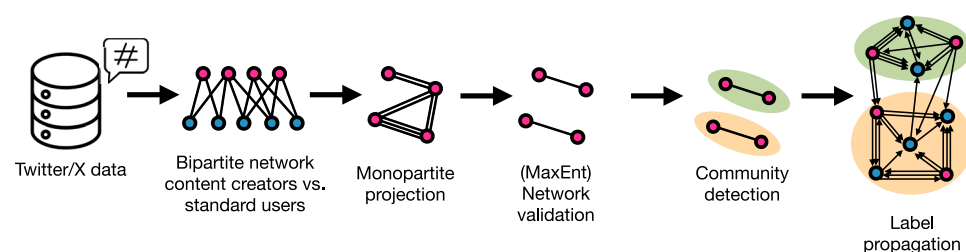


Fig. 2. BiDC's pipeline. Starting from the data from Twitter/X, we build a bipartite undirected network between the accounts of the content creators and the one of standard users; a link is present if the standard users retweeted the given content creator at least once. Then, the information of the bipartite network is projected into a monopartite network of content creators. Such a network is further statistically validated using the expectations of a maximum entropy null model for bipartite undirected networks. Finally, the so-obtained labels are propagated to standard users through a label propagation algorithm. The procedure was first proposed in¹⁸. The icon representing Twitter/X data is obtained by superimposing the icons “Data Base #7994662” by Farrih Icon and “Hashtag #4315467” by Alex Burte, both from Noun Projects.

Dataset	Unique users	Unique tweets	Annotated users	Verified users	Influential users	H-index ≥ 3 users
President ^a	119,018	711,497	292	1148	5804	1296
Crisis ^b	113,876	1,414,181	229	1004	7038	2554
Elections ^c	222,356	2,784,951	239	1555	12184	4382

Table 1. A summary of the datasets considered in the present paper. ^a Debate around the election of the President of the Republic. The political debate ended with the re-election of Sergio Mattarella on January 29, 2022. ^b Debate around the political crisis that led to the fall of the Draghi government on July 21, 2022. ^c Debate around the Italian general elections held on September 25, 2022, that led to Giorgia Meloni becoming the new Prime Minister.

Dataset	Nodes in GCC	Links	Q(parties)	Q(coalitions)
President	169	219	0.763	0.618
Crisis	145	246	0.721	0.641
Elections	154	355	0.777	0.649

Table 2. Retweet networks of annotated users and modularity (Q) of the partition induced by the manual annotation at the party and coalition level.

Among the verified accounts appearing in each dataset, we manually annotated those corresponding to politically active public figures. Annotations were performed at both the party and coalition level and provide an external ground truth against which we assess the performance of our methods.

Manual annotation of verified accounts

For each event, we manually annotated the political affiliation of verified users to construct the ground truth used in our evaluation. Only verified accounts of political parties and politicians were annotated; accounts belonging to journalists, media outlets, institutions, or entertainers were excluded. Each annotated user was assigned both a *party label* and a *coalition label*. The party label was determined based on publicly available information from the user’s Twitter profile and Wikipedia page, or from institutional sources such as party websites, parliamentary records and candidate lists. The coalition label was assigned according to the official coalitions recognized during the 2022 Italian electoral cycle.

Annotations were performed by one of the authors, while the remaining authors conducted random spot checks to assess consistency and correctness. Verified accounts with no clear political alignment were included in the construction of both MonoDC_V and BiDC_V networks, but were excluded from the ground truth and used only as unlabelled nodes in the evaluation.

Results

Verified accounts provide a strong political signal under the pre-2022 verification regime

To assess the effectiveness of our framework, we focused on three political events that sparked intense debate on Italian Twitter in 2022: the re-election of the President of the Republic, the resignation of Prime Minister Draghi, and the general elections in September. Crucially, these datasets were compiled before the platform introduced paid verification, ensuring that verified accounts retained their original role as signals of institutional recognition rather than self-declared status.

We manually annotated a subset of verified accounts (between 229 and 292 per event) based on their publicly identifiable political affiliations. When restricting the retweet networks to, first, the Greatest Connected Component (GCC), and then to annotated users and evaluating the partition induced by their known party and coalition labels, we observed highly modular structures. Modularity values exceeded 0.6 in all cases (cfr. Table 2), indicating that retweet interactions among verified users are strongly structured along political lines.

These results show that, even without any statistical validation, retweet interactions among the annotated verified users already separate clearly along party and coalition lines, indicating that well-defined political communities exist within this subset.

To extrapolate this benchmark to the broader network, we propagated community labels from verified users to the rest of the user population using a label propagation algorithm⁴¹. While these propagated labels are inherently noisier than manual annotations, they provide a principled proxy for large-scale political alignment and enable a systematic comparison of community-detection strategies over the full dataset.

Standard community detection on the full retweet network shows limited political alignment

We applied four widely used community detection algorithms—Louvain, label propagation, Infomap, and stochastic block model inference—to the complete retweet networks, defined as the directed weighted graph including all users active during the event and all observed retweet interactions. While all methods identified

non-trivial community structure, their agreement with the manual annotations was limited. V-measure scores for party affiliation ranged between 0.4 and 0.6 (cfr. Fig. 3), with substantial variability across algorithm runs.

Comparisons between algorithms further revealed significant inconsistency, with pairwise V-measures for the same dataset lying in the [0.3, 0.7] range (cfr. Fig. 4 in Supplementary Information). Qualitative inspection confirmed that standard methods often conflated ideologically distinct actors or fragmented coherent political blocs.

These results do not indicate a general failure of standard algorithms, but rather highlight the difficulty of recovering politically meaningful communities when they are applied directly to noisy retweet networks without distinguishing between heterogeneous user roles or filtering statistically expected interactions.

Leader-anchored and statistically validated networks improve community detection

Applying our two proposed methods, MonoDC_V and BiDC_V, resulted in a substantial improvement in community-detection performance. When verified users were used as anchors within statistically validated networks, both methods produced partitions closely aligned with the annotated political affiliations.

MonoDC_V achieved particularly high accuracy at the party level, reflecting the focused retweet patterns through which politicians tend to amplify content from their own party leadership. By contrast, BiDC_V more effectively captured coalition-level structures, as it relies on shared audience behaviour rather than direct interactions among leaders.

Quantitatively, V-measure scores for both methods systematically exceeded those of standard algorithms across all datasets. MonoDC_V reached values above 0.75 for party-level annotations, while BiDC_V achieved comparably strong results at the coalition level (cfr. Fig. 4). These findings remained robust across different values of the resolution parameter β (cfr. Fig. 5) and when considering propagated labels on the full network (cfr. Fig. 5 in the Supplementary Information).

As a baseline, we apply the Louvain algorithm to the complete retweet network. This representation includes neither statistical filtering nor a distinction between verified and non-verified users, and thus reflects the raw interaction structure dominated by highly heterogeneous activity patterns.

Although MonoDC_V and BiDC_V also rely on Louvain for the final partitioning step, the comparison serves a different purpose. In our framework, Louvain is applied after maximum-entropy filtering and on networks anchored to a coherent set of political leaders. Running Louvain directly on the raw retweet network therefore isolates the contribution of preprocessing and leader selection, rather than acting as a competing methodological alternative. Notably, the Louvain algorithm performs no worse than the others in recovering the annotations (cfr. Fig. 3).

Overall, these results indicate that verified accounts function as effective anchors for political discursive communities *within a statistically validated framework*. The improved performance does not stem from verified status alone, but from its combination with maximum-entropy filtering and with the institutional coherence of the verified-user set under the pre-2022 verification regime.

MonoDC captures party structure, while BiDC reveals coalitional organization

The complementary strengths of MonoDC_V and BiDC_V arise from the nature of their respective network constructions. MonoDC_V relies on direct retweet interactions among verified users, which tend to encode endorsement and intra-party signalling. As a result, its statistically validated networks retain a large fraction of edges (94–97%), indicating dense and coherent communication among political actors.

BiDC_V, by contrast, infers similarity among leaders through audience overlap. After statistical validation, only 7–12% of edges are retained, reflecting the fact that only a small subset of shared audiences carries meaningful political alignment. This filtering allows BiDC_V to surface stable coalition-level structures, particularly in phases where electoral alliances are clearly defined.

Consistently, the VM_{β} between the partitions produced by the two methods shows that MonoDC_V communities tend to be subclusters of those identified by BiDC_V, especially for the Crisis and Elections datasets (cfr. Fig. 6).

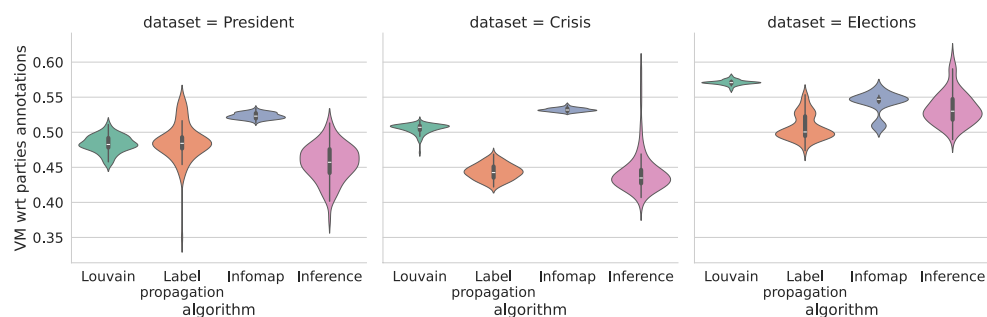


Fig. 3. Limits of standard methods. The VM_1 between partitions obtained with off-the-shelf algorithms, restricted to the annotated users, and the parties' annotations. Each violin represents the distribution over 100 independent runs. Standard methods show a limited accuracy in recognizing the affiliation of renowned politicians.

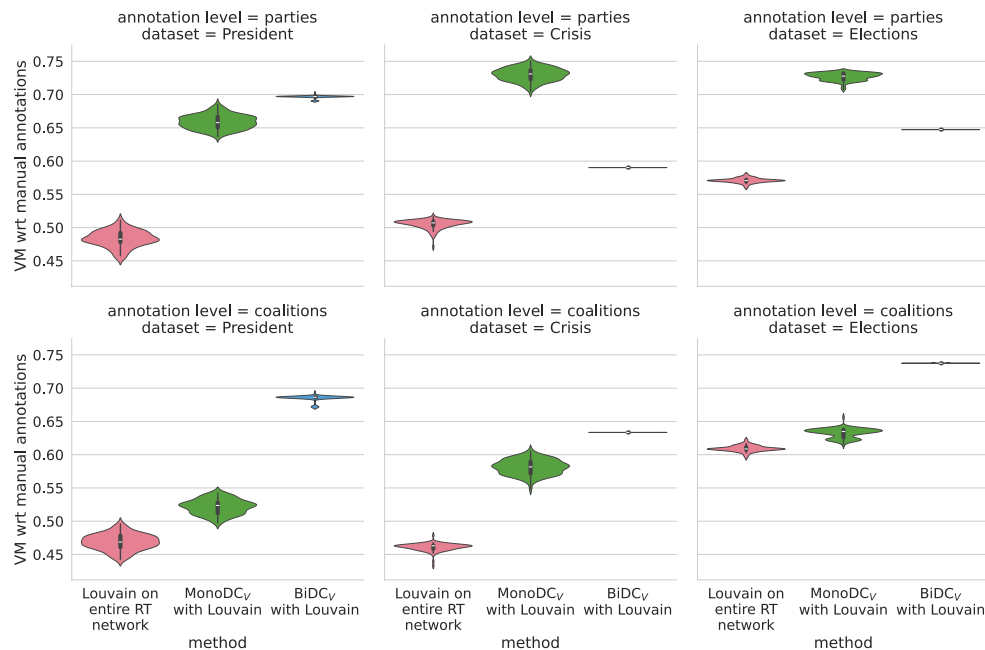


Fig. 4. Comparison of the performances of MonoDC_V and BiDC_V with standard methods. The VM₁ between partitions of the annotated users, obtained with different methods, and the annotations at party (first row) and coalition (second row) level. Each violin represents the distribution over 100 independent runs, and we use the Louvain algorithm as a benchmark because it performs no worse than other algorithms (cfr. Fig. 3). MonoDC_V and BiDC_V systematically outperform Louvain, with MonoDC_V generally preferable in identifying parties while BiDC_V in identifying coalitions.

A qualitative comparison further illustrates these differences (cfr. Fig. 7). MonoDC_V yields fine-grained partitions aligned with individual parties, whereas BiDC_V aggregates these into broader coalitional structures. In contrast, Louvain applied to the full retweet network produces a noisier mapping that conflates party- and coalition-level affiliations.

We attribute this fragmentation not to political complexity, but to the structural noise of the raw retweet network, which is substantially reduced by maximum-entropy filtering.

Activity-based definitions of leaders yield weaker political partitions

To assess the robustness of our findings, we replaced verified users with alternative activity-based definitions of the leader set $\mathcal{T}$. In the case $\mathcal{T} = I$, we selected influential users following⁴, while $\mathcal{T} = H$ was defined using an adaptation of the Hirsch index²⁴ to retweet activity: a user u 's H-index in the debate is the largest h such that at least h tweets authored by u received at least h retweets each, and we define H as the set of users having H-index ≥ 3 (cfr. Fig. 2 in Supplementary Information for a justification of the threshold).

In all cases, performance deteriorated relative to anchoring communities on verified users (cfr. Fig. 8). While activity-based criteria identify highly active accounts, their discursive behavior is more heterogeneous and less predictive of political alignment. In particular, H-index-based selections include journalists and commentators whose audiences span ideological boundaries, blurring community structure.

These results highlight that the effectiveness of our framework depends not only on selecting active users, but on identifying content creators whose publicly recognized roles constrain and stabilize their online behavior. Further comparison can be found in Section 4 of Supplementary Information.

Discussion

Digital platforms play an increasingly influential role in shaping the structure and visibility of public discourse. One crucial contribution of this work is the characterization of a general methodology for extracting meaningful discursive communities from noisy social data. We introduced a leader-based framework that explicitly accounts for the asymmetric structure of online participation. By separating content producers from their audiences, our approach allows communities to emerge more clearly than methods that treat all users symmetrically. We showed that identifying and positioning content producers is a necessary first step, after which the broader user base can be reliably assigned.

Both proposed methods—MonoDC and BiDC—filter user interactions through maximum entropy null models, thereby revealing latent organizational patterns otherwise obscured by platform dynamics or behavioral heterogeneity. In the specific case considered here, we applied this framework to datasets centered on highly politicized debates, where the expected discursive structures correspond to political parties and electoral coalitions. Within this context, the distinction between the two approaches offers analytical flexibility:

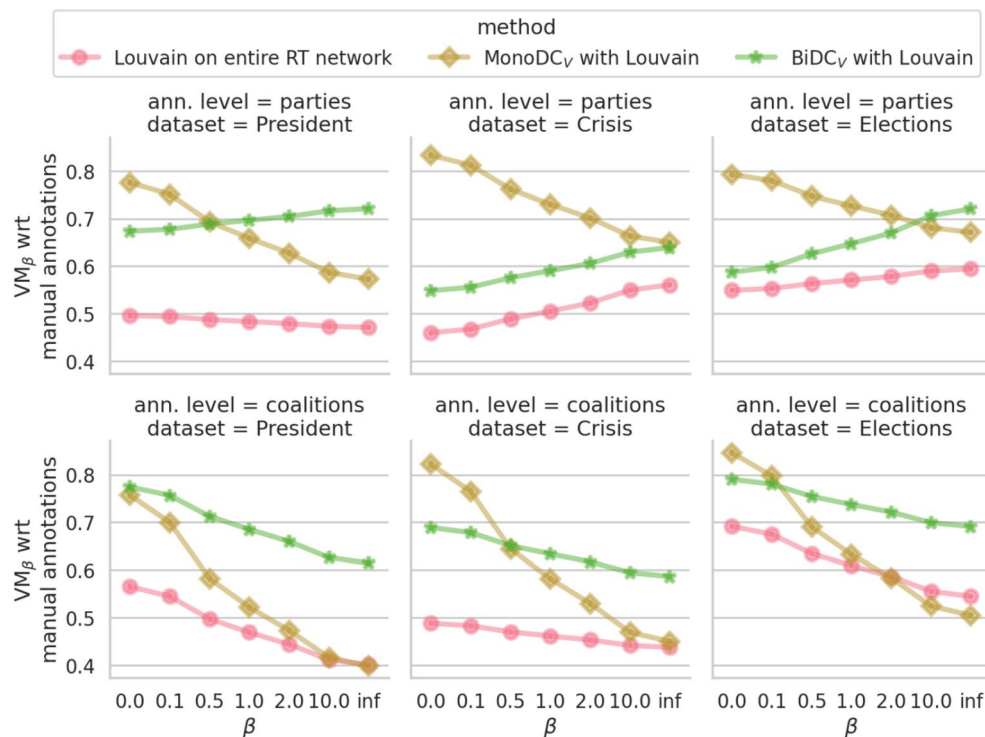


Fig. 5. Comparison of the performances of MonoDC_V and BiDC_V with standard methods at different resolution levels. The VM_{β} between partitions of the annotated users, obtained with different methods, and the annotations, for different values of β and both at the parties (top) and coalitions (bottom) level. Each point is the average over 100 runs of the method indicated by the colour and marker. Analogous results are obtained when considering the propagated annotations. When compared with manually annotated parties (top) and coalitions (bottom), MonoDC_V and BiDC_V exhibit a consistently greater average VM_{β} than standard methods even for $\beta \neq 1$, i.e. at essentially all scales at which the political spectrum can be observed. In this sense, standard algorithms, when executed on the entire retweet network, do not return sub- or super-groups of the political parties that MonoDC_V and BiDC_V cannot detect.

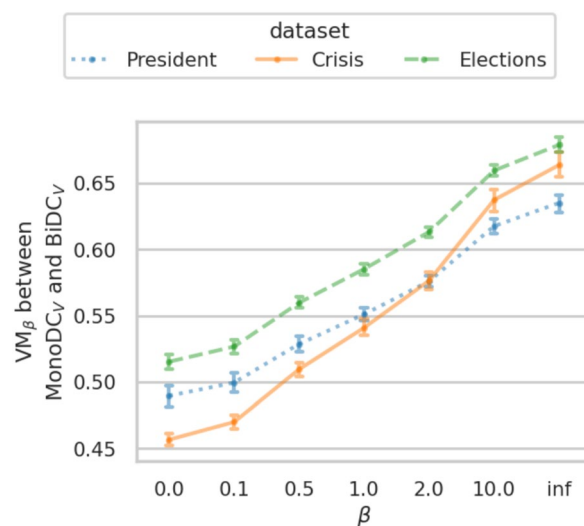


Fig. 6. Different resolution of MonoDC_V and BiDC_V. The VM_{β} , for variable β , between partitions of all users obtained running MonoDC_V and BiDC_V. The clusters found by MonoDC_V are subclusters of those found by BiDC_V, at least in part, and especially for the two datasets, Crisis and Elections, which refer to a political phase in which the coalitions were more clearly defined.

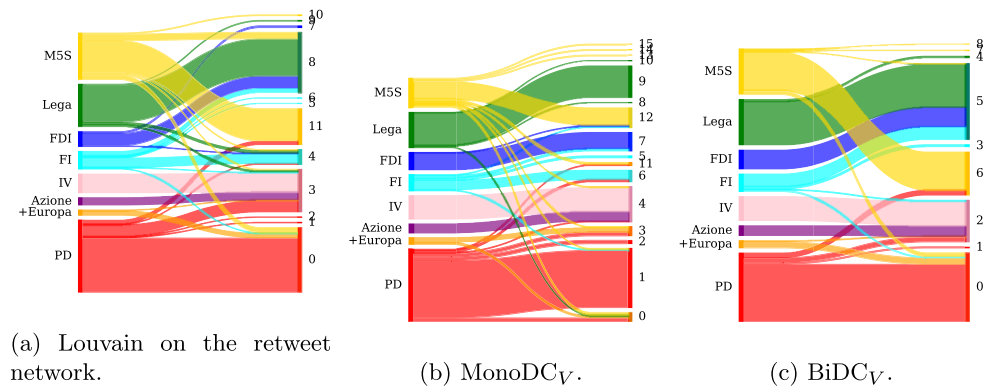


Fig. 7. Qualitative comparison of MonoDC_V and BiDC_V with Louvain on the Elections dataset. For each of the considered methods, the Sankey diagrams show the composition of the extracted communities in terms of political parties. MonoDC_V very accurately detects parties (Azione and Italia Viva ran together as a single list at the elections), plus some political movements/trends within parties; BiDC_V almost perfectly detects the official political coalitions of those elections; Louvain, executed on the entire retweet network, provides a more noisy classification that is a mix of the two levels, with Forza Italia separated from its coalitions and many politicians contributing to communities led by some of their rivals.

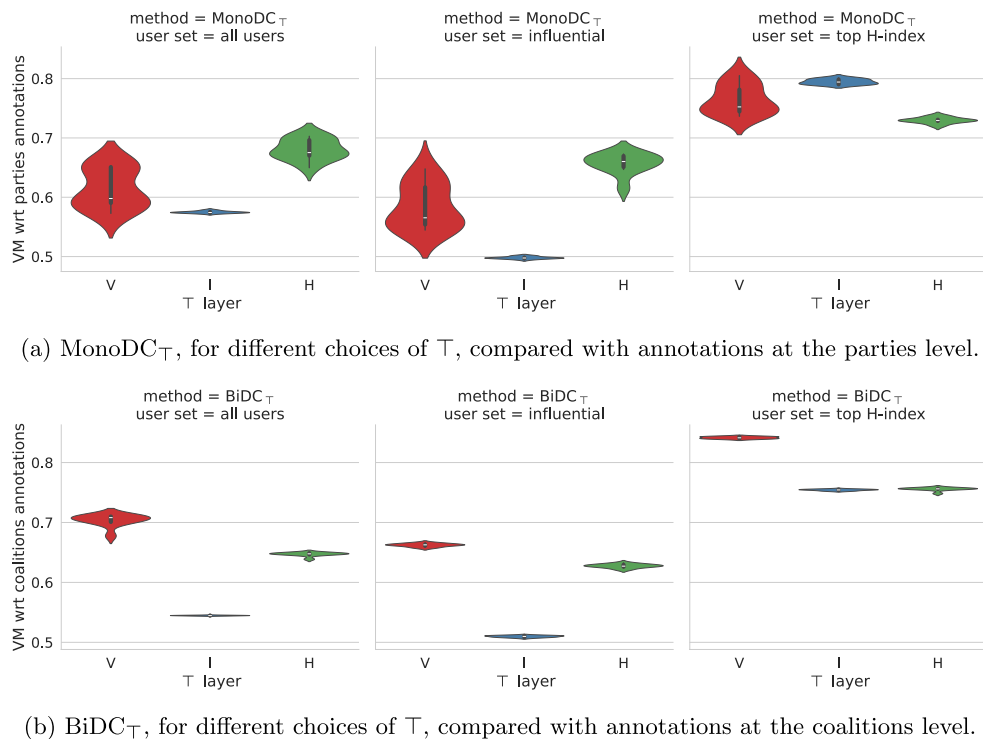


Fig. 8. Comparison of the partitions obtained with MonoDC_T and BiDC_T, for different choices of T . The VM_1 between partitions obtained with different choices of T and the annotations, for the Elections dataset. For each method, we used the level of annotation that the method more accurately reproduces, that is, parties for MonoDC_T (a) and coalitions for BiDC_T (b). The columns refer to the user set upon which the VM is computed. For instance, “user set = influential” means that in the second column we only consider how influential users have been labeled by different methods, temporarily ignoring the rest of the network.

MonoDC emphasizes direct political signaling through retweets, while BiDC captures broader audience-based alignments, including coalition-level structures.

In addition, through a systematic comparison of three alternative criteria for selecting leading users, our findings demonstrate that institutional status, rather than raw activity, provides the strongest signal for organizing online debates. We found, in fact, that pre-2022 verified accounts represent the most informative anchors for

community detection. We interpret this result as a consequence of their public recognizability and the behavioral constraints imposed by their off-platform roles, rather than as an intrinsic property of verification itself.

While the institutional meaning of verification has changed following the introduction of paid subscriptions, the methodological insights of this study remain valid. In particular, our results suggest that carefully curated sets of publicly relevant actors—identified either through institutional roles or external information—can serve as effective seeds for analyzing online discourse, at least in the well-studied and highly structured context of political debates. This insight opens the door to more efficient data-collection strategies, in which the ego networks of a limited number of high-relevance users are sufficient to recover much of the political structure of online debates, even under restrictive platform access conditions. Indeed, similar strategies have been implemented, for instance in⁴⁴. More broadly, our study reinforces the idea that the organization of digital public debate potentially depends not only on individual user behavior, but also on how platforms encode and signal authority. While the methods proposed here are technical, they contribute to a growing body of work showing that discursive structure, legitimacy, and visibility are deeply intertwined in online public spheres.

As any other product of research, our analyses have some limitations. First, our empirical results pertain exclusively to the pre-2022 verification regime, under which verified accounts corresponded to publicly recognized figures and formed coherent, politically aligned communities. In this historical setting, verification acted as a reliable proxy for public relevance and institutional embeddedness, providing a stable anchor for discursive structures. The transition to paid verification introduces a qualitatively different regime, whose effects fall outside the scope of the present analysis. Concerns about potential distortions of online discourse—such as increased visibility for actors lacking public accountability or the erosion of previously coherent community structures—should therefore be interpreted as speculative and grounded in plausible mechanisms rather than in direct empirical evidence. Whether political discourse under the new regime will remain similarly structured, fragment into less interpretable patterns, or reorganize around alternative signals of authority remains an open empirical question. Future research should explicitly address this gap by analyzing data collected after the introduction of paid verification, enabling direct comparisons between the two regimes. Such studies could assess whether the structural role once played by verified users has weakened, persisted, or been replaced by other indicators of public relevance.

Second, our empirical analysis focuses on political debates within a single national context, Italy, and on a single social media platform, Twitter/X. This choice reflects both data availability and methodological considerations: analyzing a political system that is well known to the authors allows for more reliable manual annotations and a more accurate interpretation of political actors, coalitions, and discursive dynamics. Importantly, this focus does not represent an intrinsic limitation of the proposed framework, which is grounded in structural features of online discourse—such as strong participation asymmetries and the presence of discussion leaders—that have been widely documented across platforms, languages, and cultural settings^{15,17,45}. Consistently with this perspective, previous studies have already shown that BiDC performs well in different countries^{22,29,46}, while the present work provides the first systematic investigation of MonoDC and a direct comparison between alternative leader-based constructions. Moreover, the proposed approach does not rely on platform-specific features, but rather on general interaction patterns between content producers and their audiences, making its application to other platforms—where visibility and influence may be structured through different mechanisms—a natural direction for future research.

Data availability

Preprocessed data and code used for the analysis are available on GitLab at https://gitlab.com/s.guarino/dico_analysis.

Received: 18 August 2025; Accepted: 5 February 2026

Published online: 03 March 2026

References

- Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F. & Flammini, A. Political polarization on Twitter. In *Proceedings of the International Aaai Conference on Web and Social Media*. Vol. 5. 89–96 (2011)
- Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A. & Menczer, F. Predicting the political alignment of twitter users. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. 192–199 (2011)
- Conover, M. D., Gonçalves, B., Flammini, A. & Menczer, F. Partisan asymmetries in online political activity. *EPJ Data Sci.* <https://doi.org/10.1140/epjds6> (2012).
- González-Bailón, S., Borge-Holthoefer, J. & Moreno, Y. Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.* <https://doi.org/10.1177/0002764213479371> (2013).
- Andalibi, N., Haimson, O.L., Choudhury, M.D. & Forte, A. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Trans. Comput.-Hum. Interact.* **25**(5). <https://doi.org/10.1145/3234942> (2018).
- Morini, V. et al. Online posting effects: Unveiling the non-linear journeys of users in depression communities on reddit. *Comput. Hum. Behav. Rep.* **17**, 100542. <https://doi.org/10.1016/j.chbr.2024.100542> (2025).
- Jamieson, K. & Cappella, J. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment* (Oxford University Press, 2008).
- Garrett, R. K. Echo chambers online?: Politically motivated selective exposure among internet news users. *J. Comput.-Mediat. Commun.* **14**, 265–285. <https://doi.org/10.1111/j.1083-6101.2009.01440.X> (2009).
- Del Vicario, M. et al. The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**(3), 554–559. <https://doi.org/10.1073/pnas.1517441113> (2016).
- Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G. & Shekhtman, L. Debunking in a world of tribes. *PLoS ONE* **7**(12), e0181821 (2017)
- Villa, G., Pasi, G. & Viviani, M. Echo chamber detection and analysis: A topology- and content-based approach in the covid-19 scenario. *Soc. Netw. Anal. Min.* **11**(1). <https://doi.org/10.1007/s13278-021-00779-3> (2021).

12. Morini, V., Pollacci, L. & Rossetti, G. Toward a standard approach for echo chamber detection: Reddit case study. *Appl. Sci.* **11**(12), 5390. <https://doi.org/10.3390/app11125390> (2021).
13. Pratelli, M., Saracco, F. & Petrocchi, M. Entropy-based detection of twitter echo chambers. *PNAS Nexus* **3**(5). <https://doi.org/10.1093/pnasnexus/pgae177> (2024).
14. Citraro, S. & Rossetti, G. Eva: Attribute-aware network segmentation. In *International Conference on Complex Networks and Their Applications*. 141–151 (Springer, 2019).
15. Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *J. Consum. Res.* **34**, 441–458. <https://doi.org/10.1086/518527> (2007).
16. Wu, S., Hofman, J.M., Mason, W.A. & Watts, D.J. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web. WWW '11*. 705–714. <https://doi.org/10.1145/1963405.1963504>. (Association for Computing Machinery, 2011).
17. Hilbert, M., Vázquez, J., Halpern, D., Valenzuela, S. & Arriagada, E. One step, two step, network step? Complementary perspectives on communication flows in twittered citizen protests. *Soc. Sci. Comput. Rev.* **35**(4), 444–461. <https://doi.org/10.1177/0894439316639561> (2017).
18. Becatti, C., Caldarelli, G., Lambiotte, R. & Saracco, F. Extracting significant signal of news consumption from social networks: The case of twitter in Italian political elections. *Palgrave Commun.* **5**, 1–16. <https://doi.org/10.1057/s41599-019-0300-3> (2019).
19. Caldarelli, G. et al. The role of bot squads in the political propaganda on twitter. *Commun. Phys.* **3**, 1–15. <https://doi.org/10.1038/s42005-020-0340-4> (2020).
20. Radicioni, T., Saracco, F., Pavan, E. & Squartini, T. Analysing twitter semantic networks: The case of 2018 Italian elections. *Sci. Rep.* **11**, 1–22. <https://doi.org/10.1038/s41598-021-92337-2> (2021).
21. Cimini, G. et al. The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**(1), 58–71. <https://doi.org/10.1038/s42254-018-0002-6> (2019) [arXiv:1810.05095](https://arxiv.org/abs/1810.05095) [physics.soc-ph].
22. De Clerck, B., Rocha, L. E. & Van Utterbeeck, F. Maximum entropy networks for large scale social network node analysis. *Appl. Netw. Sci.* **7**(1), 68 (2022).
23. Haman, M. & Školník, M. The unverified era: Politicians' twitter verification post-musk acquisition. *J. Inf. Technol. Politics* **22**(2), 167–171. <https://doi.org/10.1080/19331681.2023.2293868> (2023).
24. Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102> (2005).
25. Garimella, K., Morales, G.D.F., Gionis, A. & Mathioudakis, M. Quantifying controversy on social media. *Trans. Soc. Comput.* **1**(1). <https://doi.org/10.1145/3140565> (2018).
26. Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci.* **118**(9), 2023301118. <https://doi.org/10.1073/pnas.2023301118> (2021).
27. Lorenz-Spreen, P., Oswald, L., Lewandowsky, S. & Hertwig, R. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nat. Hum. Behav.* **7**, 74–101. <https://doi.org/10.1038/s41562-022-01460-1> (2023).
28. Saracco, F. et al. Inferring monopartite projections of bipartite networks: An entropy-based approach. *N. J. Phys.* <https://doi.org/10.1088/1367-2630/aa6b38> (2017).
29. Mattei, M., Pratelli, M., Caldarelli, G., Petrocchi, M. & Saracco, F. Bow-tie structures of twitter discursive communities. *Sci. Rep.* **12**, 1–23. <https://doi.org/10.1038/s41598-022-16603-7> (2022).
30. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. 65–74. <https://doi.org/10.1145/1935826.1935845> (2011).
31. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 181–218. <https://doi.org/10.1103/PhysRev.106.620> (1957).
32. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**(6), 66117. <https://doi.org/10.1103/PhysRevE.70.066117> (2004).
33. Garlaschelli, D. & Loffredo, M. I. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **78**, 1–5. <https://doi.org/10.1103/PhysRevE.78.015101> (2008).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300. <https://doi.org/10.2307/2346101> (1995).
35. Caldarelli, G., De Nicola, R., Petrocchi, M., Pratelli, M. & Saracco, F. Flow of online misinformation during the peak of the COVID-19 pandemic in Italy. *EPJ Data Sci.* **10**, 1–23. <https://doi.org/10.1140/EPJDS/S13688-021-00289-4> (2021).
36. Radicioni, T., Squartini, T., Pavan, E. & Saracco, F. Networked partisanship and framing: A socio-semantic network analysis of the Italian debate on migration. *PLOS ONE* **16**, 0256705. <https://doi.org/10.1371/JOURNAL.PONE.0256705> (2021).
37. Mattei, M., Caldarelli, G., Squartini, T. & Saracco, F. Italian twitter semantic network during the COVID-19 epidemic. *EPJ Data Sci.* **10**, 1–27. <https://doi.org/10.1140/EPJDS/S13688-021-00301-X> (2021).
38. Guarino, S., Pierri, F., Giovanni, M. D. & Celestini, A. Information disorders during the COVID-19 infodemic: The case of Italian Facebook. *Online Soc. Netw. Media* **22**, 100124. <https://doi.org/10.1016/J.OSNEM.2021.100124> (2021).
39. Bruno, M., Lambiotte, R. & Saracco, F. Brexit and bots: Characterizing the behaviour of automated accounts on twitter during the UK election. *EPJ Data Sci.* **11**, 1–24. <https://doi.org/10.1140/EPJDS/S13688-022-00330-0> (2022).
40. De Clerck, B., Van Utterbeeck, F., Petit, J., Lauwens, B., Mees, W. & Rocha, L.E. Maximum entropy networks applied on twitter disinformation datasets. In *Complex Networks & Their Applications X. Vol 2. Proceedings of the Tenth International Conference on Complex Networks and Their Applications* Complex Networks 2021. Vol. 10. 132–143 (Springer, 2022).
41. Raghavan, U.N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* <https://doi.org/10.1103/PhysRevE.76.036106> (2007).
42. Rosenberg, A. & Hirschberg, J. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. 410–420 (2007).
43. Meilă, M. Comparing clusterings—An information based distance. *J. Multivar. Anal.* **98**(5), 873–895. <https://doi.org/10.1016/j.jmva.2006.11.013> (2007).
44. Brugnoli, E., Gravino, P., Lo Sardo, D.R., Loreto, V. & Prevedello, G. Fine-grained clustering of social media: How moral triggers drive preferences and consensus. In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*. Vol. 3. AWA1. 1405–1412. <https://doi.org/10.5220/0012595000003636> (SciTePress, INSTICC, 2024).
45. Bracciale, R., Martella, A. & Visentin, C. From super-participants to super-echoed. Participation in the 2018 Italian electoral twittersphere. *PaCO Vol. 11*(2) (2018). In *Special Issue: From Big Data in Politics to the Politics of Big Data*. Vol. 11. 361–393. <https://doi.org/10.1285/I20356609V11I2P361> (2018).
46. De Clerck, B. et al. Detecting coordinated and bot-like behavior in twitter: The Jürgen Conings case. *EPJ Data Sci.* **13**(1), 40. <https://doi.org/10.1140/epjds/s13688-024-00477-y> (2024).

Acknowledgements

The authors are thankful to Lorenzo Federico for interesting discussions.

Author contributions

S.G. and F.S. conceived and designed the study. F.S. and A.M. collected the data. S.G., F.S. and A.M. performed the modeling and data analysis with input from G.C. S.G. and A.M. contributed to data preprocessing and visualization. S.G. and F.S. interpreted the results and drafted the manuscript. All authors contributed to revisions and approved the final version of the manuscript.

Funding

S.G. and F.S. were partially supported by the project “CODE – Coupling Opinion Dynamics with Epidemics”, funded under PNRR Mission 4 “Education and Research” - Component C2 - Investment 1.1 - Next Generation EU “Fund for National Research Program and Projects of Significant National Interest” PRIN 2022 PNRR, grant code P2022AKRZ9.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-39477-5>.

Correspondence and requests for materials should be addressed to S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026