PAPER • OPEN ACCESS

Quantifying the drivers behind collective attention in information ecosystems

To cite this article: Violeta Calleja-Solanas *et al* 2021 *J. Phys. Complex.* **2** 045014

View the article online for updates and enhancements.

Journal of Physics: Complexity

PAPER

Quantifying the drivers behind collective attention in information ecosystems

Violeta Calleja-Solanas[1,*] , Emanuele Pigani[2,6] , María J Palazzi[3] ,
Albert Solé-Ribalta[3,4] , Samir Suweis[2,5] , Javier Borge-Holthoefer[3] and
Sandro Meloni[1,*]

1 IFISC, Institute for Cross-Disciplinary Physics and Complex Systems (CSIC-UIB), 07122, Palma de Mallorca, Spain
2 LIPh Lab, Physics and Astronomy Department, University of Padova, via Marzolo 8, 35131, Padova, Italy
3 Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Catalonia, Spain
4 URPP Social Networks, University of Zurich, Zurich, Switzerland
5 Padova Neuroscience Center, University of Padova, Via Orus 2/B, 35131 Padova, Italy
6 Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, 80121, Italy
* Authors to whom any correspondence should be addressed.

E-mail: violeta@ifisc.uib-csic.es and sandro@ifisc.uib-csic.es

Abstract

Understanding human interactions in online communications is of paramount importance for our society. Alarming phenomena such as the spreading of fake news or the formation of echo-chambers can emerge in unhealthy communication environments and, ultimately, undermine the democratic discourse. In this context, unveiling the individual drivers that give rise to collective attention can help to conserve the health of our information ecosystems. Here, following a recently proposed analogy between natural and information ecosystems, we explore how competition for attention in online social networks and the strategies adopted by the users to maximize their visibility shape our communication dynamics. Specifically, by analyzing large-scale datasets from the micro-blogging platform Twitter and performing numerical modeling of the system dynamics, we are able to measure the amount of competition for attention experienced by users and how it changes when exogenous events captivate collective attention. The work relies on topic modeling to extract users' interests and memes context from the data and a framework based on ecological niche theory to quantify the strength of negative (competitive) and positive (mutualistic) interactions for both users and memes. Interestingly, our findings show two different behaviors. While memes undergo a sharp increase in competition during exceptional events that can lead to their extinction, users perceive a decrease in effective competition due to a stronger effect of mutualistic interaction, explaining the focus of collective attention around specific topics. Finally, to confirm our results we reproduce the observed shifts with a data-driven model of species dynamics.

1. Introduction

Online social networks (OSNs) have not only transformed the way in which we communicate, but also how we access and process information. In fact, since their appearance, they have played a double role: as spaces to build social interactions and as news media platforms. Consequently, our communication model has switched over from a centralized mass media environment and face to face interactions, to an era when all the actors are, at the same time, information sources and receivers. This duality and the new paradigm it induced, make also OSNs a perfect example of social information processing, leading to emergent phenomena such as viral information spreading [1], fake news [2] and the formation of echo-chambers [3, 4].

These changes have also exposed our cognitive limitations. Our brain passed from hearing few broadcast information sources, to be bombed by millions of messages demanding our attention. This has led to a variety of phenomena, such as an acceleration of social dynamics [5], and terms like 'competition for attention' entered

our everyday vocabulary. Moreover, OSNs are often designed to captivate our brain and maximize the time we spend on them by providing reinforcing feedbacks [6] and instant gratification [7, 8]. These limitations along with the huge amount of information we produce every second, induce competition between ideas/memes for visibility, and users tend to adopt different strategies to increase the chances that their messages would be read.

To understand how these low-level drivers shape the way our society processes information, in recent years, different approaches have been proposed. From a data analytic perspective, several works demonstrated the role of competition in information diffusion [9, 10], our social interactions [11–14] and the quality of the information we share [15]. All these results also found confirmation from theoretical models, for example for the distribution of memes popularity [16–18] or how echo-chambers emerge [19].

This strong emphasis on the effects of competition and on how users respond to it, also led researchers to draw an analogy between information and natural ecosystems [5, 20–22]. Actors (e.g. users and memes) of OSNs are seen as species in ecological communities where they seek to maximize their abundance—visibility in this case—and all are competing for limited resources (e.g. individuals' attention). Moreover, the communication strategies adopted by the users like, for example, the use of specific hashtags to provide context to their messages can be represented as mutualistic, as they favor both visibility of users and growth of certain memes.

One example of how this analogy can lead to insights on the functioning of information ecosystems has been the work of Borge-Holthoefer *et al* [20], who studied the organization of online discussions around social protests in Spain in 2011 as an ecological network, finding that its structure evolved toward a nested architecture, very close to the typical organization of natural mutualistic assemblages [23, 24]. Moreover, building on these results, in a recent work Palazzi *et al* [21] proposed an ecology-inspired model [25, 26] to explain the structural flexibility showed by OSNs in response to external shocks such as breaking news. In [5], instead, the authors were able to explain the acceleration in collective attention they found in the digital streams and cultural items by employing a mathematical model based on Lotka–Volterra dynamics—a theoretical framework often used in population dynamics and theoretical ecology [27, 28]. Finally, from a more theoretical perspective, models based on the concept of ecological neutrality [29, 30], were able to reproduce several emergent patterns found in online communications [22].

Here, following this research line, we exploit the similarities between natural and information ecosystems to understand the main drivers that shape collective attention. Specifically, analyzing different datasets extracted from the online platform Twitter and an ecology-inspired numerical model, we are able to quantify the intensity of the competitive and mutualistic interactions experienced by each user, and how these change when exogenous events focus collective attention. We start by relying on topic modeling to study the evolution of users' interests over time, along with the diversity of the discussion around the different topics. In this way, we can measure the similarity between users and memes and then, employing a modeling framework based on ecological niche-theory [26], build interaction networks similar to ecological communities with both competitive and mutualistic terms. We then analyze the obtained networks with the aim of understanding how variations in the amount of competition and mutualism experienced by the users can explain the focus of collective attention around one or few dominating topics. These analyses allow us to explain these shifts in attention in terms of a reduction of the effective competition experienced by the users. We then confirm this finding by reproducing the patterns we observe in the data with numerical simulations of an ecological model based on species abundance maximization [21]. Our results not only shed light on how information processing drives users' behaviors in social media but, in a broader sense, they spotlight the opportunities that an ecological approach can offer to the study of information ecosystems for understanding collective phenomena in techno-social systems

The rest of the work is organized as follows. In the next section, we introduce the datasets we will employ in the analysis, the topic modeling technique used to extract information topics from Twitter discussions, the ecologically inspired numerical model and the methods to quantify the strength of competitive and mutualistic interactions from the data. Then, in sections 3.1 and 3.2 we present our results for the time evolution of topics and users' interests respectively while, in section 3.3, we study the user–meme interaction networks. Moreover, in section 3.4, we confirm our previous results with the numerical simulations of our model. Finally, with section 4 we summarize the main findings of this work and draw our conclusions.
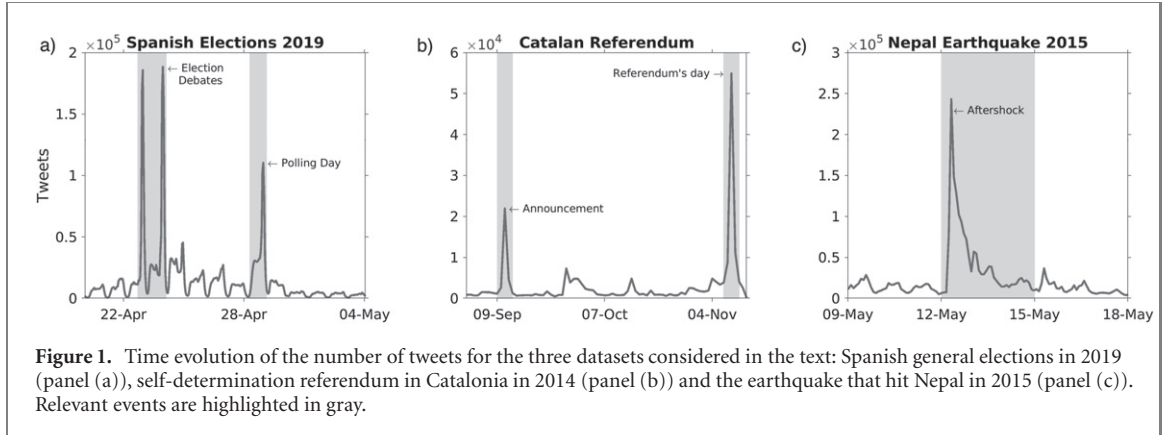
2. Methods and data

2.1. Twitter events datasets
Since we are interested in quantifying the changes in collective attention during exceptional events, we decided to focus on data from the online social platform Twitter. Twitter, along with being a social network and due to its microblogging nature, is also often considered as a news and information media. Thus, social, political

Table 1. Summary of the datasets used in the study.

Dataset	Starting date	Ending date	# of tweets	# of unique hashtags	# of unique users
Spanish elections	20-04-2019	04-05-2019	4882 546	2952	41 314
Catalan referendum	01-09-2014	13-11-2014	220 364	18116	78 270
Nepal earthquake	09-05-2015	18-05-2015	5641 719	41 589	1706 559

Figure 1. Time evolution of the number of tweets for the three datasets considered in the text: Spanish general elections in 2019 (panel (a)), self-determination referendum in Catalonia in 2014 (panel (b)) and the earthquake that hit Nepal in 2015 (panel (c)). Relevant events are highlighted in gray.

or natural events will be likely reflected in its users' activity, allowing us to record both the change in their interests and in their interactions.

2.2. Information topics

For our analyses, we focus on three large datasets describing Twitter activity: the consultations and protests around the self-determination referendum organized by the Catalan government in November 2014 [21]; the general elections held in Spain in 2019 [21] and the response to the earthquake that hit Nepal in 2015 [31]. All the datasets have been collected through the Twitter streaming API and are publicly available (see appendix A for details on the data collection process and their availability). Table 1 summarizes the main features of the datasets used.

We choose those datasets because they cover both expected and unexpected events [32, 33]: from political discussions with a fixed timeline (Spanish elections), to political/social protests like the Catalan referendum with a mixture of organized activities and improvised protests, and an inherently unexpected natural disaster (Nepal's earthquake). In particular, since our aim is to study how exceptional events shape the communication dynamics in online discussions, we focus only on events that are exogenous to the classical OSNs dynamics and where we can have a detailed view on their evolution—e.g. salient periods can be easily identified from the news and other media.

In order to extract information topics from the tweets, for all the datasets, we only considered tweets with at least one hashtag (with the only exception of the Nepal earthquake dataset, where due to the extremely large size we had to restrict our analysis to only hashtags that have been posted at least 100 times). From each tweet, the information considered is: the timestamp of the tweet, the user-ID (previously anonymized) of the user who posted the tweet and all the hashtags present in it. In this way, we were able to extract users' interests and their evolution throughout the discussion, along with hashtags co-occurrences, fundamental to infer the information topics.

Figure 1 presents the time evolution of the number of tweets for the datasets considered. In all the cases, it is easy to see peaks in the activity, usually related to external events, like the television debate between candidates or the polling day for the Spanish elections or the referendum day for the Catalan referendum. These high activity periods are usually preceded and followed by calmer periods characterized by a lower and constant activity. Following this separation, we split each dataset into different parts, each corresponding to a 'peak' or 'rest' period, and study how users' interests and competition for attention varies between them. In particular, our aim in splitting the datasets is to have periods small enough to clearly distinguish between peaks, containing only a single event, but also to have enough data during the rest intervals. Thus, for each dataset, we identified the largest spike in activity and made its duration the size of the period. For example, for the Catalan referendum, since the largest spike lasts around 10 days, we decided to divide the dataset in intervals of exactly 10 days. We applied a similar reasoning for the Nepal earthquake dataset, in which the activity peak lasts 3 days so, we have 3 periods of 3 days each. However, for the Spanish elections dataset, due to its short duration and

Figure 2. Example of topics extracted from the Spanish general elections dataset. The word-clouds represent the most abundant hashtags for four selected topics detected with our method. Hashtags' size in each figure is proportional to the number of times it has been tweeted.
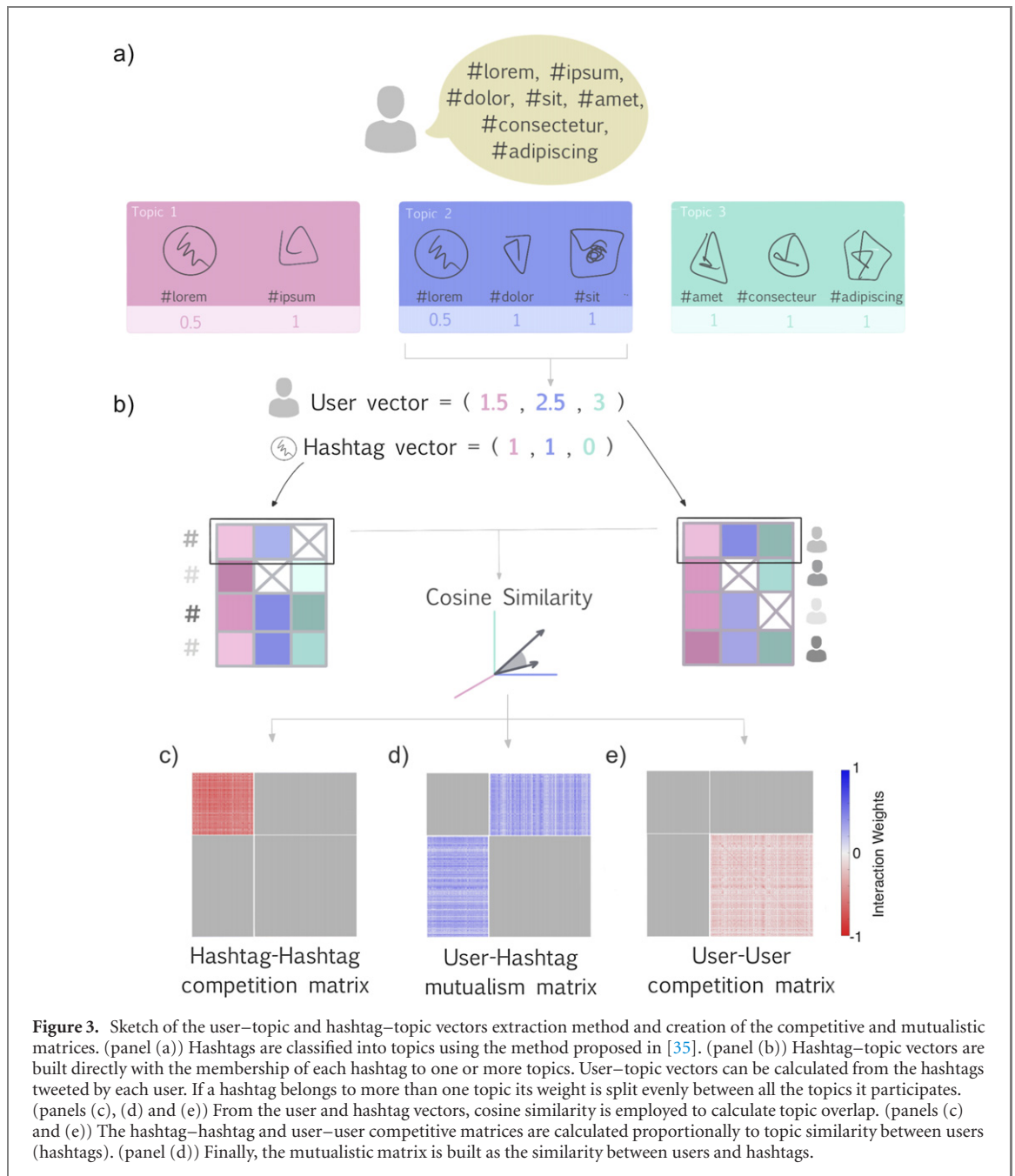
the effects of circadian rhythms, we were forced to have slightly uneven intervals to guarantee enough data during the resting periods.

Once the datasets and the different periods are defined, we focus on how to extract users' interests from the raw data. This step is fundamental to measure the similarity between users and, eventually, quantifying the competition for attention that both users and memes experience over the development of the events. To extract users' interests from their timeline, we rely on a network theory approach based on hashtags co-occurrence, designed specifically to infer information topics in Twitter [34–36]. Hashtags are used in many social media as keywords to indicate the content of a message and they often represent memes, whose meaning is known to the users. They thus provide a concise indication of a tweet's semantic context. In this way, the appearance of two or more hashtags in the same tweet is often the sign of a semantic association between them—they belong to the same topic—as it happens to words co-occurring frequently in texts [37, 38]. Based on this association, we extract information topics as cohesive clusters of hashtags that significantly appear together in different tweets.

To do so, for each period considered in the data, we build a weighted co-occurrence network between hashtags, where a link is laid if two hashtags appeared together in the same tweet and its weight represents the number of different tweets where they co-occurred. In order to eliminate spurious links and assure the significance of the semantic associations between the hashtags, we then remove all the links which weight is equal or smaller than 3. Once the networks have been built, information topics can be detected as clusters of densely connected hashtags—i.e. communities in the graph. Following the same procedure proposed in [35, 36], we thus employ the OSLOM tool [39] to detect communities in the different networks. We decided to use OSLOM for community detection because it is able to extract overlapping communities where nodes can belong to more than one community at a time. Thus, in our case, it is able to catch the fact that hashtags with different meanings can be part of several topics. Figure 2 represents an example of topics extracted from the Spanish elections dataset. In the four topics the names of the distinct national parties are clearly identifiable along with abbreviations of the most important dates.

Once the information topics are extracted from the data, we can use them to detect users' interests from the hashtags they posted, and their membership to the different topics. For each user, we build a feature vector **u** whose entries u_i quantify user's interest toward topic i. Specifically, u_i accounts for the number of times the user posted a hashtag belonging to topic i in their tweets. Moreover, since hashtags can be part of different topics, we split the weight of each hashtag according to the number of communities it belongs to. For example, a hashtag part of only one topic t will contribute to the pertinent u_t by adding one. On the contrary, one belonging to

Figure 3. Sketch of the user–topic and hashtag–topic vectors extraction method and creation of the competitive and mutualistic matrices. (panel (a)) Hashtags are classified into topics using the method proposed in [35]. (panel (b)) Hashtag–topic vectors are built directly with the membership of each hashtag to one or more topics. User–topic vectors can be calculated from the hashtags tweeted by each user. If a hashtag belongs to more than one topic its weight is split evenly between all the topics it participates. (panels (c), (d) and (e)) From the user and hashtag vectors, cosine similarity is employed to calculate topic overlap. (panels (c) and (e)) The hashtag–hashtag and user–user competitive matrices are calculated proportionally to topic similarity between users (hashtags). (panel (d)) Finally, the mutualistic matrix is built as the similarity between users and hashtags.

two topics will contribute 0.5 to each of the two corresponding entries of $\mathbf{u}$. In figures 3(a) and (b) we provide a diagram showing how the hashtags posted by the users are classified into topics and then used to build the user vector $\mathbf{u}$. The same procedure can be then applied to hashtags to build an equivalent hashtag vector $\mathbf{h}$ for each of them, accounting for their participation in the different topics.

User vectors quantify how individual attention is distributed across the different topics and how external events can focus it over a specific theme. However, individual interests are also a proxy of competition for attention. If everyone is interested in a specific topic, hashtags on that topic will experience a stronger competition for spaces in the users' screens. In the same way, users compete with their peers to get their messages read. On the other side, a user can decide to post about popular topics to have the advantage of reaching a potentially larger audience, leading to a sort of mutualistic interaction.

To quantify how much competition users experience in peak and resting periods in our datasets, we measure how similar the user vectors are one to the other. For each couple of user vectors $\mathbf{u}$ and $\mathbf{v}$ we calculate the cosine similarity:

$$\text{sim}_{\cos}(u, v) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}. \tag{1}$$

In this way, we have a measure that ranges from 0 (users with totally disjoint interests) to 1 (totally aligned users). Finally, hashtag competition can be estimated in the same way using the hashtag vectors **h** instead of **u** (see section 2.4 for details).

2.3. Dynamical model of users' attention

Defining users and hashtags similarity allows us to quantify the amount of competition and mutualism experienced by each node. However, to get further insights on the driving mechanisms behind these changes in collective attention, we also employ an ecology-inspired visibility optimization model proposed to explain structural changes in the user–hashtag interaction networks [21] over the course of an event.

Following the analogies between ecological communities [25, 26] and information ecosystems [21, 22], the model is based on the assumption that the attention dynamics recorded in the data is the result of an optimization process where users aim to maximize their visibility. Users and hashtags are thus seen as species of a mutualistic assemblage belonging to two different classes—e.g. plants and pollinators in natural ecosystems. Competition takes place between species of the same guild (user–user or hashtag–hashtag), while mutualistic interactions occur between species of different guilds (user–hashtag), with the species dynamics modeled by Lotka–Volterra equations with a Holling-type II functional response [24, 25]:

$$
\frac{dn_i^U}{dt} = n_i^U \left(\rho_i^U - \sum_j \beta_{ij}^{UU} n_j^U + \frac{\sum_k \gamma_{ik}^{UH} \theta_{ik}^{UH} n_k^H}{1 + h \sum_k \theta_{ik}^{UH} n_k^H} \right),
$$

$$
\frac{dn_i^H}{dt} = n_i^H \left(\rho_i^H - \sum_j \beta_{ij}^{HH} n_j^H + \frac{\sum_k \gamma_{ik}^{HU} \theta_{ik}^{UH} n_k^U}{1 + h \sum_k \theta_{ik}^{HU} n_k^U} \right),
$$

$$(2)$$

where n_i^U and n_i^H stand for the abundance (visibility) of species i part of the users' (U) or hashtags' (H) guild, while ρ_i^U and ρ_i^H represent the respective growth rates and h the handling time of the Holling-type II mutualistic functional response. The intensity of the users and hashtags competitive and mutualistic interactions is defined by matrices β^{UU}, β^{HH} and γ^{UH}, respectively. Finally, θ is the adjacency matrix of the mutualistic interactions accounting for the hashtags produced by each user in their posts.

The optimization process at the basis of the model is the same as the one proposed in Suweis *et al* [25], where species (users) rewire their mutualistic connections to randomly selected partners (hashtags) and the new links are kept only if they lead to an increase in abundance (visibility). Otherwise, the original connection is restored. Specifically, at constant time intervals, a random user U is selected and one of its existing connections to hashtag H is rewired to a new hashtag H'. The link to be rewired is selected with probability $p_{UH} \propto 1 - k_H^{-1}$ where k_H is the degree of H. After the rewiring, we let the system evolve until the equilibrium is reached. If at the end of this period U's abundance is greater than the previous one, the link is kept; otherwise, the previous configuration is restored.

It is important to notice that in our model only users maximize their abundance (visibility) since they choose the hashtags to post in their tweets. Thus, changes in the hashtag networks are due only to users' actions.

Finally, information topics are modeled as ecological niches [40] associated to each species. For the users, they represent the set of individual interests while, for the hashtags, they define their semantic context.

2.4. Estimating competition and mutualism from data

Once the species dynamics are defined, we need to estimate the interaction strength among species (β^{UU}, β^{HH} and γ^{UH}) from our data. To do so, we follow the approach proposed by Cai *et al* [26], where competition and mutualistic strengths are proportional to the niche overlap between species of the same (competition) and opposite (mutualism) guild. This way, we can estimate niche overlap $G_{ij}^{gg'}$ between species i belonging to guild g and species j of g' from the data as the cosine similarity between the topics vectors of i and j. Then, the elements of the competition matrices β_{ij}^{HH} and β_{ij}^{UU} are simply proportional to the ovelap: $\beta_{ij}^{HH} \propto \Omega_c \cdot G_{ij}^{HH}$ and $\beta_{ij}^{UU} \propto \Omega_c \cdot G_{ij}^{UU}$, with Ω_c a global factor to tune the absolute strength of the interaction. In a similar way, we can define the mutualistic strength as: $\gamma_{ik}^{UH} \propto \Omega_m \cdot G_{ik}^{UH}$, with the only difference that, in the model, the mutualistic strength γ_{ik}^{UH} is then multiplied by θ_{ik}^{UH} to include the fact that user i may ($\theta_{ik}^{UH} = 1$) or may not ($\theta_{ik}^{UH} = 0$) interact with hashtag k. Finally, to be able to compare the two strengths we impose $\Omega_c = \Omega_m$.

3. Results

We start our analyses by dividing our datasets in different periods, depending on their activity (i.e., the overall number of tweets produced). We thus distinguish between what we define as peak periods, when external events produce spikes in Twitter activity, such as the television debate day for the Spanish elections dataset
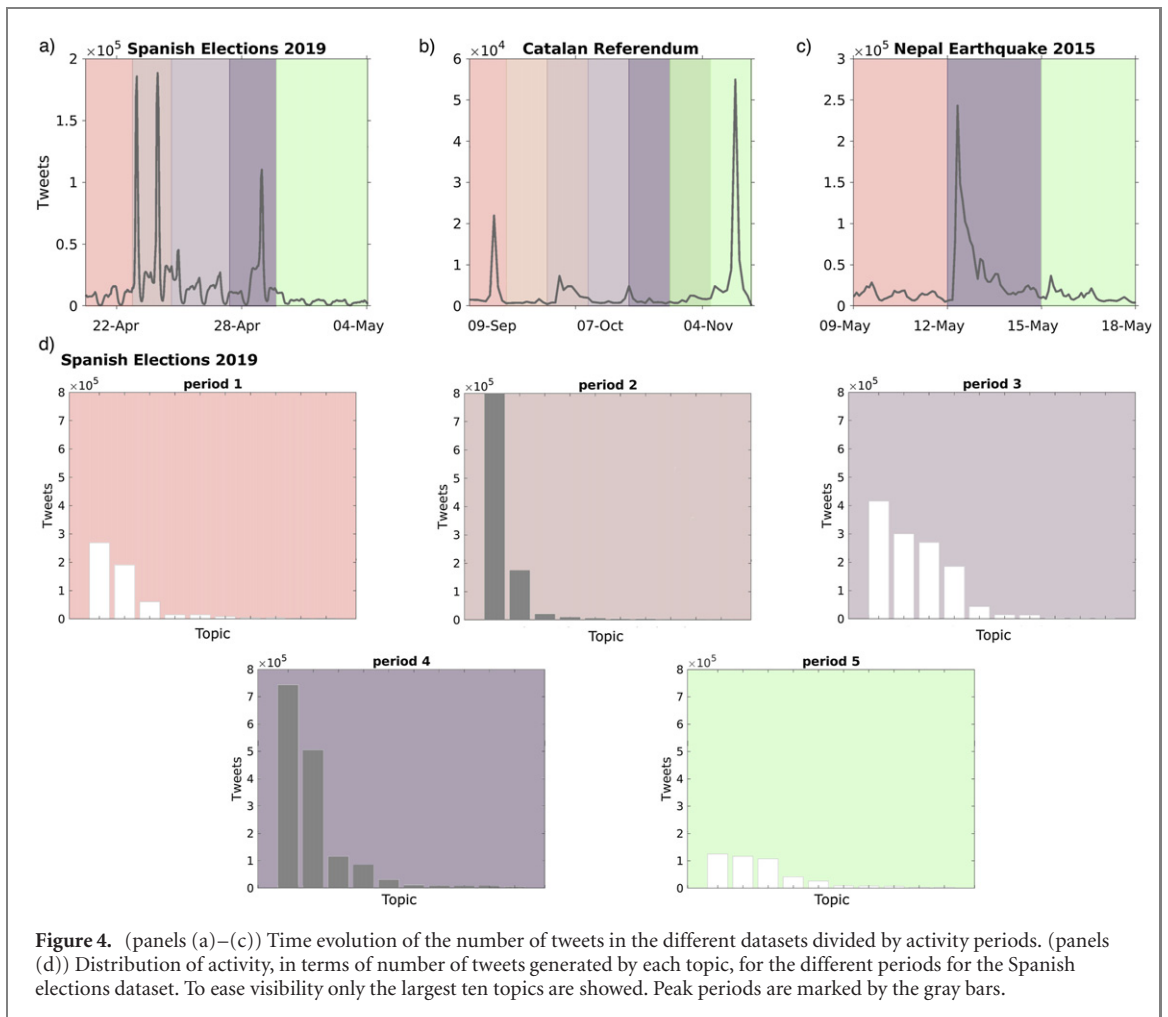
Figure 4. (panels (a)–(c)) Time evolution of the number of tweets in the different datasets divided by activity periods. (panels (d)) Distribution of activity, in terms of number of tweets generated by each topic, for the different periods for the Spanish elections dataset. To ease visibility only the largest ten topics are showed. Peak periods are marked by the gray bars.

or the referendum day for the Catalan dataset; and resting periods, where the discussion is still active but not driven by major external events. In this way, we can use the resting periods to define a baseline for activity and collective attention and compare them with the peak periods.

Figures 4(a)–(c) show the split of the three datasets we are considering. For the Spanish elections dataset (figure 4(a)) we define five different periods with the second and fourth covering large events, while the first and fifth represent our baseline. The third one is a mixture of peak and resting periods, since it is characterized by an increased activity but only smaller events took place. We adopt a similar split also for the Catalan self-determination referendum dataset. In this case, we defined seven periods, with the first and seventh as peaks, while the second, fourth and sixth as resting states. The remaining two (third and fifth intervals) are classified as mixed. Finally, in the Nepal earthquake dataset we identify three periods, where the first one represents the resting state, the second the highest activity and the third one a mix of peak and rest.

Once defined our baseline and active periods, we apply the topics extraction procedure to each period and study how the hashtags communities and users' interests evolve through time. To do so, for each period we compute the hashtag co-occurrence network and extract the most significant communities from it, defining our topics and hashtags' and users' membership as described in section 2.2.
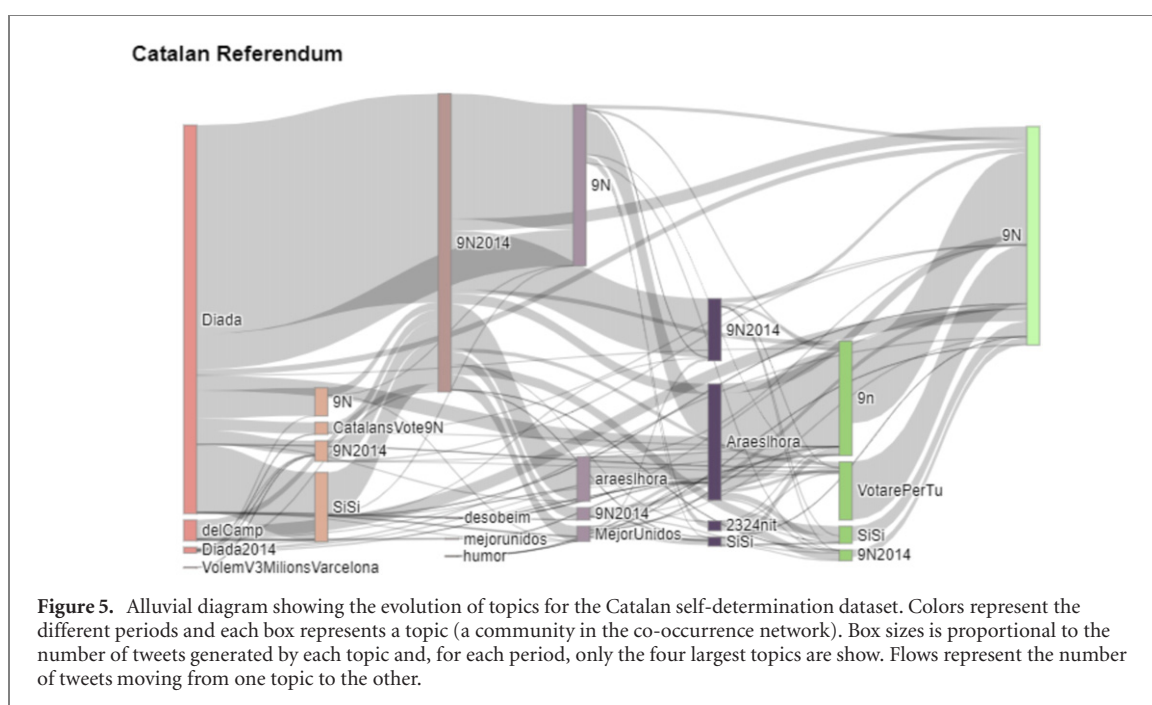
3.1. Topics evolution

From the networks it is possible to study how information topics evolve, and the corresponding discussions re-arrange around the external events. Table 2 summarizes the main features of the co-occurrence networks obtained from the data. A result that first catches the eye is the fact that, even if the networks for different periods have different sizes—e.g., the number of nodes/hashtags for the Catalan referendum dataset ranges from 310 to 1602 between the least and most active periods—the number of communities/topics is more or less constant. This contrasts with our assumption that, during relevant events, collective attention focuses only on one or a few 'important topics'. This discrepancy is solved if we consider not only the number of topics but also the activity around them, in terms of the number of tweets they generate. Figure 4(d) shows how tweets are distributed among the ten largest topics for the different periods of the Spanish elections dataset. Although activity in general is highly heterogeneous, however, during peak periods (marked by gray bars) this tendency

Table 2. Features of the co-occurrence networks extracted from each considered dataset.

Dataset	Period	# of hashtags	# of links	# of topics
Spa. elec.	1	1344	6866	75
	2	1714	11587	68
	3	1899	12026	77
	4	1736	12373	63
	5	1313	5114	60
Cat. ref.	1	703	3170	9
	2	310	863	20
	3	586	2070	10
	4	350	1037	21
	5	338	911	16
	6	362	1112	13
	7	1602	5837	12
Nepal	1	471	1564	29
	2	771	4739	27
	3	670	2921	33

Figure 5. Alluvial diagram showing the evolution of topics for the Catalan self-determination dataset. Colors represent the different periods and each box represents a topic (a community in the co-occurrence network). Box sizes is proportional to the number of tweets generated by each topic and, for each period, only the four largest topics are show. Flows represent the number of tweets moving from one topic to the other.
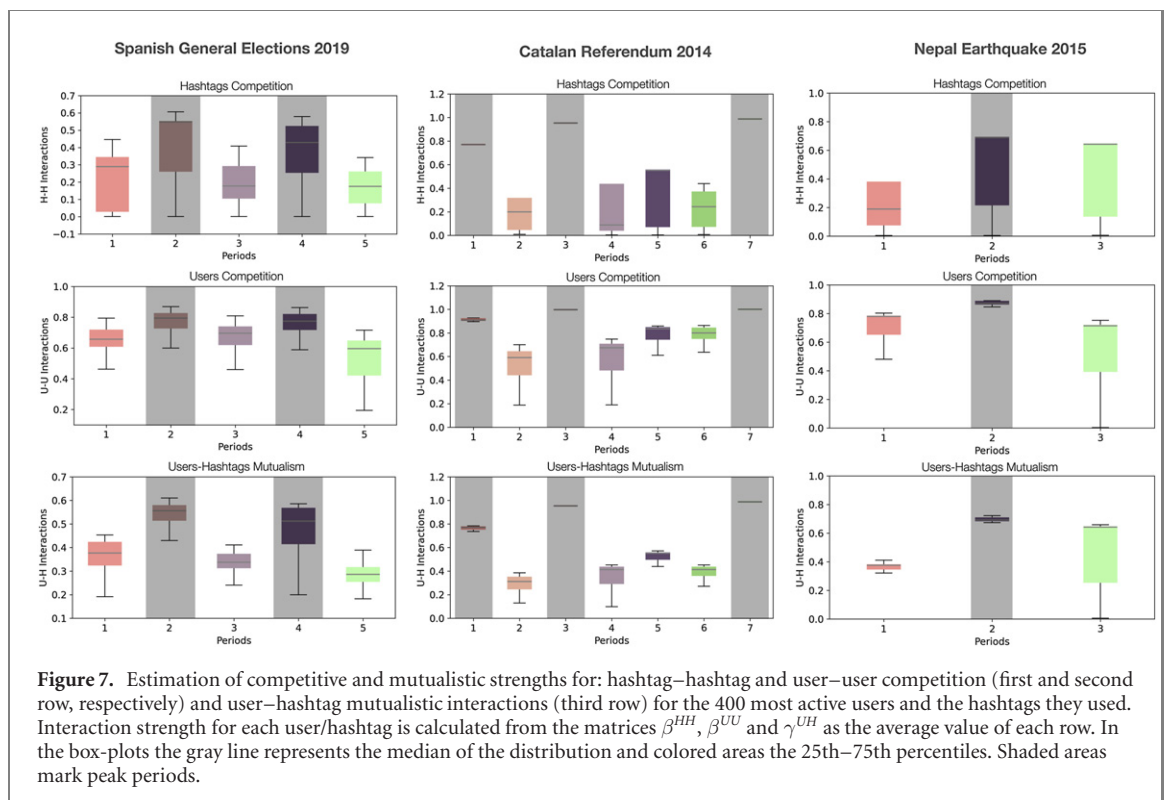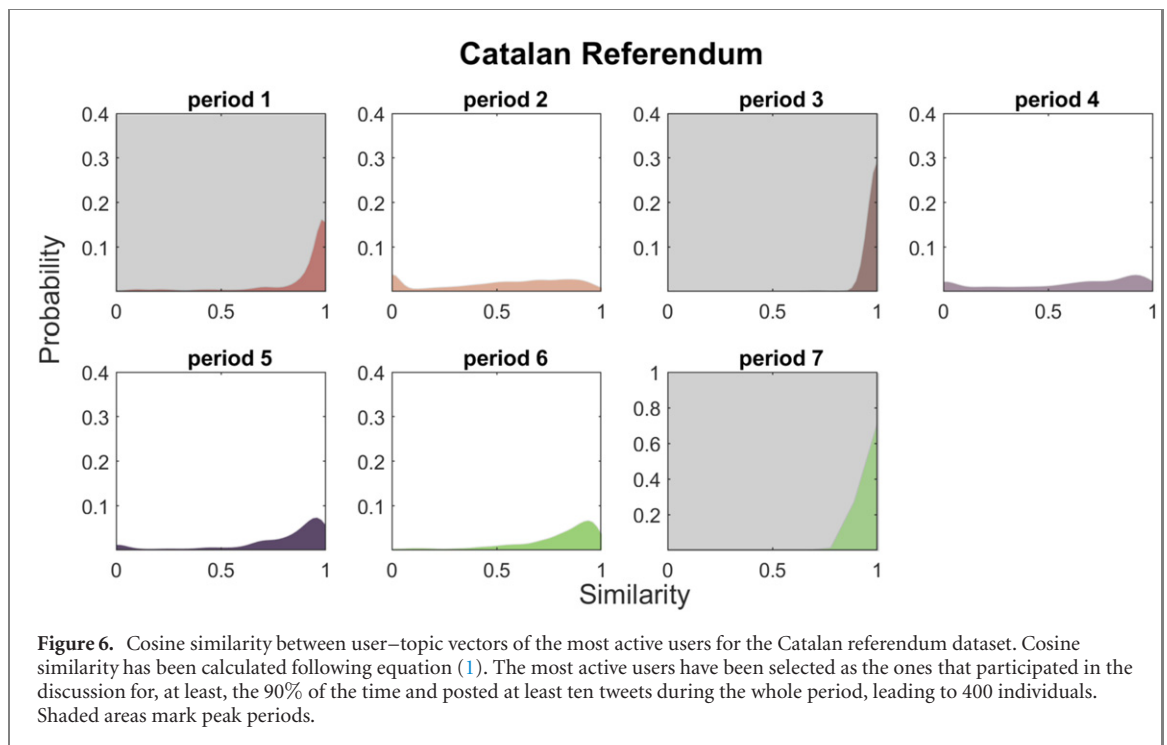
exacerbates with one or two topics generating almost the 90% of the tweets (we obtain similar results for the other two datasets, not shown).

Taken together, the last two results (i.e., the almost constant number of topics over different periods, as shown in table 2, and the higher heterogeneity in attention during peak events presented in figure 4(d)) depict a clear scenario. During normal activity periods, users' interests are focused on specific topics and hashtags compete for attention inside individual discussions. When an external event captures collective attention, original topics remain active but generate way less activity than before. To visualize how attention flows between the different topics over time, figure 5 shows an alluvial diagram of the activity around the topics in the Catalan referendum dataset (we obtain similar results for the other datasets, not shown). In each block (relative to a time period) the size, in terms of number of tweets, and the flows of activity for the four largest topics are shown. As a confirmation of the results in figures 4(c) and (d), during peak periods (first, third and seventh) all flows converge around one single topic. When external events fade out, flows split back to different discussions or, some hashtags stop to be used during calm intervals (e.g. the second period) to regain interest in high activity periods as shown, for example, by the large flows moving from the first period to directly the third.
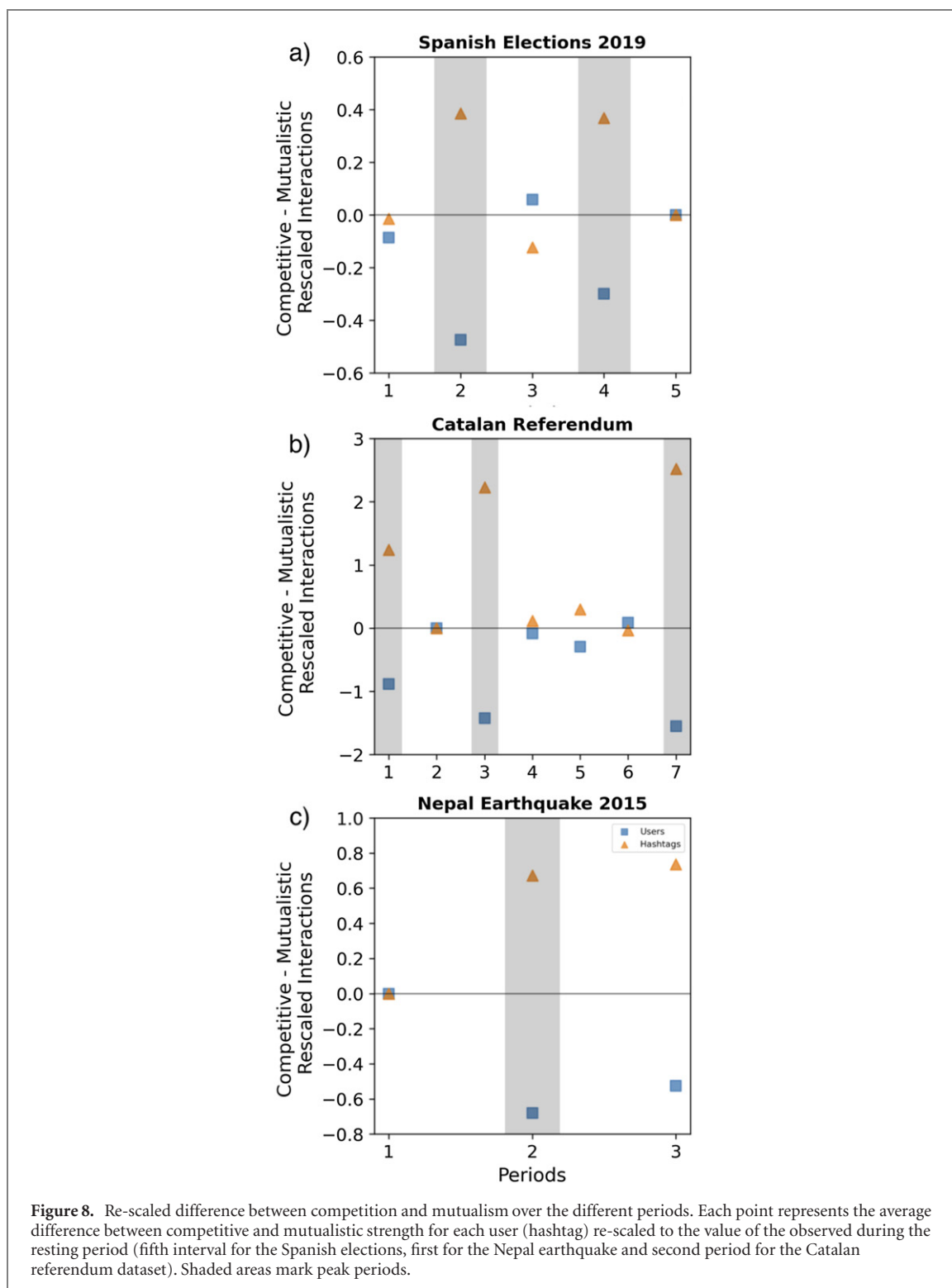
3.2. Users' similarity

Once studied how exceptional events shape hashtags dynamics, we can now focus on the other side of the coin: users' behavior. We use the cosine similarity between user–topic vectors to quantify interests overlap between

Figure 6. Cosine similarity between user–topic vectors of the most active users for the Catalan referendum dataset. Cosine similarity has been calculated following equation (1). The most active users have been selected as the ones that participated in the discussion for, at least, the 90% of the time and posted at least ten tweets during the whole period, leading to 400 individuals. Shaded areas mark peak periods.

Figure 7. Estimation of competitive and mutualistic strengths for: hashtag–hashtag and user–user competition (first and second row, respectively) and user–hashtag mutualistic interactions (third row) for the 400 most active users and the hashtags they used. Interaction strength for each user/hashtag is calculated from the matrices β^{HH}, β^{UU} and γ^{UH} as the average value of each row. In the box-plots the gray line represents the median of the distribution and colored areas the 25th–75th percentiles. Shaded areas mark peak periods.
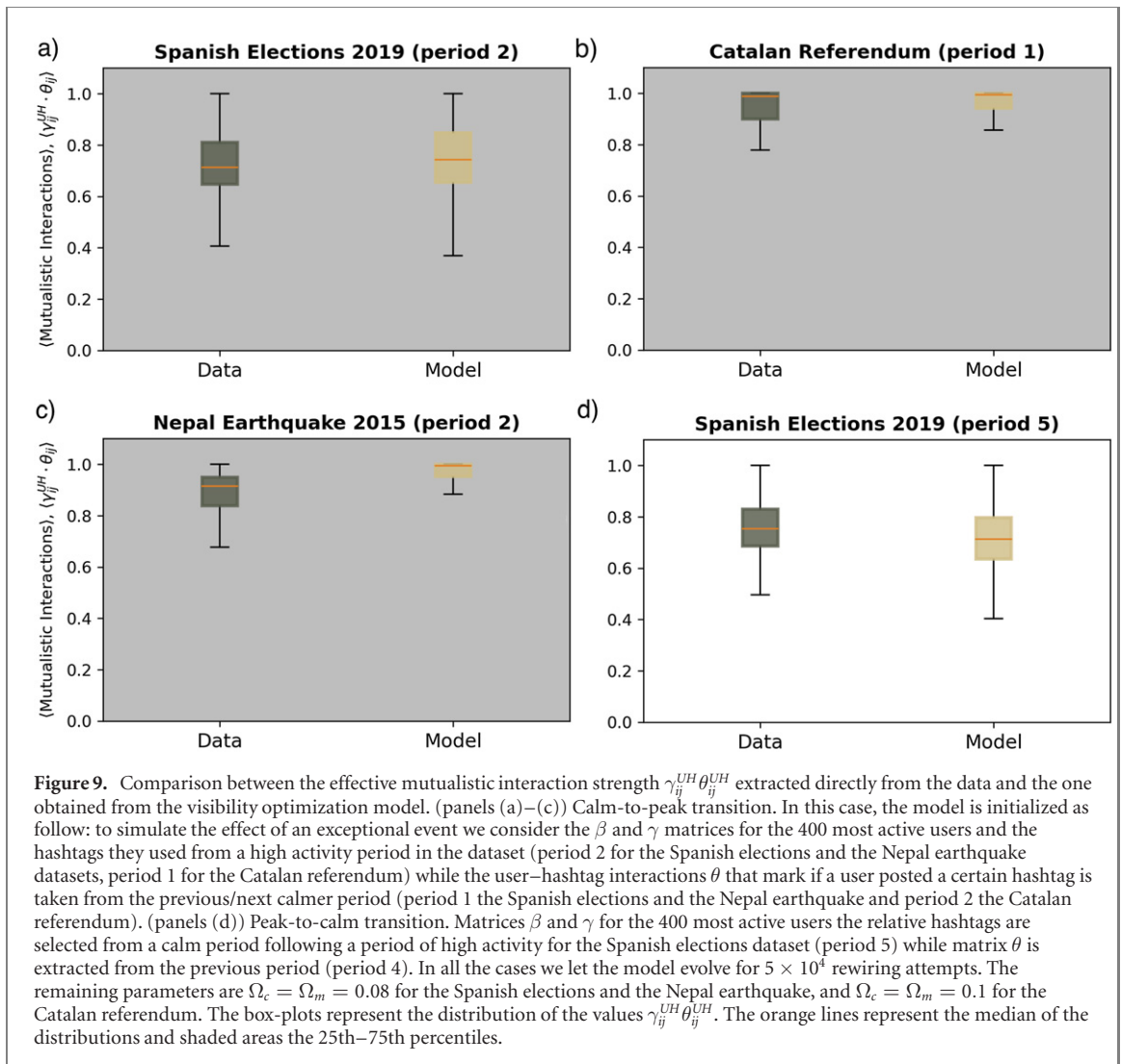
pairs of users. A higher average similarity will point to a focus of collective attention around few specific topics and, in turn, increased competition.

In figure 6 we present the distribution of similarity between all the possible pairs of the most active users (active at least 90% of the time and who posted at least ten tweets with hashtags for a total of 400 users) for the Catalan referendum dataset. As expected, the results match what we found for the hashtags. During low activity intervals (i.e. periods 2 and 4) users' similarity is practically uniformly distributed between 0 and 1 meaning that separated discussions are taking place, each with an—almost—equivalent activity. On the other side, high activity periods are characterized by an almost perfect alignment between users with the similarity distribution peaked around exactly one: everyone is talking about the same topic.

Figure 8. Re-scaled difference between competition and mutualism over the different periods. Each point represents the average difference between competitive and mutualistic strength for each user (hashtag) re-scaled to the value of the observed during the resting period (fifth interval for the Spanish elections, first for the Nepal earthquake and second period for the Catalan referendum dataset). Shaded areas mark peak periods.

3.3. Quantifying competition and mutualism during exceptional events

The results so far in figures 4–6 can be easily understood in terms of visibility maximization. In normal times, users tend to focus on their specific interests and aim for visibility inside their communities. Important events can captivate everyone's attention, leading users to join the conversation. This simple response driven by individuals' strive for visibility has some complex consequences, however. On one side, an increase in competition for attention. With a huge amount of messages covering one single topic, the probability for each user's ideas to be heard decreases. On the other side, the use of hashtags related to relevant topics can lead to a mutualistic (positive) interaction as they allow to reach a larger audience. Our goal is to quantify the strength of both these interactions to understand which one dominates the dynamics in the different periods. To do so, as detailed in section 2.4, from users' and hashtags' similarities we build three interaction networks—two for

Figure 9. Comparison between the effective mutualistic interaction strength $\gamma_{ij}^{UH}\theta_{ij}^{UH}$ extracted directly from the data and the one obtained from the visibility optimization model. (panels (a)–(c)) Calm-to-peak transition. In this case, the model is initialized as follow: to simulate the effect of an exceptional event we consider the β and γ matrices for the 400 most active users and the hashtags they used from a high activity period in the dataset (period 2 for the Spanish elections and the Nepal earthquake datasets, period 1 for the Catalan referendum) while the user–hashtag interactions θ that mark if a user posted a certain hashtag is taken from the previous/next calmer period (period 1 the Spanish elections and the Nepal earthquake and period 2 the Catalan referendum). (panels (d)) Peak-to-calm transition. Matrices β and γ for the 400 most active users the relative hashtags are selected from a calm period following a period of high activity for the Spanish elections dataset (period 5) while matrix θ is extracted from the previous period (period 4). In all the cases we let the model evolve for 5×10^4 rewiring attempts. The remaining parameters are $\Omega_c = \Omega_m = 0.08$ for the Spanish elections and the Nepal earthquake, and $\Omega_c = \Omega_m = 0.1$ for the Catalan referendum. The box-plots represent the distribution of the values $\gamma_{ij}^{UH}\theta_{ij}^{UH}$. The orange lines represent the median of the distributions and shaded areas the 25th–75th percentiles.

user–user (β^{UU}) and hashtag–hashtag (β^{HH}) competition, and one for user–hashtag mutualism (γ^{UH})—and analyze changes in nodes strength throughout time. Using niche overlap to estimate competitive and mutualistic strengths [26] has the advantage of re-scaling both terms making them comparable. This way we can quantify how much each one weighs in the dynamics of the system.

The box-plots in figure 7 show how strength is distributed in the different periods for the three datasets, and the 400 most active users and the hashtags they used. In all the cases we find an increase in both competition and mutualism during high activity events. However, there are differences between interaction types. Competition experienced by users is quite high even during low activity periods. The arrival of an external event, and the subsequent shift in attention, only slightly increases its strength. Hashtag competition instead is, in general, lower during resting periods, but shows a larger increase during events. Something similar happens with mutualistic interactions: with moderate values (with a median of around 0.4) for normal periods and sharp increases up to almost 1.0 in peak periods.

These differences, in both basal and aroused levels, lead to the question of which mechanism weights the most during the different periods. To answer it, in figure 8 we plot the average difference between competitive and mutualistic strength for each of the 400 users and associated hashtags, re-scaled to the value they experience during the resting periods. Interestingly, we find two opposite behaviors for users and hashtags. During the peak intervals, the effective competition perceived by the users decreases considerably as mutualistic strength increases more than competitive strength. This effect pushes users to actively participate in the discussion as they see that the ideas they post receive more attention. On the contrary, during events, hashtags on average experience an even stronger competition with respect to normal periods. This is probably due to the fact that hashtags not related to important topics will receive extremely low attention, while the ones affiliated with the major topics have to endure the increased pressure generated by users' interests.

3.4. Modeling attention switches during exceptional events

After quantifying the effective competition experienced by the agents, we conclude our analyses by trying to understand what are the major drivers behind the re-organization of the information topics we see in the data. To do that, we employ the visibility optimization model introduced in section 2.3. In particular, while in [21] the authors run the same model on synthetic data to reproduce the structural reorganization of the user–hashtag interaction networks during exogenous events, here we go one step further and inform the model with the matrices obtained directly from the events datasets. This way, we can test the hypothesis that visibility optimization, and thus the reduced perceived competition, is the main driver behind the focusing of collective attention.

We do so by running the model for two consecutive periods of our data: a calmer one followed by a peak one—e.g., periods 1 and 2 in the Spanish elections dataset—and then comparing the user–hashtag network optimized by the model, $\gamma_{ij}^{UH}\theta_{ij}^{UH}$ with the one extracted from the data. Specifically, to simulate the arrival of an external event, we initialize our simulations with matrices β^{UU}, β^{HH} and γ^{UH} from the peak period, but with the user–hashtag adjacency matrix θ from the previous, calmer one. In this way, θ represents the interests users had before the arrival of the event, while β and γ denote the push that the new topic exerts on the agents. We then let the model evolve, rewiring the connections in θ, until a steady state in species abundance has been reached (see the caption of figure 9 for the model parameters). At the end of the simulation, we compare the effective mutualistic interactions $\gamma_{ij}^{UH}\theta_{ij}^{UH}$ obtained from the data of the peak period with the one optimized by the model. In addition, we also check if visibility optimization is able to explain not only the focus of collective attention around a specific event, but also the opposite transition when, at the end of a peak period, users focus back on their original topics of interest. Consequently, we initialize the model with matrices β^{UU}, β^{HH} and γ^{UH} from a calm period (period 5 for the Spanish elections), and then optimize matrix θ from the previous peak period (period 4).

In figure 9, the box-plots show the comparison between the data and the model, for both the calm-to-peak transitions for all the three datasets (panels (a)–(c)) and the peak-to-calm (panel (d)) transition for the Spanish elections. In all the cases, we find a quite good agreement between matrices $\gamma_{ij}^{UH}\theta_{ij}^{UH}$, demonstrating that the optimization process alone reproduces the effective strength of the user–hashtag interaction networks extracted from the data. This further confirms our previous results, i.e., that users focusing on the trending topics during exogenous events experience a lower effective competition and, once the event ends, visibility optimization drives them back to their original interests.

4. Discussion and conclusions

The advent of OSNs changed the way in which we, as a society, process information. We moved from the classical one emitter—multiple receivers scheme typical of the mass media era to an intricate network where everyone is a producer and consumer at the same time. Thus, understanding the drivers that shape our collective behavior in this new scenario is of paramount importance to avoid distortions of our information processing such as the spreading of fake news or the formation of echo-chambers that can alter the democratic debate.

In this work, we built on the analogy between natural and information ecosystems to quantify the competition for attention experienced by the agents of an OSN during exceptional events and how they respond to it. In particular, employing a modeling framework based on ecological niche-theory, we were able to measure the intensity of competitive and mutualistic interactions between users and hashtags. Our results show that, although during the peak of the event, competition sensibly increases for both users and hashtags, it is also associated with a sharp increase in users' mutualism. Interestingly, for users—that, on average, already experience higher levels of competition—this translates into a positive gain in terms of interactions, effectively reducing the net competition. On the other side, this effect is limited for the hashtags that end up experiencing stronger competition. Moreover, those patterns seem to be universal as they do not depend on the nature of the event—political, social or natural—nor on the way tweets have been collected.

Our findings also have some profound implications in terms of information processing and collective behaviors as they bring light to the mechanisms behind the emergence of collective attention in online social media. In [21], employing the same visibility optimization model, the authors showed that exogenous events cause a structural change in the user–hashtag interaction network. Here, informing the model with real data, we understand why this new configuration is more advantageous for the users. The agreement between the data and the numerical simulations suggests that the seek for visibility drives users' behavior: individuals participating in online discussions tend to maximize their visibility inside theirs topics of interest leading to high levels of competition inside each topic. At the arrival of an event, the same visibility maximization invites users

to follow the dominating topic, increasing competition but also creating larger audiences. This reorganization, however, is stable because mutualistic advantages exceed the stronger competition.

Although the success of our modeling in shedding light on users' behavior is encouraging, it still represents an idealization of the complex dynamics taking place in online social systems. For example, we are implicitly assuming, as in the vast majority of ecological models, that mutualistic (competitive) interactions only take place between species of different (same) guilds. However, this assumption may overlook some second order dynamics, like intra-guild mutualism that has been recently discussed in the ecological modeling literature [41, 42]. In Twitter, these effects translate into positive interactions between users or hashtags, for example the boost in visibility obtained by one user when their messages are retweeted by another one or the co-occurrence of hashtags that allow each of them to reach larger communities. The inclusion of such mechanisms could help to understand the emergence of influential users and the endogenous dynamics of Twitter events. Finally, another simplifying assumption in our framework is that the interaction strength Ω is set equal for competitive and mutualistic interactions ($\Omega_c = \Omega_m$). This is a common assumption in ecological modeling to reduce the parameters space and, in our framework, it is needed to correctly compare the interaction matrices $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Nevertheless, inferring Ω from the data would allow for a better characterization of competition dynamics in OSNs.

In conclusion, we believe that our work demonstrates, once more, how an ecological approach to the study of information ecosystems can provide useful tools for understanding collective and social dynamics.

Acknowledgments

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: 10.17605/OSF.IO/J5QWX and https://figshare.com/articles/dataset/Twitter_event_datasets_2012-2016_/5100460.

Appendix A. Data and code availability

All the Twitter datasets are publicly available. The Catalan self-determination referendum and the Spanish national elections datasets can be downloaded from https://osf.io/j5qwx/ (accessed 20 September 2021). For the Catalan self-determination referendum, data were collected by selecting all the tweets containing at least one hashtag from a list of around 70 predetermined hashtags, or posted by one account from a list of approximately 50 users active in the referendum process and Catalan independence movements. A similar filtering was employed to collect the Spanish national elections dataset, where tweets were collected if containing at least one keyword from a total set of 300 relevant terms that could either be Twitter usernames or hashtags related to the electoral process such as candidates' names, parties or campaign hashtags. The 2015 Nepal earthquake dataset was firstly collected in [31] and is available at https://figshare.com/articles/dataset/Twitter_event_datasets_2012-2016_/5100460 (accessed 20 September 2021). In this case, the keywords used for filtering were: 'nepal', 'earthquake', 'nepalearthquake'.

Finally, the code for the abundance optimization model can be found at the COSIN3 group Github page: https://github.com/COSIN3-UOC/dynamical-niche-model (DOI 10.5281/zenodo.4555890).

ORCID iDs

Violeta Calleja-Solanas ⬤ https://orcid.org/0000-0001-7917-8984
Emanuele Pigani ⬤ https://orcid.org/0000-0002-0566-0759

María J Palazzi ⓘ https://orcid.org/0000-0002-4894-4110
Albert Solé-Ribalta ⓘ https://orcid.org/0000-0002-2953-5338
Samir Suweis ⓘ https://orcid.org/0000-0002-1603-8375
Javier Borge-Holthoefer ⓘ https://orcid.org/0000-0001-9036-8463
Sandro Meloni ⓘ https://orcid.org/0000-0001-6202-3302

References

[1] Weng L, Menczer F and Ahn Y-Y 2013 Virality prediction and community structure in social networks *Sci. Rep.* **3** 2522
[2] Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley H E and Quattrociocchi W 2016 The spreading of misinformation online *Proc. Natl Acad. Sci. USA* **113** 554−9
[3] Del Vicario M, Vivaldo G, Bessi A, Zollo F, Scala A, Caldarelli G and Quattrociocchi W 2016 Echo chambers: emotional contagion and group polarization on facebook *Sci. Rep.* **6** 37825
[4] Cota W, Ferreira S C, Pastor-Satorras R and Starnini M 2019 Quantifying echo chamber effects in information spreading over political communication networks *EPJ Data Sci.* **8** 35
[5] Lorenz-Spreen P, Mønsted B M, Hövel P and Lehmann S 2019 Accelerating dynamics of collective attention *Nat. Commun.* **10** 1−9
[6] Sherman L E, Hernandez L M, Greenfield P M and Dapretto M 2018 What the brain 'likes': neural correlates of providing feedback on social media *Soc. Cogn. Affect. Neurosci.* **13** 699−707
[7] Fareri D S and Delgado M R 2014 Social rewards and social networks in the human brain *Neuroscientist* **20** 387−402
[8] Malik A, Dhir A and Nieminen M 2016 Uses and gratifications of digital photo sharing on facebook *Telemat. Inform.* **33** 129−38
[9] Weng L, Flammini A, Vespignani A and Menczer F 2012 Competition among memes in a world with limited attention *Sci. Rep.* **2** 335
[10] Altman E, Kumar P, Venkatramanan S and Kumar A 2013 Competition over timeline in social networks *2013 IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining (ASONAM 2013)* (Piscataway, NJ: IEEE) pp 1352−7
[11] Gonçalves B, Perra N and Vespignani A 2011 Modeling users' activity on twitter networks: validation of Dunbar's number *PLoS One* **6** e22656
[12] Romero D M, Galuba W, Asur S and Huberman B A 2011 Influence and passivity in social media *Joint European Conf. Machine Learning and Knowledge Discovery in Databases* (Berlin: Springer) pp 18−33
[13] Borge-Holthoefer J *et al* 2011 Structural and dynamical patterns on online social networks: the Spanish may 15th movement as a case study *PLoS one* **6** e23883
[14] Omodei E, De Domenico M and Arenas A 2015 Characterizing interactions in online social networks during exceptional events *Front. Phys.* **3** 59
[15] Qiu X, Oliveira D F, Shirazi A S, Flammini A and Menczer F 2017 Limited individual attention and online virality of low-quality information *Nat. Hum. Behav.* **1** 0132
[16] Gleeson J P, Ward J A, O'sullivan K P and Lee W T 2014 Competition-induced criticality in a model of meme popularity *Phys. Rev. Lett.* **112** 048701
[17] Gleeson J P, O'Sullivan K P, Baños R A and Moreno Y 2016 Effects of network structure, competition and memory time on social spreading phenomena *Phys. Rev. X* **6** 021019
[18] Asur S, Huberman B A, Szabo G and Wang C 2011 Trends in social media: persistence and decay *5th Int. AAAI Conf. Weblogs and Social Media*
[19] Baumann F, Lorenz-Spreen P, Sokolov I M and Starnini M 2020 Modeling echo chambers and polarization dynamics in social networks *Phys. Rev. Lett.* **124** 048301
[20] Borge-Holthoefer J, Baños R A, Gracia-Lázaro C and Moreno Y 2017 Emergence of consensus as a modular-to-nested transition in communication dynamics *Sci. Rep.* **7** 41673
[21] Palazzi M J, Solé-Ribalta A, Calleja-Solanas V, Meloni S, Plata C A, Suweis S and Borge-Holthoefer J 2021 An ecological approach to structural flexibility in online communication systems *Nat. Commun.* **12** 1941
[22] Plata C A, Pigani E, Azaele S, Calleja-Solanas V, Palazzi M J, Solé-Ribalta A, Borge-Holthoefer J, Meloni S and Suweis S 2021 Neutral theory for competing attention in social networks *Phys. Rev. Res.* **3** 013070
[23] Bascompte J, Jordano P, Melián C J and Olesen J M 2003 The nested assembly of plant-animal mutualistic networks *Proc. Natl Acad. Sci.* **100** 9383−7
[24] Bastolla U, Fortuna M A, Pascual-García A, Ferrera A, Luque B and Bascompte J 2009 The architecture of mutualistic networks minimizes competition and increases biodiversity *Nature* **458** 1018−20
[25] Suweis S, Simini F, Banavar J R and Maritan A 2013 Emergence of structural and dynamical properties of ecological mutualistic networks *Nature* **500** 449−52
[26] Cai W, Snyder J, Hastings A and D'Souza R M 2020 Mutualistic networks emerging from adaptive niche-based interactions *Nat. Commun.* **11** 5470
[27] Leigh E R 1968 The ecological role of Volterra's equations *Some Mathematical Problems in Biology* (Providence, RI: American Mathematical Society)
[28] Hofbauer J and Sigmund K 1998 *Dynamical Systems and Lotka−Volterra Equations* (Cambridge: Cambridge University Press)
[29] Hubbell S P 2001 *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)* (Princeton, NJ: Princeton University Press)
[30] Azaele S, Suweis S, Grilli J, Volkov I, Banavar J R and Maritan A 2016 Statistical mechanics of ecological systems: neutral theory and beyond *Rev. Mod. Phys.* **88** 035003
[31] Zubiaga A 2018 A longitudinal assessment of the persistence of twitter datasets *J. Assoc. Inf. Sci. Technol.* **69** 974−84
[32] Lehmann J, Gonçalves B, Ramasco J J and Cattuto C 2012 Dynamical classes of collective attention in twitter *Proc. 21st Int. Conf. World Wide Web* (New York: ACM) pp 251−60
[33] Borge-Holthoefer J, Perra N, Gonçalves B, González-Bailón S, Arenas A, Moreno Y and Vespignani A 2016 The dynamics of information-driven coordination phenomena: a transfer entropy analysis *Sci. Adv.* **2** e1501158
[34] Weng L and Menczer F 2015 Topicality and impact in social media: diverse messages, focused messengers *PLoS One* **10** 1−17
[35] Cardoso F M, Meloni S, Santanchè A and Moreno Y 2019 Topical alignment in online social systems *Front. Phys.* **7** 58

[36] Mussi Reyero T, Beiró M G, Alvarez-Hamelin J I, Hernández L and Kotzinos D 2021 Evolution of the political opinion landscape during electoral periods *EPJ Data Sci.* **10** 31

[37] Turney P D and Pantel P 2010 From frequency to meaning: vector space models of semantics *J. Artif. Int. Res.* **37** 141−88

[38] Martinez-Romo J, Araujo L, Borge-Holthoefer J, Arenas A, Capitán J A and Cuesta J A 2011 Disentangling categorical relationships through a graph of co-occurrences *Phys. Rev.* E **84** 046108

[39] Lancichinetti A, Radicchi F, Ramasco J J and Fortunato S 2011 Finding statistically significant communities in networks *PLoS One* **6** 1−18

[40] Williams R J and Martinez N D 2000 Simple rules yield complex food webs *Nature* **404** 180−3

[41] Crowley P H and Cox J J 2011 Intraguild mutualism *Trends Ecol. Evol.* **26** 627−33

[42] Stanton M L 2003 Interacting guilds: moving beyond the pairwise perspective on mutualisms *Am. Nat.* **162** S10−23