Online appendix of Affording Fragmented Audiences: Multi-Platform Deliberation within the Five Star Movement

Francesco Marolla (Università degli Studi di Milano), Marilù Miotto (Erasmus University Rotterdam), C

2025-05-22

Ethical Approval

This study was approved by the REDC 2022.65, under the ethics approval code of the Tilburg University.

Platform Data Descriptives

This table presents descriptive statistics for each of the four main platforms included in our analysis—Beppe Grillo's Blog, Facebook, the M5S Forum, and Meetup. These figures provide important context for understanding the structural and functional differences across platforms that underpin our affordance-based analysis.

For each platform, and based on the dataset collected by the authors and used in the analysis, we report: the total number of unique users; the number of discussion threads (comprising opening posts and their associated comments); the total number of postings (including both initial posts and comments); the total number of organised events (for Meetup only); the date of the first post; and the average and median word count for both opening posts and comments.

Beppe Grillo's Blog, the first platform to go online in what would eventually become the Five Star Movement, stands out as the largest in terms of both unique users and volume of content. However, since user-generated content is limited to comments on Grillo's own posts, the number of discussion threads is comparatively low.

Notably, the average length of comments is similar on the Blog and the Forum, but significantly shorter on Facebook. This suggests differing levels of user engagement across platforms, possibly reflecting variations in how each platform structures participation and fosters deliberation.

Table 1: Descriptive statistics of the original data by platform.

	Blog	**Facebook**	**Forum**	**Meetup**
Number of Users	426,362	68,011	64,451	20,139
Number of Threads	5,607	68,011	24,097	—
Number of Postings	3,605,072	853,065	134,062	—
Number of Events	—	—	—	37,049
First posting	2005-01-16	2008-12-09	2009-02-11	2005-06-11
Opening post Words	Avg: 3951.83 Med: 2,176	Avg: 368.73 Med: 211	Avg: 1360.79 Med: 833	Avg: 755.31 Med: 479
Comment Words	Avg: 492.61 Med: 284	Avg: 164.28 Med: 84	Avg: 447.13 Med: 263	—

Cross-platform behaviour

This figure shows the proportion of users active on each of the other three platforms, as a function of their level of activity within a given platform. Each row corresponds to one of the four platforms (Blog, Facebook, Forum, Meetup), and each column differentiates between candidates (left column) and all users (right column). The x-axis indicates the number of edges within the platform, used here as a proxy for user activity (i.e., interactions or posts). The y-axis represents the proportion of users who are also active on another platform.
Each colored line represents cross-platform presence relative to a specific destination platform (as indicated in the legend on the right). For example, in the Facebook candidate panel, the red line shows the share of Facebook candidates who are also active on the blog, as their Facebook activity increases.
Several key patterns emerge:

- Not surprisingly, candidates (left panels) tend to show higher levels of cross-platform activity than the general user population (right panels).
- Among the general user base, there is no consistent linear relationship between within-platform activity and the likelihood of being active on other platforms. In fact, a positive association between engagement within a platform (x-axis) and cross-platform activity is observable only at lower levels of engagement, particularly for users on Facebook and the Forum.
- A clear pattern emerges in terms of cross-platform flows: regardless of their level of activity, users on Facebook and the Forum are most likely to also engage with the Blog, suggesting it serves as a central hub for broader participation.

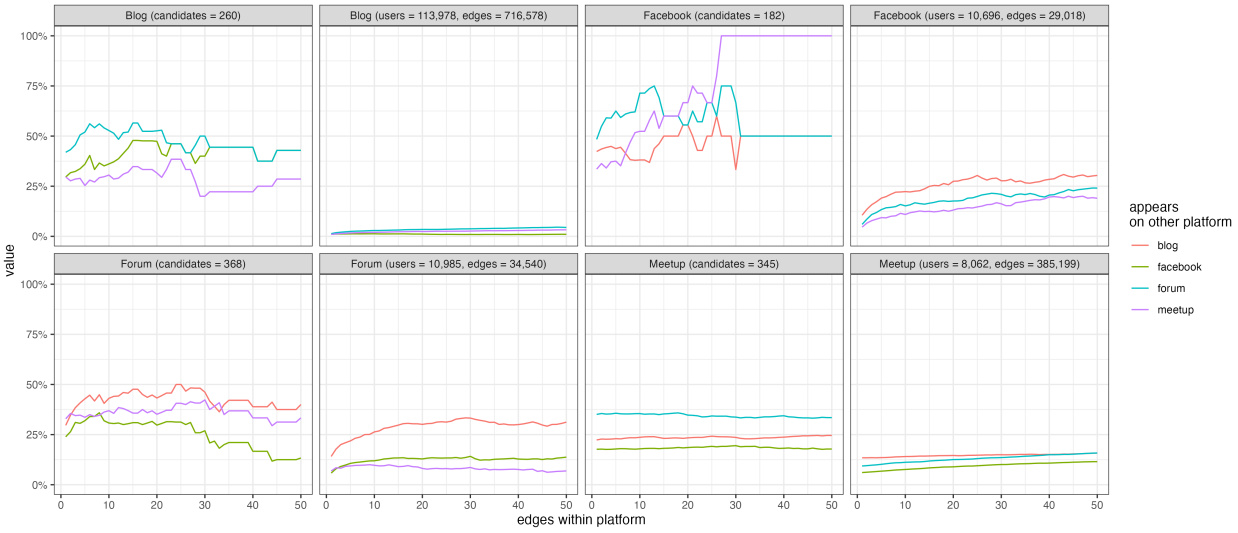

Figure 1: For the platform of users' first activity, the proportion of users active multiple times across each platform.

Table 2: Weekly N-grams percentage by platform.

Platform	Platform sample Avg. # of Tokens	Uni-gram Percentage	Bi-gram Percentage	Tri-gram Percentage
Blog	593246	0.8974	0.2342	0.0053
Facebook	93732	2.2448	0.1425	0.0047
Forum	49215	0.8886	0.0788	0.0004
MeetUp	60536	2.9362	0.4526	0.0094

Measure of Communicative Power by Platform

For each post by Grillo on the blog, we collected comments from the same platform, posts from the forum, Facebook posts, and MeetUp descriptions published within the following seven days. To measure the overlap between the leader's post and content from these platforms, we calculated the frequency of uni-, bi-, and tri-grams from Grillo's post that also appeared in the platform posts or descriptions.

To account for the possibility that some n-grams were commonly used across platforms regardless of the leader's influence, we sampled 10% of posts from each platform to establish a baseline frequency for each n-gram. This baseline helped distinguish n-grams that were widely used from those likely influenced by the leader's communication. All texts were preprocessed to remove punctuation and links.

Next, we calculated the percentage of n-grams from Grillo's post that reappeared on each platform. This percentage was determined as the ratio of n-grams that occurred at least three times more often during the specific week than in the baseline sample (suggesting they were likely adopted following the leader's usage) to the total number of uni-, bi-, or tri-grams. During this process, we excluded uni- and bi-grams containing stopwords and tri-grams containing more than one stopword. For n-grams present in Grillo's post but absent or rare in the baseline sample, we considered them more frequent than average if their frequency during the seven days exceeded three occurrences.

Table 3: Within-platform topic-based cosine distance.

Platform	Mean	Median	Variance	SD
Blog	0.56	0.56	0.05	0.21
Facebook	0.51	0.50	0.07	0.27
Forum	0.47	0.45	0.06	0.25
MeetUp	0.53	0.53	0.06	0.23

Data processing

Topic Models Generation

Two topic models were trained, one encoding the base topics and the second the leadership one. The topic models were obtained using , which can capture syntactic and semantic information of the words and documents and handle polysemy and meaning variations using Large Language Models as encoders. The two models differ in the training dataset: i) the leadership model is trained on all the posts published on Beppe Grillo's blog before the Parlamentarie election; ii) the base model corpus consists of a sample of five comments for each post published on the same blog, to capture a representative sample of what activists wrote in response to the leadership posts. CTMs require a pre-trained encoder model to represent input documents as embeddings: we worked with sentence-bert-base-italian-uncased[1], which showed competitive performance on standard NLP tasks. The input corpora were preprocessed by removing the stopwords and case-lowering the text. Furthermore, because of the model's upper limit of 128 tokens per text, longer texts were divided into buckets of 100 words.
Two authors independently labelled the obtained topics (26 topics for both models, determined using the average coherence score of five runs for each number of topics between 15 and 30) based on the first 25 most representative words for each topic. Only those topics where the labelers assigned a similar label (as determined by a third author not involved in the labeling) were kept, while the others were discarded, yielding 12 topics for the "base" model and 21 for the "leadership" one.
We then measured the topic distributions on all texts whose authors' usernames passed the name selection used for the general network. Each text that was longer than 100 words was divided into chunks of 100 words each: after estimating the topic distribution on each chunk, the whole text's topic distribution was the chunks' mean. For those texts that were shorter than 15 words, in order to improve the prediction accuracy, their length was increased by appending to the text's string a copy of it (e.g., text = *"This is a string."*; final text = *"This is a string. This is a string."*). To obtain the topic distribution of each user, the topic distribution of all the texts produced by each username was averaged. This was calculated using each topic model separately to gauge the association between each user and the topics discussed by the leadership as well those discussed by the activists. Usernames that matched the name and surname combination of the candidates to the 2012 Parlamentarie elections were considered candidates.

Topics Robustness To test the robustness of our models, we conducted a similar procedure using BERTopic models trained on the same data as the CTMs. Like the CTMs, these models used the sentence-bert-base-italian-uncased embedding model. We began by calculating the average coherence score across 25 topic models for each topic number that was a multiple of 5, ranging from 5 to 35. Observing that coherence scores were highest for topic numbers between 20 and 30, we refined our analysis by training 25 models for each topic number within this range. This analysis confirmed that a topic number of 26 consistently achieved one of the highest coherence scores, further validating the robustness of our models. As an additional robustness check, the analysis was first performed with topic vectors from CTM and then repeated with topics from BERTopic. Repeating the analysis with topics generated by different methods did not change the final results.

Measure of Topic Dispersion by Platform Although not included in the main analysis, we also measure the topic dispersion across different platforms. Topic dispersion is derived from the topic vectors generated by the CMT models. For each post, we first calculated the topic distribution and then computed the cosine distance between the topic vector of that post and the vectors of other posts on the same platform. Specifically, for each week from the time M5S began using the platform until the 2012 parliamentary elections (December 2012), we calculated the cosine distance for 10% of the vectors from posts (forum, Facebook), comments (blog, forum, Facebook), and event descriptions (MeetUp) published on each platform. Finally, platform topic dispersion was measured by calculating the average cosine distance across the considered time span.

Topic Measures

Topics heterogeneity
The Gini coefficient was calculated on each candidate topic's distribution to measure the heterogeneity of the topics discussed by the candidates, using both topic models.
Distance from the Leadership
The cosine distance between each candidate's topic distribution and the leadership distribution was computed to measure the distance from the leadership in terms of topics discussed. The topic distribution of the leadership was obtained for each topic model by measuring the topic distribution of each text published by Beppe Grillo's account on the four platforms and calculating the mean of the topics' distribution.

[1]https://huggingface.co/nickprock/sentence-bert-base-italian-uncased

Vanity metrics measures

Because of the different characteristics of the platforms, specific vanity metrics were obtained for each platform to quantify users' activity. For Beppe Grillo's blog, we derived the total number of comments and the number of posts commented on; Facebook vanity metrics consisted in the number of posts published by each user along with the number of comments written and received, the number of likes sent and received and the number of days since the user had been registered on the platform before the Parlamentarie; for the Forum, the number of threads created, the likes, dislikes, score of those threads and the number of comments received by the created threads were part of the platform vanity metrics along with the number of comments posted on the platform and the number of likes, dislikes, and reports that those comments received; finally, the Meetup.com vanity metrics were the number of events a user took part in and the number of days since the registration on the platform. All the vanity metrics correspond to the users' activity before the Parlamentarie of December 2012.

Users' network

This table presents key descriptive statistics for the user interaction networks across each of the four main platforms—Blog, Facebook, Forum, and Meetup—as well as for the combined cross-platform network.
For each network, we report:

- The total number of nodes (users) and edges (replies to a user's post or in the case of Meet co-participation in an event),
- Measures of centrality commonly used in network analysis: degree, in-degree, out-degree, eigenvector, betweenness, and closeness centrality.

The platform-specific networks are typically sparse and vary significantly in scale. For example, the Blog and Facebook have the largest user bases, while Meetup and the Forum are smaller but may exhibit denser or more reciprocal interactions.
Some notable patterns include:

- Meetup displays the highest average degree centrality and eigenvector centrality among the individual platforms, reflecting more tightly connected and possibly event-driven interactions.
- Forum users show relatively high in- and out-degree centrality, pointing to more reciprocal engagement than Facebook and Blog.
- In the cross-platform network, which merges interactions across platforms, the centrality distributions are substantially more varied (as indicated by large standard deviations), reflecting heterogeneous connectivity and the presence of a few highly central actors.

Overall, these metrics provide insight into the structure of engagement within and across platforms, helping to identify patterns of influence, centrality, and connectivity in the broader Five Star Movement digital ecosystem.

Table 4: Descriptive statistics of the users network by platform network and cross-platform one.

	Blog	**Facebook**	**Forum**	**MeetUp**	**Cross Platform***
N. Of nodes	180,571	184,237	23,289	11,869	168,627
N. of Edges	617,143	218,902	46,610	357,505	263,308
Degree centrality	Av: $3.16*10^{-5}$ SD: $2.33*10^{-4}$ Mdn: $4.9*10^{-6}$	Av: $1.29*10^{-5}$ SD: $2.04*10^{-3}$ Mdn: $5.4*10^{-6}$	Av: $1.60*10^{-4}$ SD: $7.66*10^{-4}$ Mdn: $4.11*10^{-5}$	Av: $2.51*10^{-3}$ SD: $3.88*10^{-3}$ Mdn: $1.19*10^{-3}$	Av: 8.07 SD: 341.15 Mdn: 1.0
In-degree centrality	Av: $1.58*10^{-5}$ SD: $1.13*10^{-4}$ Mdn: $4.9*10^{-6}$	Av: $6.5*10^{-6}$ SD: $2.04*10^{-3}$ Mdn: 0.0	Av: $7.97*10^{-5}$ SD: 6.6*10−4 Mdn: 0.0	—	Av: 2.41 SD: 352.18 Mdn: 0.0
Out-degree centrality	Av: $1.58*10^{-5}$ SD: $1.4*10^{-4}$ Mdn: $4.9*10^{-6}$	Av: $6.5*10^{-6}$ SD: $5.6*10^{-6}$ Mdn: $5.4*10^{-6}$	Av: $7.97*10^{-5}$ SD: $3.41*10^{-4}$ Mdn: $4.11*10^{-5}$	—	Av: 2.41 SD: 10.62 Mdn: 1.0
Eigenvector centrality	Av: $3.1*10^{-4}$ SD: $2.2*10^{-3}$ Mdn: 0.0	Av: $7.27*10^{-5}$ SD: $2.33*10^{-3}$ Mdn: 0.0	Av: $7.66*10^{-4}$ SD: $6.36*10^{-3}$ Mdn: 0.0	Av: $1.77*10^{-3}$ SD: $7.3*10^{-3}$ Mdn: $5.23*10^{-5}$	Av: $1.31*10^{-3}$ SD: $8.69*10^{-3}$ Mdn: $5.99*10^{-4}$
Betweenness centrality	—	—	—	—	Av: $1.69*10^{-5}$ SD: $2.3*10^{-6}$ Mdn: $1.77*10^{-5}$
Closeness centrality	—	—	—	—	Av: 17609.52 SD: 2067680.68 Mdn: 0.0

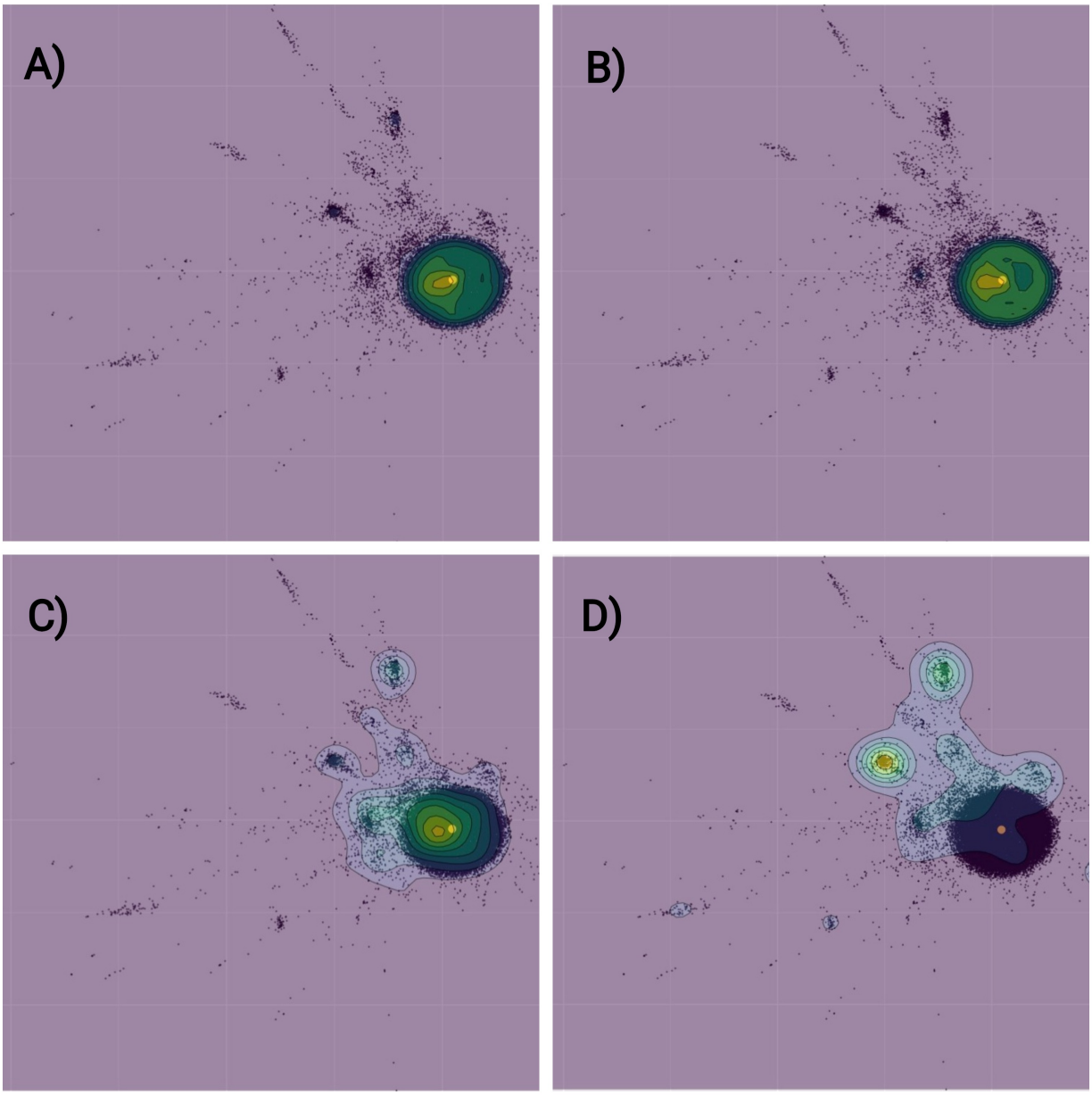

Figure 2: Cross platform networks (only nodes are represented).The figures represent the position of the users' platform in the cross-platform network. A) Blog users; B) Facebook users; C) Forum users; D) MeetUp users.

This figure visualizes the cross-platform user network, using a two-dimensional layout. The layout is based on a force-directed or dimensionality-reduction algorithm, meaning that nodes positioned closer together have a higher frequency of interaction or structural similarity in the network. In other words, spatial proximity in the figure reflects relatively more interaction among those users. Each panel (A–D) displays a density contour plot showing the distribution of users from a specific platform.
Clear differences in user distribution patterns emerge across platforms:

- Blog and Facebook users (Panels A and B) exhibit strong central clustering, indicating that their interactions are concentrated around a dense core in the cross-platform network. This suggests these users are tightly embedded in the most interconnected portions of the broader network.
- In contrast, Forum and Meetup users (Panels C and D) are more diffusely spread across the network. These users appear in multiple smaller clusters or more peripheral regions, pointing to more distributed patterns of engagement and potentially less overlap with the core user communities.

Overall, this spatial distribution highlights the varying degrees of integration and centrality of users from different platforms within the overall network structure. Blog and Facebook users dominate the network's core, while Forum and Meetup users are more scattered and peripheral.

Topic models

Leadership's Topics

Topic 1: italia, nucleare, tav, susa, val, centrali, nucleari, francia, euro, enel, inceneritori, energia, valle, torino, telecom, italiani, costi, centrale, europa, piemonte, rifiuti, gas, societa, aziende, referendum
Topic 2 [EXCLUDED]: blog, beppe, grillo, libro, morte, morto, ragazzo, intervista, carcere, lettera, marco, figlio, scritto, caro, passaparola, storia, testo, mail, giornalista, comprimi, espandi, padre, morti, pubblicato, giuseppe
Topic 3 [EXCLUDED]: day, video, immagine, libera, gt, foto, aprile, clicca, inserisci, informazione, tag, blog, casta, post, ps, giornali, testo, sito, mail, rete, internet, libero, stampa, nome, beppe
Topic 4: grecia, debito, euro, tremorti, pubblico, italia, pil, default, crisi, ue, titoli, europa, paesi, economia, crescita, germania, banche, europei, aumento, valore, italiani, pubblici, economica, europea, interessi
Topic 5: stare, bambino, donne, mattina, so, bambini, posso, qua, donna, madre, devo, strada, figli, figlio, papa, famiglia, occhi, lavorare, vedere, devi, aiuto, macchina, pensare, uscire, notte
Topic 6 [EXCLUDED]: milano, calabria, san, genova, reggio, polizia, zona, torino, metri, cemento, notte, roma, val, piazza, tav, area, auto, aria, milanese, susa, giovanni, valle, velocita, poliziotti, parco
Topic 7: mafia, ciancimino, palermo, falcone, provenzano, trattativa, stragi, borsellino, strage, mafioso, spatuzza, mafiosi, utri, padre, rapporti, carabinieri, massimo, cose, berlusconi, uomini, magistrati, sicilia, parlare, antimafia, boss
Topic 8: cina, petrolio, pianeta, uniti, paesi, terra, guerra, mondiale, armi, energia, militare, risorse, prezzo, produzione, nucleari, nucleare, livello, internazionale, popolazione, sviluppo, crescita, nato, usa, gas, accordo
Topic 9: parlamento, legge, pulito, referendum, popolare, senato, firme, cittadini, elettorale, democrazia, costituzione, iniziativa, proposta, partiti, condannati, parlamentari, voto, schifani, commissione, camera, leggi, italiani, presidente, definitiva, votare
Topic 10 [EXCLUDED]: italia, italiani, guerra, sud, italiano, mafia, europa, germania, nord, americani, storia, morti, politica, italiana, nato, sicilia, politici, uniti, criminalita, paesi, pace, nazione, repubblica, militari, uomini
Topic 11: rifiuti, raccolta, ambientale, riduzione, gestione, fonti, costruire, energia, progetto, parco, comune, territorio, inceneritori, produzione, acqua, pubblici, piano, prodotti, metri, servizio, ambiente, porta, qualita, salute, posti
Topic 12: pd, partito, sinistra, centro, destra, voti, pdl, votato, lega, partiti, votare, berlusconi, maggioranza, elettori, bersani, pdmenoelle, referendum, elezioni, casini, politica, neanche, candidato, fini, nucleare, voto
Topic 13: processo, corte, tribunale, cassazione, giudici, giudice, sentenza, appello, giudizio, condannato, presidente, parlamento, procura, associazione, pm, ministri, grado, condanna, richiesta, imputato, costituzionale, magistrati, giustizia, milano, definitiva
Topic 14: berlusconi, legge, lodo, costituzionale, alfano, presidente, corte, legittimo, napolitano, capo, costituzione, decreto, leggi, fini, ministri, vuole, conflitto, processi, cose, repubblica, sentenza, maggioranza, giudici, silvio, comprimi'
Topic 15: partiti, cittadini, politica, movimento, ms, soldi, pubblici, elettorali, stelle, finanziamenti, partito, pubblico, regionali, politici, euro, elezioni, democrazia, consiglieri, pubblica, finanziamento, comuni, acqua, cittadino, giornali, programma
Topic 16: reato, prescrizione, processo, pena, processi, reati, penale, commesso, giudizio, legge, breve, condanna, carcere, grado, galera, giudice, cassazione, sentenza, norma, intercettazioni, viene, giustizia, bisogna, possono, vengono
Topic 17: giornali, pubblicita, rete, giornale, rai, tg, televisione, giornalisti, corriere, quotidiano, stampa, informazione, internet, tv, sera, televisioni, giornalista, direttore, libero, pagina, notizie, mediaset, media, blog, sito
Topic 18: de, magistris, genchi, magistrati, salerno, mastella, procura, indagine, indagini, procuratore, inchiesta, carte, magistrato, indagati, atti, capo, indagato, segreto, pm, segreti, polizia, telefono, viene, colleghi, carabinieri
Topic 19: politica, cittadini, legge, politici, giustizia, magistratura, polizia, leggi, sicurezza, diritti, magistrati, diritto, reati, intercettazioni, responsabilita, potere, verita, costituzione, cittadino, forze, parlamento, pubblico, magistrato, politico, istituzioni
Topic 20: presidente, partito, senatore, berlusconi, pdl, deputato, regionale, milano, consigliere, direttore, roma, regione, antonio, pd, segretario, paolo, nord, sindaco, bersani, lega, de, politica, gianni, craxi, giunta
Topic 21: politica, espandi, comprimi, grillo, beppe, cose, gente, dobbiamo, vogliamo, sistema, cambiare, vivere, bisogna, crisi, penso, credo, problema, cambiamento, cultura, movimento, possiamo, democrazia, potere, blog, punto
Topic 22: berlusconi, fininvest, soldi, societa, mills, finanza, guardia, processo, craxi, previti, tangenti, silvio, milano, corruzione, sentenza, giudici, bilancio, giudice, conto, reato, prescrizione, fondi, falso, quei, conti
Topic 23: euro, pensione, pagare, stipendio, pensioni, contributi, soldi, tasse, mese, paga, spese, stipendi, lavorare, dipendenti, parlamentari, lavoratori, fiscale, eta, pagato, famiglia, contratto, figli, italiani, spesa, giovani
Topic 24: stelle, movimento, lista, grillo, bologna, cinque, beppe, liste, emilia, comune, consigliere, ms, comunale, firenze, maggio, marzo, comuni, regionale, genova, programma, piazza, sindaco, consiglieri, candidati, reggio
Topic 25: soldi, banche, telecom, banca, debiti, societa, borsa, aziende, imprese, azioni, pagare, debito, fondi, euro, tasse, azienda, mercato, titoli, pubblico, valore, investimenti, comprare, grandi, fiat, capitale
Topic 26 [EXCLUDED]: epoca, com, opinione, unione, conviene, passate, estate, arrenderanno, entra, quell, finale, effetto, occasione, organizzazione, piena, operazione, azione, riferimento, morale, ndrangheta, ultima, maroni, sanita, certamente, vergogna

Base's topics

Topic 1: legge, referendum, cittadini, elettorale, parlamento, firme, democrazia, popolare, votare, voto, elezioni, partiti, costituzione, movimento, proposta, eletti, lista, politica elettori, diritto, partito, iniziativa, leggi, diretta, proposte
Topic 2 [EXCLUDED]: genitori, figli, bambini, figlio, madre, notte, padre, terra, cuore, vivere, occhi, gente, nanon, strada, morire, amore, paura, moglie, vedere, cose, animali, voglio, andare, posso, persona
Topic 3 [EXCLUDED]: on, of, to, the, and, na, nai, day, roma, san, nail, napoli, notte, milano, polizia, nala, aprile, settembre, genova, link, nae, stelle, vaticano, naa, mafia

Topic 4: berlusconi, presidente, legge, repubblica, premier, napolitano, silvio, roma, decreto, corte, de, magistrati, capo, processo, giustizia, commissione, reato, costituzione, magistratura, direttore, na, polizia, milano, italia, articolo
Topic 5 [EXCLUDED]: unione, nasono, aspettare, arrivato, cercando, nasi, nase, domande, mette, segue, appunto, permesso, fara, nacome, esattamente, siccome, uguale, avra, creato, aprire, vorrebbe, nomi, simile, cerca, alternativa
Topic 6: politica, popolo, politici, cittadini, gente, potere, giustizia, casta, democrazia, mafia, classe, societa, italiani, bisogna, mafiosi, ordine, dobbiamo, istituzioni, legge, sociale, forze, polizia, politico, difendere, italia
Topic 7 [EXCLUDED]: politici, soldi, pensioni, stipendi, euro, pensione, parlamentari, cittadini, tasse, pubblici, stipendio, privilegi, pagare, lavoratori, spese, dipendenti, partiti, sindacati, mese, pubblico, contributi, parlamento, italiani, politica, dirigenti
Topic 8: blog, grillo, post, commenti, commento, beppe, scrivere, scritto, leggere, cose, letto, articolo, sito, ciao, argomento, risposta, mail, credo, link, so, saluti, nacaro, scrive, travaglio, parlare
Topic 9: italia, grecia, debito, banche, euro, moneta, banca, europa, italiani, paesi, crisi, germania, soldi, debiti, economia, europei, italiano, mondiale, cina, europea, pubblico, politica, sistema, usa, conti
Topic 10 [EXCLUDED]: soldi, pagare, gente, tasse, andare, banche, lavorare, devi, bisogna, nama, cazzo, vuole, devono, fame, quei, vogliono, far, ricchi, vedere, dare, figli, dobbiamo, politici, vanno, pensione
Topic 11 [EXCLUDED]: sito, internet, blog, cose, informazioni, credo, problema, post, link, telecom, mail, scritto, pagina, so, commento, web, pubblicita, possibile, numero, leggere, argomento, errore, informazione, rete, dati
Topic 12: pd, pdl, pietro, grillo, partito, idv, votato, sinistra, votare, bersani, voti, berlusconi, voto, lega, prodi, elezioni, movimento, destra, alema, beppe, casini, stelle, partiti, ms, credo
Topic 13 [EXCLUDED]: chiesa, religione, dio, donne, dura, uomini, storia, donna, papa, amore, fede, verita, pace, video, guarda, popolo, morte, liberta, violenza, guerre, figli, vaticano, parole, giustizia, societa
Topic 14 [EXCLUDED]: euro, mese, pagare, soldi, stipendio, pensione, moglie, contratto, pagato, devo, italia, tasse, km, milano, andare, settimana, costo, auto, almeno, spese, paga, roma, lavorare, naio, vado
Topic 15 [EXCLUDED]: italia, italiani, italiano, estero, popolo, europa, guerra, sud, paesi, roma, italiana, figli, terra, merda, gente, mafia, nazione, germania, culo, storia, nae, nord, francia, vero, america
Topic 16 [EXCLUDED]: ciao, nagrazie, saluti, saluto, giornata, buona, paolo, francesco, link, naun, piacere, complimenti, nami, anch, presto, commento, speriamo, dispiace, buon, mail, volevo, rispondere, risposta, scusa, messaggio
Topic 17: movimento, politica, stelle, gente, sinistra, ms, grillo, destra, politici, idee, cose, partiti, cambiamento, cambiare, bisogna, tv, potere, democrazia, rivoluzione, partito, popolo, rete, sistema, votare, credo
Topic 18: evasione, iva, fiscale, reddito, euro, tasse, azienda, attivita, imprese, pagare, settore, aziende, societa, spese, lavoratori, dipendenti, banche, paga, dipendente, contratto, moneta, valore, privato, mercato, costo
Topic 19 [EXCLUDED]: beppe, grillo, cose, italia, italiani, gente, politica, cambiare, italiano, credo, nacaro, far, tanti, so, veramente, blog, persona, davvero, popolo, capire, dobbiamo, parlare, ciao, penso, parole
Topic 20: nucleare, energia, centrali, rifiuti, acqua, nucleari, centrale, italia, petrolio, km, produzione, costruire, problema, auto, nord, aria, produrre, sud, mare, ambiente, prodotti, napoli, torino, treno, zona
Topic 21: berlusconi, nord, mafia, lega, sinistra, bossi, leghisti, destra, politica, italia, sud, gente, italiani, politici, mafiosi, merda, soldi, culo, partito, vero, cazzo, fini, mafioso, popolo, votato
Topic 22 [EXCLUDED]: ah, de, er, eh, nano, ce, so, cazzo, na, merda, sti, mica, culo, naa, sai, ciao, nama, sa, nanon, vede, vai, qua, vaffanculo, sera, pare
Topic 23: tv, rai, italia, canale, rete, tg, internet, televisione, giornali, berlusconi, beppe, informazione, giornale, italiani, notizie, grillo, giornalisti, blog, notizia, politica, gente, pubblicita, giornalista, soldi, media
Topic 24 [EXCLUDED]: beppe, alvise, naciao, day, grillo, stelle, movimento, ciao, milano, roma, blog, nacaro, manifestazione, sindaco, nagrazie, lista, video, piazza, nabeppe, ragazzi, caro, saluto, spettacolo, firme, sito
Topic 25 [EXCLUDED]: nama, nae, merda, nache, coglioni, nanon, vaffanculo, sti, culo, naa, frega, cervello, nabeppe, nai, mica, vai, schifo, galera, prendono, cazzate, bravi, andiamo, cazzo, figura, leghisti
Topic 26: guerra, armi, pianeta, guerre, animali, mondiale, sviluppo, risorse, cina, umano, petrolio, sistema, umani, ricerca, produzione, societa, ambiente, umana, paesi, usa, salute, pace, natura, terra, internazionale

Topic selection

Table 5: Topic selection Leadership model

Topic number	Topic name	Labeler 1	Labeler 2	Passed
1	italia_nucleare_tav	Environmental Protection, Sustainability	Environment	1
2	blog_beppe_grillo	Digital Democracy, Leadership	Undefined	0
3	day_video_immagine	Digital Democracy, Anti-Establishment	Undefined, Digital communication	0
4	grecia_debito_euro	Anti-Imperialism, Foreign Financial Influence	European debt crisis	1
5	stare_bambino_donne	Maybe People-Centrism	Family	1
6	milano_calabria_san	Environmental Protection, Sustainability	Undefined	0
7	mafia_ciancimino_palermo	Political Corruption, Anti-Establishment, Conspiracy theories	Mafia-state	1
8	cina_petrolio_pianeta	Anti-Imperialism, Anti-Establishment	International conflicts	1
9	parlamento_legge_pulito	Political Corruption, Anti-Establishment, Direct Democracy	Democratic institutions	1
10	italia_italiani_guerra	Anti-Imperialism, Anti-Establishment	Mafia	0
11	rifiuti_raccolta_ambientale	Environmental Protection, Sustainability	Environment, Waste management	1
12	pd_partito_sinistra	Political Corruption, Anti-Elitism	Italian party politics	1
13	processo_corte_tribunale	Political Corruption, Judicial Populism	Justice	1
14	berlusconi_legge_lodo	Political Corruption, Anti-Elitism - Populism	Berlusconi's legal issues	1
15	partiti_cittadini_politica	Civic Mindedness, Bottom-Up Activism	Elections-voting	1
16	reato_prescrizione_processo	Political Corruption, Judicial Populism	Justice	1
17	giornali_pubblicita_rete	Traditional Media, Anti-Establishment	New and old media	1
18	de_magistris_genchi	Political Corruption, Judicial Populism	Justice	1
19	politica_cittadini_legge	Political Corruption, Judicial Populism	Justice	1
20	presidente_partito_senatore	Political Corruption, Anti-Establishment, Anti-Elitism	Party politics	1
21	politica_espandi_comprimi	Civic Mindedness, Bottom-Up Activism, Digital Democracy	MS High politics	1
22	berlusconi_fininvest_soldi	Political Corruption, Anti-Establishment, Conspiracy theories	Berlusconi- Fininvest- Corruption	1
23	euro_pensione_pagare	Economic issues, Undefined	Microeconomics	1
24	stelle_movimento_lista	Civic Mindedness, Bottom-Up Activism	MS	1
25	soldi_banche_telecom	Anti-Imperialism, Foreign Financial Influence	Finance	1
26	epoca_com_opinione	Political Corruption	Undefined	0

Table 6: topic selcetion base model

Topic number	Topic name	Labeler 1	Labeler 2	Passed
1	legge_referendum_cittadini	Direct Democracy, Populism	Participation, Electoral institutions	1
2	genitori_figli_bambini	Mostly undefined, some terms may refer to People-centrism	Family	0
3	on_of_to	Mostly undefined, some terms may refer to Anti-Establishment	Undefined	0
4	berlusconi_presidente_legge	Political Corruption, Anti-Establishment, Judicial Populism	Berlusconi, judicial issues	1
5	unione_nasono_aspettare	Undefined	Undefined	0
6	politica_popolo_politici	People-Centrism, Populism	Politics and Democracy	1
7	politici_soldi_pensioni	Political Corruption, Anti-Elitism, Populism	Microeconomy	0
8	blog_grillo_post	Digital Democracy Leadership	Digital communication, Blog	1
9	italia_grecia_debito	Anti-Imperialism, Foreign Financial Influence	European sovereign debt crisis	1
10	soldi_pagare_gente	Undefined, maybe Anti-Establishment	Microeconomy	0
11	sito_internet_blog	Digital Democracy, Leadership	Undefined	0
12	pd_pdl_pietro	Political Corruption, Anti-Elitism, Populism	Italian party politics	1
13	chiesa_religione_dio	Peace	Religion, Vatican	0
14	euro_mese_pagare	Economic Issues, Undefined	Microeconomy	0
15	italia_italiani_italiano	Anti-Imperialism, or Anti-Establishment	Undefined, (foreign affairs)	0
16	ciao_nagrazie_saluti	Undefined	Undefined	0
17	movimento_politica_stelle	Civic Mindedness, Bottom-Up Activism, Representative Democracy	Five Star Movement, Participation, Political change	1
18	evasione_iva_fiscale	Economic issues, Undefined	Corporate / Tax evasion	1
19	beppe_grillo_cose	Digital Democracy, Leadership	Undefined	0
20	nucleare_energia_centrali	Environmental Protection, Sustainability	Environment	1
21	berlusconi_nord_mafia	Political Corruption, Anti-Establishment	Politics and organised crime	1
22	ah_de_er	Undefined, maybe Anti-Politics	Rants	0
23	tv_rai_italia	Traditional Media, Anti-Establishment	Old and News Media	1
24	beppe_alvise_naciao	Civic Mindedness: Bottom-Up Activism	Undefined, (Five Star Movement politics)	0
25	nama_nae_merda	Undefined, maybe Anti-Politics	Rants	0
26	guerra_armi_pianeta	Anti-Imperialism	International conflicts	1

Blocks of variables included in the RF model

Interface: age (median centered by district); screen position (unit normalized per district); sex; logo [3 levels: no picture available, no party logo in picture, party logo in picture]; number of candidates in district (logged); percentage of women in a district; percentage of candidates showcasing a logo in a district.

Local elections: previous candidacy to local elections [2 levels: yes, no]. This variable was added to the previous batch of predictors.

Online presence: candidate present in MeetUp, candidate present on Facebook.com; candidate present on the Forum; candidate present on the leader's blog; candidate present in all 4 platforms. All variable have 2 levels: yes, no. These predictors were added to those included in the previous set of predictors.

Platform centralities: blog degree, eigenvector, indegree and outdegree centrality measures; forum degree, eigenvector, indegree, and outdegree centrality measures; MeetUp degree and eigenvector centrality measures; Facebook.com degree, eigenvector, indegree, and outdegree centrality measures. All variables are ordered factor with 4 levels: not present on the platform, in the lowest tertile of users in a district, in the second tertile of users in a district, in the top tertile of users in a district. Since these variables subsume the variable coding for online presence, we added this set of features to the model including interface and local elections predictors.

In the statistical analysis, we tested several cross-network metrics considering networks derived from ties between users and organizations, user and locations, users and people as determined using SpaCy's Named Entity Recognition tagger. These network metrics improved predictions over the model including interface predictors, offline elections, and platform centrality metrics, but did not contribute further once network metrics from the user cross-network were considered. For this reason, we report only the user cross-network in the main text: details about these models can be found in the source code.

User:loc network: betweenness, closeness, eigenvector and degree centrality, page rank, distance from the leader, presence in the same community as the leader (binary) as computed from the bipartite network connecting users to entities tagged as locations. All variables except presence in the leader's community are ordered factors with 4 levels: not present, lower tertile, mid tertile, top tertile, computed by district. Measures were computed from the giant component of the network.

User:org network: same measures reported for the user:loc network, with the same data types and levels, but computed from the bipartite network connecting users to entities tagged as organizations.

User:per network: same measures reported for the user:loc and user:org networks, with the same data types and levels, but computed from the bipartite network connecting users to entities tagged as person.

User:url network: same measures reported for the user:loc, user:org, and user:per networks, with the same data types and levels, but computed from the bipartite network connecting users to URLs.

Cross-platform network: same network measures detailed for the four previous sets, computed from the user:user network across all platforms. These 5 sets of predictors were first added individually to the model including predictors from the interface, local elections, and platform specific centrality measures. The best performing model included predictors from the cross platform network. We than included predictors from the other four sets (user:org, user:loc, user:per, user:url) separately and never observed an improvement. We thus built following models using predictors concerning the interface, local elections, platform specific centrality measures, and network measures from the cross-platform network.

Topics: cosine distance between a candidate's topic distribution and the leader's topic distribution - computed for the base and leadership topic model; GINI coefficient of the topic distribution, computed for the base and leadership topic model; association between candidate and topics for which independent labelers agreed (topics are identified by the three most relevant words, BT indicates topics from the base topic models, LT from the leadership topic model: blog_grillo_post_BT, euro_pensione_pagare_LT, stelle_movimento_lista_LT, movimento_politica_stelle_BT, berlusconi_presidente_legge_BT, politica_espandi_comprimi_LT, dist_leader_LT, pd_partito_sinistra_LT, pd_pdl_pietro_BT, parlamento_legge_pulito_LT, presidente_partito_senatore_LT, italia_grecia_debito_BT, stare_bambino_donne_LT, politica_cittadini_legge_LT, giornali_pubblicita_rete_LT, italia_nucleare_tav_LT, processo_corte_tribunale_LT, de_magistris_genchi_LT, evasione_iva_fiscale_BT, berlusconi_fininvest_soldi_LT, berlusconi_legge_lodo_LT, reato_prescrizione_processo_LT, guerra_armi_pianeta_BT, cina_petrolio_pianeta_LT, politica_popolo_politici_BT, berlusconi_nord_mafia_BT, grecia_debito_euro_LT, partiti_cittadini_politica_LT, nucleare_energia_centrali_BT, soldi_banche_telecom_LT, tv_rai_italia_BT, cosine_distance_LT, mafia_ciancimino_palermo_LT, rifiuti_raccolta_ambientale_LT, giornali_pubblicita_rete_LT. All variables were coded as ordered factors with 4 levels: did not publish on any platform, in the lower tertile by association to a topic, in the mid tertile by association to a topic, in the top tertile by association to a topic, computed by district. This set of variables was added to the model incorporating predictors from the interface, local elections, platform specific centrality measures, and measures from the cross-platform network.

We further considered measures that capture the sheer quantity of candidates' involvement on different platforms, to insert a set of control variables to exclude the possibility that what influenced electoral outcomes was not how the candidates connected with other users or how they posted but rather how much they posted. These set of predictors did not improve model predictions, and was thus not reported in the paper.

Vanity metrics: number of comments in the blog; number of posts in the blog commented; dislikes received to forum comments; number of reports to forum comments; likes received to forum comments; number of forum comments; dislikes received to forum threads created;, likes received to forum threads created; number of posts over threads created; score of forum threads; number of forum threads created; number of MeetUp events created; days active on MeetUp; number of Facebook.com posts written; number of Facebook.com comments written; number of likes given on Facebook.com; number of comments received on facebook.com; number of likes received on Facebook.com; days active on Facebook.com. All variables were coded as ordered factors with 4 levels: not present on the platform, in the lower tertile, in the mid tertile, in the top tertile, computed by district. This set of variables was added to the model incorporating predictors from the interface, local elections, platform specific centrality measures, and measures from the cross-platform network.

Analysis Plan

We used the randomForest package in R and ran 50 iterations of each model to obtain robust estimates of variance explained and feature importance, two key metrics in our study. We fitted 500 trees using one third of the available predictors for each tree. All other parameters were kept at default values. Models differ in the set of predictors considered, which we input in batches to monitor whether the inclusion of a novel set of predictors improves the amount of variance explained in the target dependent variable, i.e. the rank of each candidate in the election, unit-normalized by district such that the candidate who ranked first has a value of 0 in all districts and the candidate who ranked last has a value of 1 in all districts, regardless of the number of competing candidates. We start from a model which considers the predictors analyzed by Marolla et al (2023). We then consider i) whether a candidate featured on the ballot of a local election before 2012; ii) whether a candidate was present on the four target platforms; iii) measures of network centrality (degree and eigenvector) computed separately on each platform; iv) network measures (closeness, betweenness, eigenvector and degree centrality, page rank, distance from the leader and presence in the leader's community) computed on networks derived from the bipartite networks user-locations, user-person, user-organization, user-url, and on the giant component of the cross-platform network; v) the association between each candidate and the topics identified in the posts by each candidate as determined using the topic model trained on blog posts from the leadership and the model trained on comments from the activists; vi) an array of vanity metrics which tracked candidate's activity at a quantitative level.

Several variables have missing values that are, however, not missing at random, preventing imputation: for example, if a candidate was not found on Meetup.com, that candidate will not have a degree centrality on Meetup.com. However, limiting the analysis to candidates for which all metrics can be computed would reduce the sample size too much. For this reason all network measures, associations to topics, and vanity metrics were engineered as follows: for each district, we allotted candidates for which values were not available to a bucket called 'not present'. Candidates for which values were available, were sorted and divided in tertiles, creating three buckets: the first contains candidates with the lowest values, the second candidates with average values, and the third the candidates with the highest values. This was done per district to reflect the fact that the electoral competition in the primaries was within each district. The choice to use tertiles strikes a balance between preserving some variation while avoiding too fine-grained divisions since in some districts only few candidates were found to be active on any platform. For each random forest model we report the average variance explained in the dependent variable over 50 runs, with the corresponding 95% Confidence Interval. For the best model, we further visualize feature importance, again computed over 50 runs, and finally zoom in on important predictors to assess their relation with the target dependent variable.