Abstract

Ego networks have proved to be a valuable tool for understanding the relationships that individuals establish with their peers, both in offline and online social networks. Particularly interesting are the cognitive constraints associated with the interactions between the ego and the members of their ego network, whereby individuals cannot maintain meaningful interactions with more than 150 people, on average. In this work, we focus on the ego networks of journalists on Twitter, considering 17 different countries, and we investigate whether they feature the same characteristics observed for other relevant classes of Twitter users, like politicians and generic users. Our findings are that journalists are generally more active and interact with more people than generic users, regardless of their country. Their ego network structure is very aligned with reference models derived from the social brain hypothesis and observed in general human ego networks. Remarkably, the similarity is even higher than the one of politicians and generic users ego networks. This may imply a greater cognitive involvement with Twitter than with other social interaction means. From a dynamic perspective, journalists have stable short-term relationships that don’t change a lot over time. On the longer term, though, ego networks can be pretty dynamic, especially in the innermost circles. Moreover, the ego-alter ties of journalists are often information-driven, as they are mediated by hashtags. Finally, we observe that highly popular journalists tend to engage with other journalists of similar popularity, and vice versa. Instead, when interacting with generic users, their popularity doesn’t seem to play a major role.

Keywords: online social networks, ego networks, Twitter, journalists
1. Introduction

Online Social Networks (OSNs) have been shaping and transforming our daily lives for more than a decade. Currently, there are 4.14 billion active social media users and 4.08 billion active mobile social media users\(^1\). With easy access to the Internet, especially via mobile devices, we started to carry our new communication mediums and online social groups everywhere we go. This lets us be a part of online communities more than ever and increases the phenomenon known as cyber-physical convergence, whereby our offline and online lives become tightly intertwined. The huge volume of information produced online across different OSN platforms gives researchers new opportunities to analyze human behaviors on large-scale data, overcoming the limitations of survey-based traditional approaches. Thus, not only studies on human behavior in the offline world via OSN data but also the comparison between offline and online human behavior gained popularity. On the one hand, we are transferring and engaging with our offline relationships on online platforms. On the other hand, we are establishing new relationships, which may not have been even possible in the offline world (due, e.g., to long distances and different time zones). As a result, researchers tried to understand if and to which extent our offline and online relationships are similar. The findings showed that the structure of online relationships mirrors that of offline ones [1].

OSNs not only have shaped how we communicate socially with each other but they have also redefined professional interactions, especially for jobs that have a wide audience, such as politicians, artists, and journalists. As for the latter, social platforms have become pools of news sources, leading to instant access to news. Journalism has been under transition to keep up with the new communication medium. Twitter has become one of the main access points to news resources. According to the survey carried by the American Press Institute in collaboration with Twitter\(^2\), 86% of Twitter users engage with the platform for news. As a result, Twitter has become extremely popular among journalists to share the news articles they produce, which makes journalists an interesting community to analyze.

The studies in the literature about journalists show that they use Twitter

\(^1\)https://www.statista.com/statistics/617136/digital-population-worldwide/
\(^2\)https://www.americanpressinstitute.org/publications/reports/survey-research/how-people-use-twitter-news
as a platform for establishing their personal brands [2], and promote content from their news websites [3]. Beyond this, the profession itself as sourcing, gatekeeping, verifying, and broadcasting news has been under transition from traditional methods to new communication mediums [3, 4, 5]. Thus, journalists started to become individual brands, news and opinion hubs [4], sometimes even more popular than the media companies they work for. There are many studies analyzing this transformation of the profession [3, 4, 5], as well as several comparative studies on journalism. In [6], journalists from the UK and 10 other European countries are compared in terms of their role perception and their belief in the usefulness of the Internet. Another study [7] analyzes Latin American journalists’ view on their professional role in today’s digital media platforms and how they are organized to create a new way of content management. In [8], cross-cultural research has been carried out to assess the development of journalism culture in 66 countries.

The above studies are mainly carried out via traditional survey-based data collection, and try to understand journalism culture and journalists’ views on the transformation of the profession. In this work, we set out to investigate the nature and structure of the relationships journalists entertain on Twitter directly by studying their interactions on the platform. To this aim, we leverage a graph-based abstraction known as *ego network*. The latter describes the relationships between an individual (ego) and its peers (alters) [9, 10, 11, 12]. Grouping these relationships by their strength, a layered structure emerges in the ego network (Figure 1), with the inner circles containing the socially closest peers and the outer circles the more distant relationships. This structure originates from the limited cognitive resources that humans are able to allocate to *meaningful* socialization (i.e., beyond the mere level of acquaintances), as discussed in detail in Section 2. In this model, five layers exist within the limit of the Dunbar number, which is the maximum number of social relationships (around 150) that an individual, on average, can actively maintain [12, 13]. Beyond the Dunbar number, relationships are just acquaintances and their maintenance has a negligible effect on the cognitive resources of the ego. The ego network is an important abstraction, as it is known that many traits of social behaviour (resource sharing, collaboration, diffusion of information) are chiefly determined by its structural properties [14]. To the best of our knowledge, there are no studies on how journalists use their cognitive resources to maintain their relationships on Twitter, and comparative models of ego network-based behavioral characteristics. In this paper, we study journalists from different countries.
and continents in order to highlight similarities and invariants in their social
behaviour online, and to understand how journalists allocate their cognitive
resources among their colleagues and non-colleagues alters.

![Figure 1: Layered structure of human ego networks](image)

The key findings presented in this study are the following:

• Journalists that mostly rely on replies (such as those from Finland, The Netherlands, and Denmark) have a below-average number of alters, thus suggesting a tendency to get involved with the same group of people. This might hint at the fact that replies are a more personal/intimate communication with respect to mention and retweets, hence they consume more cognitive resources on the ego side, which, in turn, is able to interact with fewer people. Vice versa, the journalists in countries where retweets and mentions are predominant tend to interact with above-average distinct peers, hinting at the opposite effect.

• Journalists establish numerous relationships, much more than generic Twitter users. However, when we consider only those relationships that are active, their ego network size (119 alters, on average) becomes very close to the Dunbar’s number of 150. The social circle hierarchy of journalists also mirrors that observed for online networks, with the optimal number of social circles around 5.

• From a dynamic perspective, journalists have a stable short-term relationships that don’t change a lot over time. They tend to keep their most and least intimate relationships as they are through the time while they change the average intimate relationships much more. On the longer term, though, ego networks can be pretty dynamic, especially in the innermost circles. This is in contrast with the findings
about generic Twitter users [15] and suggests that the ego networks of journalists, while structurally similar to that of generic users, may be affected by the information-driven nature of journalist engagement on the platform, thus yielding to much more variability in the composition of rings.

- Journalists are topic-driven users. They establish their relationships through hashtags. A relationship is hashtag-activated if the first contact of the relationship includes a hashtag. Journalists have more hashtag-activated relationships than politicians and generic Twitter users. In addition, they tend to utilize hashtags for hashtag-activated relationships more than for the relationships that are not activated by hashtags. The percentage of hashtag usage is higher in inner circles that include more intimate relationships.

- Journalists are not selecting their journalist alters randomly. They are allocating more cognitive resources to the colleague/journalist alters who have similar or higher popularity (number of followers) with/than themselves. However, popularity is not a parameter when they are establishing non-colleague relationships.

The paper is organized as follows. In Section 2, we describe the Dunbar’s ego network model and how we construct it. In Section 3, we describe how we collected the datasets, how we cleaned and preprocessed them to determine which journalists are suitable for ego network analysis. In Section 4, we show the results of ego network analysis in four sub-sections with similarities and differences between journalists from different regions in addition to comparison with politicians and generic Twitter users.

2. The ego network model

As discussed in Section 1, ego networks are graph-based abstractions used to study the relationships between a tagged individual (called ego) and its peers (alters). Specifically, an ego network is a local subgraph consisting of the ego (in the center) and ties/links that connect the ego to its alters. The strength of these links quantifies the emotional closeness/intimacy of the ego-alter relationship, and it is commonly measured in the related literature through the contact frequency. When the ego-alter ties are grouped together [16, 1] based on their strength, intimacy layers emerge, as shown in
Figure 1. The typical sizes of each layer are 1.5, 5, 15, 50, 150, respectively, where inner circles include alters with higher emotional closeness (higher contact frequency), and this closeness decreases through outer circles. The size (150 alters) of the outermost circle is known as Dunbar’s number. It represents the maximum number of relationships we can actively maintain [12, 13]. An active relationship is defined as one having at least one contact per year (see Sec. 2.1 for more details). The relationships beyond the fifth circles are not active (just acquaintances), since we do not use our cognitive resources regularly to maintain them. This finite-size layered structure is the result of our limited information-processing capacity, and this finding from anthropology goes under the name of social brain hypothesis [16]. Since we have limited cognitive capacity and limited time, we try to optimize how we allocate our cognitive resources among our relationships [14]. The result of this optimisation is the ego network structure. An important structural invariant of ego networks is their scaling ratio, defined as the ratio between the sizes of consecutive layers. Its value is typically around 3 in the offline and online networks studied in the related literature [1] [13].

The existence of Dunbar’s ego network structure has been shown in different communication means in the offline world, including face-to-face interactions [17], letters (Christmas cards) [12], phone calls [18]. Recently, their occurrence has been confirmed also for online interactions in OSNs [1, 19], showing that the same cognitive capacity that limits our social interactions exists also in OSNs (Facebook, Twitter). In this sense, OSN become one of the outlets that is taking up the brain capacity of humans, and thus are subject to the same limitations that have been measured for more traditional social interactions, but are not capable of “breaking” the limits imposed by cognitive constraints to our social capacity. Tie strengths and how they determine ego network structures have been the subject of several additional work. For example, in [20] authors provide one of the first evidences of the existence of an ego network size comparable to the Dunbar’s number in Twitter. The relationship between ego network structures and the role of users in Twitter was analysed in [21]. In general, ego network structures are also known to impact significantly on the way information spreads in OSN, and the diversity of information that can be acquired by users [22]. In [19], the ego network structure of politicians and how they allocate their cognitive resources among the alters have been studied. In the study, they showed the similarities between politicians’ ego networks and Dunbar’s model, and differences between generic users’ ego networks by using Twitter data [1].
In addition to the characterization of the ego networks, the effect of ego network structure on information diffusion has also been investigated in the literature [23].

2.1. How to extract the ego network

Dunbar’s ego network model is based on understanding how an individual allocates their cognitive resources among their relationships. We do not allocate our cognitive resources homogeneously across all the relationships that we establish. Instead, we spend them mostly on the ones we try to keep active in our lives. In [12], an active relationship has been defined as one where at least one contact per year occurs. This idea comes from the Christmas card exchange tradition of Western societies where people put effort to contact at least once a year the people they value the most. In the related studies about online ego networks [15, 1, 19], the concept of active relation has been adapted to the online domain by calculating the frequency of direct contacts and labelling as inactive those relationships with average contact frequency smaller than one direct tweet per year (in analogy with the Christmas card exchange in offline networks). In Twitter, a direct contact is a tweet (retweet, reply, mention) where a user directly interacts with another user. Based on direct contacts, we calculate the contact frequency for a relationship between ego \( i \) and alter \( j \) as follows:

\[
    w_{ij} = \frac{N_{\text{reply}} + N_{\text{mention}} + N_{\text{retweet}}}{L_R}.
\]

where \( w_{ij} \) is contact frequency (which is a proxy for the intimacy between the ego and the alter), \( N_{\text{reply}}, N_{\text{mention}}, N_{\text{retweet}} \) are the total number of reply, mention, and retweet between the ego \( i \) and the alter \( j \), and \( L_R \) is the observed duration (in years) of the relationship between \( i \) and \( j \), calculated as the time interval between their first direct tweet until the download date of the dataset. In agreement with the related literature, we filter out the relationships where \( L_R \) is smaller than one year, since these recently established relationships may not have yet stabilized.

An active ego network of an individual consists of the relationships for which the average contact frequency \( (w_{ij}) \) is bigger than 1 direct tweet per year. In order to extract the intimacy layers as in Figure 1, these relationships must be clustered into groups. Each layer includes relationships that have similar intimacy with the ego. The number of layers, and the alters located in each of them, can be determined by means of clustering algorithms. Similarly
to [24], in this work we use the Mean Shift algorithm [25], which is able to automatically identify the optimal number of clusters without additional interventions (e.g., k-means requires using the silhouette score or the elbow method to this purpose). Thus, for each ego in our datasets, starting from the active ego network, we derive the optimal number (typically around 5 in the related literature) of intimacy layers and the alters assigned to them. Then, we can easily compute the ego network scaling ratios as the ratio between the sizes of consecutive layers (recall that this is typically an invariant, whose value is around 3).

The ego networks obtained as described above are typically referred to as static ego networks, and they will be studied in Section 4.1 for the journalists in our datasets. The static view of an ego network considers a single, aggregate snapshot of the ego and its ties. By modelling the ego-alter ties through a static ego network, we consider all the communications between the ego and the alter. This analysis provides an essential starting point for understanding the ego-alter interactions and for making quantitative comparison with other datasets studied in the related literature. At the opposite spectrum, the dynamic analysis of ego networks focuses on their evolution over time. Specifically, snapshots of one year are considered, each being shifted forward by one month with respect to the previous one. For each snapshot we focus on the alters that are members of each ring, where a ring is defined as the portion of circle that excludes its inner circles. If we denote the set of alters in consecutive circles $i$ and $i + 1$ as $C_i$ and $C_{i+1}$, respectively, ring $R_{i+1}$ is defined as $C_{i+1} - C_i$. Conventionally, $R_1$ is equivalent to $C_1$. Egos for which we observe less than two years of tweets are filtered out, as they are not observed long enough to extract robust results. The dynamic analysis of the journalists’ ego networks will be discussed in Section 4.2.

With dynamic ego networks, the intimacy between the ego and alter may decrease between two consecutive time intervals, so the alter may change its location from inner rings to outer rings. Alternatively, the intimacy may increase or stay the same, so the alter may be included in an inner ring or preserve its current ring. In order to capture the stability of rings (in terms of membership) over time, we use the Jaccard similarity, defined as the ratio between the cardinality of the intersection and that of the union set of the alters belonging to ring $i$ in consecutive time intervals. We calculate the
Jaccard index $Jaccard_i$ of ring $i$ as follows:

$$Jaccard_i = \frac{1}{T} \sum_{t=1}^{T-1} \frac{\mathcal{R}_i^{(t)} \cap \mathcal{R}_i^{(t+1)}}{\mathcal{R}_i^{(t)} \cup \mathcal{R}_i^{(t+1)}}$$

(2)

where $T$ denotes the total number of consecutive snapshots. The closer this index to 1, the better the overlapping. In case of no overlap, the Jaccard index is equal to zero. The amount of movements between rings is measured through the Jump index, which simply counts (and then averages across all alters in the ring) the number of jumps between rings. This index can be computed as follows:

$$Jump_i = \frac{1}{T} \sum_{t=1}^{T-1} \sum_{j \in \mathcal{R}_i} \Delta_{j,t+1}$$

(3)

where we denote as $\Delta_{j,t+1}$ the number of jumps of alter $j$ between snapshot $t$ and $t+1$. For example, if the alter moves from $\mathcal{R}_1$ at $t$ to $\mathcal{R}_4$ at $t+1$, then $\Delta_{j,t+1}$ will be equal to 3. Of course, it is possible that, at $t + 1$, an alters enters the active network from the outside. In this case, we assume that it comes from the last plus one ring: e.g., for an ego with $n$ social circles, the alters is assumed to come from the $(n+1)$-th ring.

The notation we use in the paper is summarised in Table 1.

3. The dataset

We downloaded the Twitter timelines of journalists belonging to 17 different countries from 8 different continental regions. We classify geographical regions (continents and continental regions) by using UN M49 (standard country or area codes for statistical use\(^3\)). The countries, which can be seen in Figure 2, were selected based on the availability of journalists lists on Twitter. Twitter lists are collections of Twitter accounts that are included in a list by Twitter users with the aim of having a pool of accounts, generally with popular, that are related to a specific topic of interest. For our study, we manually looked up lists including journalists from the same country. The Twitter lists that we used as seeds can be found in Appendix A.1. For some

\(^3\)https://unstats.un.org/unsd/methodology/m49/
Table 1: Summary of notation (for a tagged ego)

| Name                      | Notation | Definition/formula                                                                 |
|---------------------------|----------|-----------------------------------------------------------------------------------|
| Active network            | $\mathcal{A}$ | ego-centered subgraph with weights greater than 1                                  |
| Optimal number of circles | $\tau$   | the results of the clustering on the edges in the active network $\mathcal{A}$       |
| Social circle (or layer)  | $\mathcal{C}_i$ | $i$-th social circles of the tagged ego, with $i \in \{1, \ldots, \tau\}$          |
| Scaling ratio of layer $i$| $\rho_i$ | $\frac{|\mathcal{C}_i|}{|\mathcal{C}_{i-1}|}$ with $i \in \{2, \ldots, \tau\}$       |
| Ring                      | $\mathcal{R}_i$ | $\mathcal{C}_i - \mathcal{C}_{i-1}$                                              |
| Jaccard index of ring $i$ | $Jaccard_i$ | $\frac{1}{T} \sum_{t=1}^{T-1} \frac{\mathcal{R}^{(t)}_i \cap \mathcal{R}^{(t+1)}_i}{\mathcal{R}^{(t)}_i \cup \mathcal{R}^{(t+1)}_i}$ |
| Jump index of ring $i$    | $Jump_i$  | $\frac{1}{T} \sum_{t=1}^{T-1} \sum_{j \in \mathcal{R}_i} \Delta_j^{t,t+1}$            |

of the countries, we found more than one relevant list, and in this case the lists are merged into one. Our dataset contains a different number of journalists for each country, as it can be seen in Table 2 (under the before filtering column). As a result of the Twitter API limitations, only the 3200 most recent tweets of each journalist (with a public profile) have been downloaded, and timelines without any tweeting activity were removed from the dataset.
Table 2: General statistics of the datasets.

| continent  | region            | dataset          | before filtering | after filtering | fully observed(%) / partially observed(%) |
|------------|-------------------|------------------|-------------------|-----------------|------------------------------------------|
|            |                   |                  | profiles (#)     | tweets (#)      | profiles (#)                             | tweets (#)                             |                           |                           |
| Americas   | Northern America  | USA              | 1,722            | 4,671,922       | 617                                       | 1,927,183                               | 10.86 / 89.14             |
|            |                   | Canada           | 917              | 2,462,998       | 427                                       | 1,281,149                               | 20.14 / 79.86             |
| South America | Brazil         |                  | 897              | 2,489,941       | 217                                       | 1,236,620                               | 13.82 / 86.18             |
| Asia       | Eastern Asia     | Japan            | 566              | 1,361,392       | 192                                       | 521,676                                 | 34.90 / 65.1              |
|            |                   | Turkey            | 2,189            | 5,495,545       | 731                                       | 2,203,729                               | 14.64 / 85.36             |
| Europe     | Northern Europe  | UK               | 512              | 1,330,467       | 235                                       | 717,341                                 | 14.47 / 85.53             |
|            |                   | Denmark           | 3,694            | 3,725,084       | 656                                       | 1,671,441                               | 50.00 / 50.0              |
|            |                   | Finland           | 936              | 1,477,304       | 321                                       | 824,823                                 | 46.11 / 53.89             |
|            |                   | Norway            | 1,190            | 1,504,906       | 201                                       | 584,251                                 | 28.86 / 71.14             |
|            |                   | Sweden            | 761              | 1,702,971       | 275                                       | 847,196                                 | 12.36 / 87.64             |
| Southern Europe | Greece       |                  | 862              | 1,858,442       | 265                                       | 778,403                                 | 24.15 / 75.85             |
|            |                   | Italy             | 486              | 1,301,858       | 255                                       | 781,415                                 | 46.47 / 53.53             |
|            |                   | Spain             | 468              | 1,309,721       | 195                                       | 611,233                                 | 10.26 / 89.74             |
| Western Europe | France        |                  | 579              | 1,525,660       | 323                                       | 975,941                                 | 21.98 / 78.02             |
|            |                   | Germany           | 405              | 879,071         | 146                                       | 415,560                                 | 30.14 / 69.86             |
|            |                   | Netherlands       | 4,303            | 9,621,422       | 1,674                                     | 4,859,433                               | 27.18 / 72.82             |
| Oceania    | Australia and New Zealand | Australia | 956              | 220,497        | 400                                       | 1,153,010                               | 29.50 / 70.5              |
| All journalists | mean ± sd     |                  | 1,260.18         | 2,647,860.24    | 419.41                                    | 1,224,831                               | 24.87 / 75.13             |

3.1. Observability

Since the Twitter API only discloses the last 3200 tweets of any public Twitter account, we were not able to download the full timelines of journalists having more than 3200 tweets. We refer to these accounts as partially observed, since we can only observe a portion of their timelines instead of their full Twitter activity since their registration to the platform. Across our datasets, on average 75.13% of the users are partially observed (i.e., they posted more than 3200 tweets, see Table 2). We were able to fully observe only 23.7% of the users (Table 2). Please note that being partially or fully observed is an indirect measure of the user’s tweeting frequency: the more the tweets posted, the quicker the 3200 slots are saturated. Indeed, the average daily tweeting frequency, across all journalists datasets, is 0.5 tweets/day for the fully observed users and 6.1 tweets/day for the partially observed ones (note that this number is fully compatible with human activity). The observability of the journalists’ timelines changes among the different countries: most of the datasets have at least 66% of the users that are partially observed, except for the Danish and Finnish journalists where approximately
half of the users have partially observed timelines.

We define the observed timeline as the portion of the user’s timeline we can access via the Twitter API. If the 3200-tweets API limitation is not hit, the observed timeline is the same as the active timeline, which instead covers the user activity starting from the date the user registered to Twitter until the dataset was downloaded. Since the coverage of the timeline is based on the number of tweets published by the user, the observed portion of the timeline is user-based. As a result, if we measure the tweeting volume (total amount of daily tweets over time) in the datasets, we observe a striking growth in the most recent weeks (Figure 3a, gray curve). This is a spurious effect, due to the fact that the observed timelines of the partially observed users fall in the latter period of the dataset timeframe. This is evident when plotting the number of active users who produced the corresponding tweeting volume (blue curve in Figure 3a). As we can see from the figure, the plots start from the early days of Twitter\(^4\) (since we have fully observed users whose observed timelines go back to the registration date). As we follow the timelines throughout the later dates, we notice that, while the number of active users increases (approximately linearly in most of the datasets), the total tweet volume per day increases exponentially. While the behaviour observed in Figure 3a well represents the general trend across most of the countries, Northern European countries (Denmark, Finland, Norway, and Sweden) are an exception. As shown in Figure 3b, very prolific journalists do not seem to be part of the dataset for these four countries.

\(^4\)Twitter has been established in 2006.
3.2. Keeping only real journalists

The Twitter lists that we used to identify our reference pool of journalists are created by other Twitter users. The general aim of these lists is to provide a curated collection of news/information sources generally related to a specific country or region. Therefore, these lists may include not only journalists but also news companies, anchormen, radio hosts, and so on. Even though these accounts are sources of information, they are not journalists. Including these accounts may lead us to a characterization of ego networks that do not represent journalists accurately. Thus, we need to apply a classifier to decrease the noise in the datasets. In order to correctly identify “true” journalists, we have tested three different approaches (and their combinations). The first one is based on journalism-related (such as columnist, correspondent) keyword-matching in the user Twitter bio. The second one entails searching for the person’s job in the Google Knowledge Graph (GKG). The third one filters out users that are marked as bots by the Botometer bot detector. For more details, please refer to Appendix A.2. We have evaluated the performance of these three classifiers and of their AND/OR combinations on a subset of journalists and non-journalists (Italian and English ones) that we have manually labelled. Since the journalist/non-journalist classes are quite imbalanced in our labelled dataset, we used the MCC (Matthews’ Correlation Coefficient [26]) for assessing which combination performs best. The results are shown in Figure 4. The classifier jointly considering the information in the GKG together with the keywords in the bio provides the best trade-off (high MCC, fewer computational steps) on the labelled datasets. Thus, we used it for filtering out non-journalists from the remaining datasets in our study. In the end, we removed XXX non-journalists from our datasets.

3.3. Identifying the regular and active users

Of all the journalist users remaining after the filtering in Section 3.2, we want to keep in our study only those that are regular and active Twitter users. The inactive life of a user is defined as the time interval between the last tweet of the user and the download time of the dataset. Long inactive periods are a sign of decreased engagement with the platform. In [15], a user is accepted as inactive if there is no activity for the last 6 months (mimicking Twitter 6-months inactivity criterion). In this paper, we are using the intertweet time, which we have previously introduced in [27]. This method labels a user as abandoned the platform if the inactive period is longer than six months plus the longest period between two consecutive tweets of the user. The intuition
behind the formulation is, instead of fixing an a-priori period of absence, to embed the user’s personal regularity pattern by adding their longest break in their tweeting activity. For the detailed formulation of *intertweet time*, the reader may refer to [27]. Using this filter, 1169 users are marked as having abandoned Twitter at the time we downloaded the dataset.

We now focus on how regular a user activity is. We assume that a Twitter user is regular, as in [19], if the user posts at least one tweet every 3 days as an average for at least half of the observed timeline. If this is not the case, the user is accepted as sporadic. In Figure 5, we show the classification of users both based on the abandonment classifier (previous paragraph) and the regularity classifier. As it may be expected, partially observed users are much more active than fully observed one. For our analysis, we focus on regular and active users, since they engaged with the platform regularly and, thus, spend more cognitive resources than others.

At the opposite end of the spectrum of inactive users are users that tweet a lot. These users could have a significant impact on the ego network statistics, due to their intense Twitter activity, and should be studied separately. In order to detect these outliers, we have run DBSCAN on the journalists tweeting frequencies. DBSCAN [28] is a standard clustering technique that automatically identifies outliers. We show the outcome of the outlier identification process for American journalists in Figure 6 (the same findings hold also for the remaining countries, as it can be observed in Appendix A.4).
Specifically, in the figure we plot the tweeting frequency against the length of the observed timeline. All the outliers feature an observed timeline shorter than one year. This is due to the fact that, as these users produce numerous tweets per day, they saturate the Twitter API limitation fast, hence their observed timelines are too short. Given that their observed timelines are shorter than one year, their ego networks are discarded, as discussed in Section 2.1.

The last check we perform is about the stationarity of the Twitter activity of the users in our dataset. As previously shown in the literature ([23], [18], [29]), newly registered social network users add relationships into their network and interact with others at a higher rate compared to long-term
users. After a while, their activity stabilizes. Newly registered users are thus outliers with respect to the general population of users, and they should be discarded from the analysis. Again, we use the approach described in [27]. We do not observe a drastic change in tweeting activity for none of the datasets. Therefore, we do not discard any part of the timeline as a transient period. For more details on this analysis, please refer to Appendix A.6.

3.4. Dataset overview

The number of journalists in our dataset at the end of the preprocessing phase is reported, per country, in Table 2. In Table 3 we provide some summary statistics regarding the length of the their active life on Twitter, their average daily tweets, and their predominant type of tweet. With an average Twitter life between 6 and 9 years, the journalists in our datasets tend to be long-time Twitter users. These journalists tweet, on average, twice or thrice per day. 65% of the tweets we observe are social tweets (reply, retweet\textsuperscript{5} and mention), i.e., entail a direct communication between two Twitter users.

When considering the split between the different flavours of social tweets and indirect tweets, some trends emerge. Specifically, we can identify (Figure 7) three distinct groups of countries. In the cluster denoted in blue in Figure 7 (comprising Japan, Greece, Turkey, Brazil), the dominant tweet type are indirect tweets. These journalists, thus, tend to post their tweets without interacting with others, not even mentioning their news outlet. At the opposite end of the spectrum, journalists in the red cluster predominantly engage on Twitter by means of replies. Interestingly, they all belong to the same geographical area (Northern Europe). Note that replies are reactive, to what other people/news outlet have posted online. The third cluster (members colored in pink) comprises all the English-speaking countries in our dataset plus the Central and Southern European states. For these countries the prevalent mode of interaction is through retweets and mentions. This is perhaps the most expected pattern of engagement for journalists on Twitter: sharing content from other users and tagging users (among which, their news outlets) in conversations.

\textsuperscript{5}Quote tweets are treated as retweets.
Figure 7: Tweet types across countries. On the left, the average number of tweets in the different categories, computed by country. On the right, the corresponding PCA plot. The colors denote the 3 groups detected by the k-means clustering algorithm on the PCA components. The optimal number of groups has been obtained using the Silhouette method.

4. Ego network analysis

In this section, we analyze the ego network structure of the journalists that have been filtered as described in Section 3, and show the differences and invariants between different countries and regions. The frequency of direct tweets (mentions, reply, or retweets) is used to represent strength of the ties (a.k.a. intimacy or emotional closeness) between journalists (egos) and their alters as discussed in Section 2.1. Recall also that, similarly to the related literature, a relationship is considered active if the ego and the alter have at least one contact (direct tweet) per year.

4.1. Static properties of ego networks

The static view of an ego network is a single, aggregate snapshot of the ego and its ties. It is a representation of the ego network that includes the whole observed timeline of the journalist. This analysis allows us to understand the general distribution of cognitive resources through all the alters and to compare results with other findings from the related literature.
Table 3: Summary statistics

| Dataset | Statistic Type | Active Life [years] | Total Tweets | Observed Tweets | Tweet/Day | % Social | % Replies | % Retweets | % Mention |
|---------|----------------|---------------------|--------------|-----------------|-----------|----------|----------|-----------|-----------|
| US      | mean           | 9.08                | 13199.12     | 3123.47         | 3.24      | 65.41    | 25.27    | 43.30     | 31.42     |
|         | sd             | 0.97                | 13467.95     | 318.19          | 2.02      | 16.89    | 19.56    | 22.28     | 18.60     |
| Canada  | mean           | 8.04                | 10544.96     | 3008.35         | 3.30      | 61.35    | 23.45    | 49.47     | 27.08     |
|         | sd             | 1.43                | 13027.79     | 511.59          | 2.19      | 19.47    | 19.26    | 22.49     | 18.92     |
| Brazil  | mean           | 8.58                | 14874.91     | 3079.88         | 3.14      | 56.09    | 33.96    | 37.64     | 28.39     |
|         | sd             | 1.59                | 16840.28     | 394.39          | 2.05      | 18.97    | 24.43    | 23.69     | 21.42     |
| Japan   | mean           | 6.42                | 7416.97      | 2717.06         | 2.81      | 45.37    | 18.16    | 67.73     | 14.11     |
|         | sd             | 2.33                | 7156.52      | 817.50          | 1.99      | 19.14    | 20.08    | 25.74     | 17.41     |
| Turkey  | mean           | 7.49                | 11167.86     | 3014.68         | 2.94      | 58.67    | 24.57    | 56.35     | 19.17     |
|         | sd             | 1.32                | 10190.01     | 465.07          | 1.80      | 19.73    | 20.45    | 24.22     | 16.550    |
| UK      | mean           | 8.45                | 11369.24     | 3052.51         | 3.18      | 67.39    | 33.92    | 43.89     | 22.28     |
|         | sd             | 1.07                | 10756.58     | 460.84          | 2.12      | 15.96    | 19.81    | 20.70     | 14.62     |
| Denmark | mean           | 7.14                | 5895.30      | 2547.93         | 2.01      | 67.63    | 47.02    | 36.04     | 16.93     |
|         | sd             | 1.83                | 5795.11      | 822.68          | 1.59      | 14.80    | 42.10    | 22.11     | 13.57     |
| Finland | mean           | 6.36                | 5141.04      | 2569.54         | 2.28      | 62.18    | 39.51    | 44.65     | 15.84     |
|         | sd             | 1.78                | 4791.04      | 340.97          | 1.77      | 16.93    | 22.59    | 22.34     | 13.93     |
| Norway  | mean           | 8.59                | 6993.00      | 2606.72         | 1.78      | 64.29    | 43.98    | 36.78     | 19.24     |
|         | sd             | 1.26                | 7383.59      | 586.30          | 1.21      | 17.66    | 22.74    | 22.44     | 15.63     |
| Sweden  | mean           | 8.57                | 12847.46     | 3088.71         | 2.50      | 66.29    | 44.65    | 36.81     | 18.54     |
|         | sd             | 1.08                | 12224.01     | 424.71          | 1.70      | 16.22    | 22.09    | 21.03     | 14.66     |
| Greece  | mean           | 7.33                | 10426.97     | 2937.37         | 2.45      | 53.10    | 27.41    | 51.67     | 20.92     |
|         | sd             | 1.29                | 14919.70     | 576.30          | 1.74      | 20.90    | 21.56    | 26.07     | 20.78     |
| Italy   | mean           | 7.61                | 10536.59     | 3064.37         | 2.89      | 66.34    | 21.38    | 52.52     | 26.10     |
|         | sd             | 1.30                | 8622.80      | 415.26          | 1.84      | 19.34    | 17.64    | 20.02     | 19.14     |
| Spain   | mean           | 8.43                | 17651.02     | 3134.53         | 3.45      | 65.80    | 28.49    | 41.41     | 30.10     |
|         | sd             | 1.07                | 14219.35     | 274.32          | 2.06      | 16.03    | 19.88    | 20.36     | 18.51     |
| France  | mean           | 8.03                | 10269.97     | 3021.49         | 3.06      | 71.92    | 21.34    | 52.53     | 26.44     |
|         | sd             | 1.30                | 9634.66      | 445.22          | 2.16      | 14.49    | 18.84    | 21.22     | 17.29     |
| Germany | mean           | 8.21                | 7810.44      | 2846.30         | 2.17      | 67.05    | 32.13    | 43.06     | 24.81     |
|         | sd             | 1.32                | 7969.02      | 667.03          | 1.62      | 14.80    | 18.53    | 20.42     | 15.05     |
| Netherlands | mean  | 8.07          | 9889.23      | 2902.89         | 2.34      | 63.35    | 41.54    | 35.87     | 22.49     |
|         | sd             | 1.52                | 12482.73     | 617.79          | 1.72      | 17.19    | 22.22    | 21.84     | 15.25     |
| Australia | mean         | 7.85          | 7146.85      | 2882.53         | 2.62      | 72.06    | 23.82    | 46.35     | 29.83     |
|         | sd             | 1.43                | 5984.29      | 612.12          | 1.80      | 15.36    | 16.13    | 19.44     | 18.09     |
| All Journalists | mean | 9.08    | 13199.12     | 3123.47         | 3.24      | 65.41    | 25.27    | 43.30     | 31.42     |
|         | sd             | 0.97                | 13396.81     | 317.94          | 2.02      | 16.88    | 19.54    | 22.26     | 19.59     |

in a quantitative manner (since the analysis of static ego networks is the starting point in all related works).

First, we show the distribution of the average number of alters (with confidence intervals) in Figure 8. This number captures how many peers the ego user interacts with during the last 3200 observed tweets, and it corresponds to the size of the ego network when both active and inactive relationships are considered (Section 2). As expected, the countries whose journalists engage on Twitter predominantly via indirect tweets (Figure 7) tend to feature fewer alters than their more social counterparts. Figure 8 also shows that journalists that mostly rely on replies (such as those from Finland, the Netherlands, and Denmark) have a below-average number of alters, thus suggesting a ten-
dency to get involved with the same group of people. This might hint at the fact that replies are a more personal/intimate communication with respect to mention and retweets, hence they consume more cognitive resources on the ego side, which, in turn, is able to interact with fewer people. Vice versa, the journalists in countries from the pink cluster in Figure 7 tend to interact with above-average distinct peers, hinting at the opposite effect. The distribution of the number of alters per ego in each dataset, instead of dataset averages, can be seen in Appendix B. Even though there are differences for journalists from different regions and countries, still the average number of alters of journalists are much bigger than those of the generic Twitter users analyzed in [1] where more than 90% of users have less than 100 relationships. This supports the claim that journalists establish more relationships and they are prominent users on Twitter.

The number of alters showed in Figure 8 include relationships that egos don’t allocate their cognitive resources regularly. We now show the distributions of average active relationship numbers in our datasets (Figure 9). When we consider only active relationships instead of all relationships, the average number of alters decreases due to the cognitive limitations to maintain relationships actively. The difference among the average active network sizes decrease a lot with respect to Figure 8, with sizes varying between 75 alters (Japanese journalists) and 147 (Spanish journalists). Turkish, Greek and Japanese journalists still have the smallest network sizes. The difference between journalists and generic Twitter users still exists. The active network size of generic Twitter users was 88 alters in [1], while the average for journalists is 119. Hence, we can infer that the ego networks sizes of the journalists mimic offline human ego network sizes predicted by Dunbar’s number, and it is closer to Dunbar’s number more than the generic Twitter users.  

Active relationships of the egos can be grouped as layers (also known as Dunbar’s circles [30]) based on their emotional closeness with the ego. It has been shown that there are four layers in offline human ego networks [13]. This number is extended to five layers for OSNs [1]. The optimal number of

---

While the initial filtering used in [1] is different from the one discussed in this paper (Section 3), the comparison between the two sets of results is still meaningful. In fact, the filtering in [1] (whereby only users with an average of more than 10 interactions per month are kept) tends to retain only very active users, thus it might overestimate the social interactions of generic users. However, our results indicate that journalists tend to be even more engaged than this active subset of generic users.
circles can be determined by non-parametric clustering algorithms. To this aim, we used Mean Shift algorithm [25]. Figure 10 shows the distribution of circle number per ego for all journalists (the distribution of circle numbers per dataset can be seen in Appendix D). The mode value of optimal circle numbers is 5, mean value is 5.6, and median value is 6. All of the datasets have a mean and median value of circle numbers around five (Appendix D) except Japanese journalists (which is not surprising as it is mentioned before that active ego network sizes of Japanese journalists are smaller than the other ones). On the other hand, even though the mean and median values of circle numbers are around five, there are a few countries where the mode values of circle size is not five, such as French journalists where the mode value is 6. However, since almost all of the datasets have a mean, median, and mode value of 5 and to be compatible with the related literature, we consider egos with five circles for the rest of the paper and discard the ones who don’t have five circles until Section 4.4.

Table 4 shows the circle sizes (alter number per inclusive Dunbar’s circles). As we can see from the last row of the table, the average of all journalists, the circle sizes of journalists are very close to Dunbar’s circle sizes (1.5, 5, 15, 50, and 150) shown for OSN [1] with slight variations. Although the circle sizes shown for OSN users tend to be smaller [1], journalists have more alters than offline human ego network circles except for the most outer circle (5th circle). Still, the composition of the circles of journalists is quite similar to offline
human ego network circles and larger than the other OSN users analyzed: generic Twitter users [1] and politicians [19]. The active network size (5th circle), on average $\sim 119$ alters, is quite close to Dunbar’s number (150). In Table 4, we also show the scaling ratio between consecutive circles which has been shown to be around 3 in the literature [13]. Again, the scaling ratio of the journalists is quite consistent with the finding of the literature.

In this section, we have shown that journalists tend to have more active relationships than generic Twitter users and politicians, therefore their ego network sizes are closer to Dunbar’s number. Turkish, Greek and Japanese journalists have smaller ego networks compared the other countries. British, French, and American journalists establish more relationships on the platform while Spanish journalists maintain more active relationships. Journalists’ cognitive resource allocation through the alters they have mimic the offline human ego networks. The circles sizes and the scaling ratio between the circle sizes are consistent with Dunbar’s model and number of alters per circle is higher than generic Twitter users and politicians.

4.2. Dynamic properties of ego networks

In this section, we analyze the evolution of ego networks through time and changes in the relationships. If we divide the observed timeline of a user into equal time intervals, the intimacy of the relationships doesn’t have to stay at the same level for all time intervals. An ego may allocate more cognitive resources for an alter in the early time intervals while may decrease
this intimacy in the latter ones, or vice versa. An ego also may prefer to keep the same alters at the same intimacy level through all time intervals. To understand how the configuration of the circles changes, we consider one-year time intervals with 1-month steps sizes (11 months overlap between two consecutive time windows) as in [23]. Each time window is a snapshot of the ego networks where we focus on rings (the portion of circles excluding the inner circles) and how they change over time.

As discussed in Section 2.1, we study the dynamic changes in ego networks by means of the Jaccard similarity and the Jump index. The Jaccard index is obtained dividing the cardinality of the intersection set by the cardinality of the union set. The closer this index to 1, the better the overlapping. The amount of movements between rings is measured through the Jump index, which simply counts (and then averages across all alters in the ring) the number of jumps between rings. We consider two time frames for our analysis, capturing respectively short-term and long-term changes. Let us start with the short-term perspective. Tables 5 and 6 show the values of Jaccard and jump indices with 1-year time interval and 1-month step size for the five rings. With these approach, we compare how ego networks change on a month-by-month basis. The same analysis was carried out in [19] for politicians. We can see (Table 5) that the Jaccard indices are higher at the innermost (R1) and the outermost (R5) rings, while middle rings have lower index values (U-shaped pattern). This shows that the most inner and outer rings
in the inner layers are more likely to jump less rings (intimacy level) than signalling that movements mostly occur between adjacent rings. The alters the Jump index (Table 6). The Jump indices increase through the outer are extremely consistent across countries. We observe a different trend for that will lead to change the alters more in the middle rings. And these results establish new relationships, terminate the old ones or change the intimacy start new relationships a lot over time. However, they are more likely to relationships and least intimate relationships steady and don’t terminate or don’t change much over time. Journalists are keeping their most intimate

| dataset | C1 ratio | C2 ratio | C3 ratio | C4 ratio | C5 ratio |
|---------|----------|----------|----------|----------|----------|
| US | 3.24 ± 0.21 | 2.96 ± 0.21 | 2.58 ± 0.11 | 2.56 ± 0.08 | 48.42 ± 2.52 | 124.56 ± 5.78 |
| Canada | 3.43 ± 0.27 | 2.63 ± 0.27 | 2.54 ± 0.12 | 2.46 ± 0.12 | 114.26 ± 7.53 |
| Brazil | 3.14 ± 0.31 | 2.98 ± 0.33 | 2.45 ± 0.12 | 2.32 ± 0.1 | 105.3 ± 9.43 |
| Japan | 2.98 ± 0.31 | 2.72 ± 0.3 | 2.35 ± 0.13 | 2.31 ± 0.1 | 96.2 ± 9.93 |
| Turkey | 3.1 ± 0.16 | 2.64 ± 0.15 | 2.66 ± 0.09 | 2.51 ± 0.07 | 2.79 ± 0.09 |
| UK | 3.5 ± 0.39 | 2.74 ± 0.32 | 2.72 ± 0.21 | 2.72 ± 0.32 | 52.65 ± 4.71 | 134.07 ± 11.56 |
| Denmark | 3.63 ± 0.22 | 2.92 ± 0.17 | 2.5 ± 0.09 | 2.33 ± 0.06 | 2.41 ± 0.08 |
| Finland | 3.74 ± 0.37 | 2.74 ± 0.25 | 2.61 ± 0.17 | 2.49 ± 0.14 | 53.57 ± 4.5 |
| Norway | 3.55 ± 0.42 | 2.91 ± 0.19 | 2.36 ± 0.16 | 2.45 ± 0.13 | 115.56 ± 11.14 |
| Sweden | 3.38 ± 0.35 | 2.93 ± 0.27 | 2.61 ± 0.17 | 2.49 ± 0.13 | 49.1 ± 9.7 | 122.77 ± 9.94 |
| Greece | 2.92 ± 0.28 | 2.49 ± 0.23 | 2.34 ± 0.16 | 2.44 ± 0.14 | 34.5 ± 3.36 | 91.76 ± 7.68 |
| Italy | 2.81 ± 0.29 | 2.97 ± 0.37 | 2.73 ± 0.19 | 2.65 ± 0.18 | 2.94 ± 0.23 |
| Spain | 3.2 ± 0.3 | 2.74 ± 0.27 | 2.42 ± 0.16 | 2.59 ± 0.16 | 46.97 ± 5.02 | 124.64 ± 10.65 |
| France | 3.2 ± 0.32 | 2.81 ± 0.16 | 2.74 ± 0.21 | 2.7 ± 0.13 | 2.78 ± 0.17 |
| Germany | 3.39 ± 0.54 | 3.1 ± 0.41 | 8.99 ± 1.35 | 22.8 ± 2.97 | 52.99 ± 6.21 | 127.25 ± 13.74 |
| Netherlands | 3.28 ± 0.12 | 2.88 ± 0.11 | 2.49 ± 0.06 | 2.45 ± 0.05 | 2.55 ± 0.05 |
| Australia | 3.19 ± 0.29 | 2.82 ± 0.23 | 2.63 ± 0.15 | 2.63 ± 0.15 | 49.17 ± 3.83 | 128.7 ± 8.67 |
| All journalists | 3.24 ± 0.05 | 2.96 ± 0.05 | 2.58 ± 0.03 | 2.56 ± 0.02 | 2.72 ± 0.02 | 124.56 ± 1.35 |

Let us now study the long-term dynamics of ego networks. To this aim, we consider 1-year time intervals for observing ego networks, and 1-year
step sizes (there is no intersection between two consecutive observation time intervals). Results are shown in Tables 7 and 8. The U-shape pattern across rings for the Jaccard coefficient is still there, but the similarity values are much lower. This means that, on a yearly basis, the rings of ego networks tend to change significantly. In the central rings, especially, the turnover in the rings is total, as indicated by the Jaccard index close to zero. Surprisingly, the most stable ring is R5, which contains the weakest active social links. We observe slightly more variability across countries with respect to the short-term results in Table 5. This suggests that, as expected, long-term dynamics tend to be more interesting than short-term ones. The same comment holds for the Jump index in Table 8. In this case, the pattern observed for the Jump index is completely different from the previous case (Table ??): the index decreases as we move from R1 to R5, while previously it was increasing. We see that alters come to the first ring after moving from non-adjacent circles, but this effect wanes as we move towards the outermost circles.

In this section, we have shown that journalists have a stable short-term relationships that don’t change a lot over time. They tend to keep their most and least intimate relationships as they are through the time while they change the average intimate relationships much more. On the longer term, though, ego networks can be pretty dynamic, especially in the innermost circles. This is in contrast with the findings about generic Twitter users [15] and suggests that the ego networks of journalists, while structurally similar to that of generic users, may be affected by the information-driven nature of journalist engagement on the platform, thus yielding to much more variability in the composition of rings.

4.3. Social tweets and hashtags

Considering that journalists are promoting their work, establishing their personal brands, and trying to satisfy the user demands for news, we may expect them to be more topic-driven than generic Twitter users. The typical way of searching for and publishing tweets about a specific topic is to use hashtags. Thus, we expect journalists to be using hashtags at a higher rate than regular users. To look into this aspect, we provide hashtag-related statistics about the journalist datasets. In Figure 11, the percentage of hashtag activated relationships are shown. A relationship is labelled hashtag activated if the first contact (direct tweet) of the relationship includes a hashtag. For all journalists, the average hashtag activated relationship percentage is 23% while it is 6% for generic Twitter users [1] and approximately 15% for
Table 5: Jaccard coefficients of egos with optimal circle number 5 with 1-month step size for 1-year intervals.

| dataset    | jaccard coefficient |
|------------|---------------------|
|            | R1     | R2     | R3     | R4     | R5     |
| US         | 0.83 ± 0.01 | 0.61 ± 0.01 | 0.59 ± 0.01 | 0.72 ± 0.02 | 0.86 ± 0.02 |
| Canada     | 0.81 ± 0.02 | 0.61 ± 0.02 | 0.59 ± 0.02 | 0.71 ± 0.02 | 0.85 ± 0.03 |
| Brazil     | 0.8 ± 0.02  | 0.61 ± 0.03  | 0.58 ± 0.02  | 0.7 ± 0.03  | 0.85 ± 0.03 |
| Japan      | 0.8 ± 0.02  | 0.6 ± 0.02  | 0.58 ± 0.02  | 0.7 ± 0.02  | 0.85 ± 0.03 |
| Turkey     | 0.82 ± 0.01 | 0.61 ± 0.01 | 0.58 ± 0.01 | 0.7 ± 0.01  | 0.84 ± 0.01 |
| UK         | 0.79 ± 0.04 | 0.59 ± 0.03 | 0.58 ± 0.03 | 0.72 ± 0.04 | 0.86 ± 0.05 |
| Denmark    | 0.79 ± 0.01 | 0.59 ± 0.01 | 0.59 ± 0.01 | 0.74 ± 0.01 | 0.88 ± 0.02 |
| Finland    | 0.81 ± 0.01 | 0.61 ± 0.01 | 0.6 ± 0.01  | 0.76 ± 0.01 | 0.92 ± 0.02 |
| Norway     | 0.8 ± 0.02  | 0.6 ± 0.02  | 0.6 ± 0.02  | 0.77 ± 0.02 | 0.92 ± 0.02 |
| Sweden     | 0.82 ± 0.02 | 0.62 ± 0.02 | 0.6 ± 0.02  | 0.74 ± 0.02 | 0.88 ± 0.02 |
| Greece     | 0.82 ± 0.03 | 0.62 ± 0.02 | 0.59 ± 0.02 | 0.7 ± 0.03  | 0.85 ± 0.03 |
| Italy      | 0.84 ± 0.02 | 0.64 ± 0.02 | 0.61 ± 0.02 | 0.72 ± 0.02 | 0.85 ± 0.03 |
| Spain      | 0.82 ± 0.03 | 0.61 ± 0.03 | 0.58 ± 0.02 | 0.7 ± 0.03  | 0.83 ± 0.03 |
| France     | 0.83 ± 0.02 | 0.62 ± 0.02 | 0.6 ± 0.02  | 0.73 ± 0.02 | 0.86 ± 0.03 |
| Germany    | 0.82 ± 0.02 | 0.61 ± 0.02 | 0.6 ± 0.02  | 0.76 ± 0.02 | 0.89 ± 0.03 |
| Netherlands| 0.81 ± 0.01 | 0.61 ± 0.01 | 0.59 ± 0.01 | 0.72 ± 0.01 | 0.86 ± 0.01 |
| Australia  | 0.84 ± 0.01 | 0.64 ± 0.02 | 0.6 ± 0.01  | 0.72 ± 0.02 | 0.86 ± 0.02 |
| All journalists | 0.83 ± 0.12 | 0.61 ± 0.12 | 0.59 ± 0.12 | 0.72 ± 0.14 | 0.86 ± 0.17 |

This shows that journalists establish more topic-driven relationships. All of the journalist datasets have a higher hashtag-activated percentage than generic Twitter users, and most of them also higher than the politicians. However, when we compare the countries, there is a high variability, with values ranging from 12% in Japanese journalists to 36% in Italian journalists, in terms of hashtag-activated relationship percentages.

Only showing the percentage of relationship activated by hashtag does not give us information about the usage of hashtag based on rings, hence on the dependency on intimacy. In Figure 12, the percentage of hashtag-activated relationships is shown for selected countries, in order to highlight three different characteristics featured in the datasets. American, Finnish, French, and Netherlander journalists use the same amount of hashtags across the rings. Hence, the percentage of their relationships activated by hashtags is independent of the intimacy level of relationships. German, Greek, and Norwegian journalists activate their less intimate relationships (outer rings)
Table 6: Jump indices of egos with optimal circle number 5 with 1-month step size for 1-year intervals.

| dataset | jump index  |
|---------|-------------|
|         | R1          | R2          | R3          | R4          | R5          |
| US      | 0.49 ± 0.03 | 0.89 ± 0.03 | 1.05 ± 0.02 | 1.46 ± 0.04 | 0.36 ± 0.04 |
| Canada  | 0.51 ± 0.04 | 0.87 ± 0.05 | 1.04 ± 0.03 | 1.46 ± 0.06 | 0.36 ± 0.06 |
| Brazil  | 0.56 ± 0.06 | 0.92 ± 0.05 | 1.09 ± 0.04 | 1.37 ± 0.07 | 0.43 ± 0.07 |
| Japan   | 0.58 ± 0.06 | 0.89 ± 0.05 | 1.07 ± 0.03 | 1.44 ± 0.06 | 0.43 ± 0.08 |
| Turkey  | 0.49 ± 0.03 | 0.86 ± 0.03 | 1.03 ± 0.02 | 1.41 ± 0.03 | 0.44 ± 0.04 |
| UK      | 0.51 ± 0.05 | 0.91 ± 0.06 | 1.07 ± 0.05 | 1.51 ± 0.09 | 0.31 ± 0.08 |
| Denmark | 0.59 ± 0.03 | 0.96 ± 0.02 | 1.1 ± 0.02  | 1.52 ± 0.03 | 0.28 ± 0.04 |
| Finland | 0.6 ± 0.05  | 0.99 ± 0.03 | 1.14 ± 0.03 | 1.58 ± 0.05 | 0.22 ± 0.05 |
| Norway  | 0.66 ± 0.06 | 1.01 ± 0.04 | 1.16 ± 0.04 | 1.53 ± 0.06 | 0.22 ± 0.06 |
| Sweden  | 0.53 ± 0.04 | 0.96 ± 0.04 | 1.12 ± 0.03 | 1.47 ± 0.06 | 0.32 ± 0.06 |
| Greece  | 0.48 ± 0.04 | 0.83 ± 0.05 | 1.02 ± 0.04 | 1.23 ± 0.07 | 0.42 ± 0.06 |
| Italy   | 0.47 ± 0.05 | 0.85 ± 0.06 | 1.08 ± 0.03 | 1.48 ± 0.06 | 0.42 ± 0.08 |
| Spain   | 0.51 ± 0.06 | 0.85 ± 0.05 | 1.05 ± 0.04 | 1.38 ± 0.07 | 0.48 ± 0.08 |
| France  | 0.49 ± 0.05 | 0.86 ± 0.04 | 1.06 ± 0.03 | 1.47 ± 0.06 | 0.39 ± 0.07 |
| Germany | 0.54 ± 0.06 | 0.92 ± 0.05 | 1.1 ± 0.04  | 1.58 ± 0.06 | 0.22 ± 0.06 |
| Netherlands | 0.51 ± 0.02 | 0.9 ± 0.02 | 1.08 ± 0.01 | 1.44 ± 0.02 | 0.39 ± 0.02 |
| Australia | 0.49 ± 0.05 | 0.84 ± 0.04 | 1.05 ± 0.02 | 1.46 ± 0.05 | 0.43 ± 0.06 |
| All journalists | 0.49 ± 0.27 | 0.89 ± 0.29 | 1.05 ± 0.21 | 1.46 ± 0.36 | 0.36 ± 0.37 |

by hashtags more than more intimate layers (inner rings). The rest of the datasets have higher hastag-activated relationship percentage in the inner rings. All of the datasets, which can be seen in Appendix E, have these three characteristics.

The number of hashtags used per relationship is much higher for the relationships that are activated by hashtag as can be seen in Figure 13 for a selected dataset (since all the datasets have the same characteristics, we have shown only the results for one country; the others are shown in Appendix F). On the other hand, cognitive resources allocated (measured in terms of contact frequency) per ego aren’t affected if the relationship has been activated by a hashtag or not, as shown in Figure 14 (results for the remaining countries are reported in Appendix G).
Table 7: Jaccard coefficients of egos with optimal cycle number 5 with 1-year step size for 1-year intervals.

| dataset  | jaccard coefficient |  |  |  |  |
|----------|---------------------|---|---|---|---|
|          | R1                  | R2 | R3 | R4 | R5 |
| US       | 0.25 ± 0.03         | 0.07 ± 0.01 | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.49 ± 0.05 |
| Canada   | 0.24 ± 0.04         | 0.07 ± 0.01 | 0.05 ± 0.01 | 0.07 ± 0.01 | 0.51 ± 0.07 |
| Brazil   | 0.17 ± 0.04         | 0.04 ± 0.01 | 0.04 ± 0.01 | 0.06 ± 0.02 | 0.4 ± 0.08  |
| Japan    | 0.18 ± 0.04         | 0.07 ± 0.02 | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.41 ± 0.09 |
| Turkey   | 0.21 ± 0.02         | 0.06 ± 0.01 | 0.05 ± 0.01 | 0.07 ± 0.01 | 0.44 ± 0.04 |
| UK       | 0.19 ± 0.04         | 0.06 ± 0.02 | 0.04 ± 0.01 | 0.06 ± 0.01 | 0.54 ± 0.09 |
| Denmark  | 0.22 ± 0.02         | 0.07 ± 0.01 | 0.06 ± 0.01 | 0.08 ± 0.01 | 0.63 ± 0.04 |
| Finland  | 0.23 ± 0.03         | 0.06 ± 0.01 | 0.05 ± 0.01 | 0.08 ± 0.01 | 0.72 ± 0.07 |
| Norway   | 0.2 ± 0.04          | 0.06 ± 0.01 | 0.05 ± 0.01 | 0.09 ± 0.02 | 0.67 ± 0.08 |
| Sweden   | 0.27 ± 0.04         | 0.08 ± 0.02 | 0.06 ± 0.01 | 0.09 ± 0.02 | 0.56 ± 0.07 |
| Greece   | 0.23 ± 0.04         | 0.07 ± 0.02 | 0.05 ± 0.02 | 0.08 ± 0.02 | 0.45 ± 0.07 |
| Italy    | 0.26 ± 0.05         | 0.06 ± 0.01 | 0.06 ± 0.01 | 0.06 ± 0.01 | 0.51 ± 0.08 |
| Spain    | 0.24 ± 0.05         | 0.06 ± 0.01 | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.41 ± 0.08 |
| France   | 0.28 ± 0.05         | 0.07 ± 0.02 | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.48 ± 0.08 |
| Germany  | 0.3 ± 0.05          | 0.09 ± 0.02 | 0.08 ± 0.02 | 0.1 ± 0.02  | 0.68 ± 0.09 |
| Netherlands | 0.25 ± 0.01        | 0.08 ± 0.01 | 0.07 ± 0.0 | 0.08 ± 0   | 0.51 ± 0.03 |
| Australia| 0.27 ± 0.04         | 0.08 ± 0.02 | 0.06 ± 0.01 | 0.07 ± 0.01 | 0.5 ± 0.07  |
| All journalists | 0.25 ± 0.23       | 0.07 ± 0.08 | 0.06 ± 0.06 | 0.07 ± 0.08 | 0.49 ± 0.42 |

4.4. Popularity of the alters and intimacy level of the relationships

Different motivations of journalists for using Twitter have been highlighted in the literature. Journalists use Twitter as a tool to establish and promote their personal brands [2] [31] and to promote content from their news website [2]. According to the study in [31], journalists use each other as sources by retweeting each other’s tweets. In this section, considering all the behavioral characteristics that are underlined in the literature about Twitter usage of journalists, we investigate the relationship between the popularity of a journalist and the popularity of the journalists they engage with on Twitter. Our goal is to show if popular journalists interact with similarly popular journalists (a sort of assortativity in popularity) and if there is a correlation between the popularity of alters and the intimacy level of the relationships (i.e., if popular alters tend to be in the innermost vs outermost circles)? In the following, we measure popularity in terms of the number of followers for the account, similarly to XXX.
Table 8: Jump indices of egos with optimal cycle number 5 with 1-year step size for 1-year intervals.

| dataset   | jump index |        |        |        |        |
|-----------|------------|--------|--------|--------|--------|
| US        | 1.84 ± 0.15| 1.78 ± 0.12| 1.57 ± 0.11| 1.4 ± 0.09| 0.28 ± 0.04 |
| Canada    | 1.77 ± 0.19| 1.8 ± 0.17| 1.62 ± 0.14| 1.42 ± 0.12| 0.28 ± 0.06 |
| Brazil    | 1.69 ± 0.26| 1.67 ± 0.24| 1.43 ± 0.19| 1.17 ± 0.16| 0.26 ± 0.07 |
| Japan     | 2.07 ± 0.28| 1.86 ± 0.25| 1.62 ± 0.2 | 1.4 ± 0.17| 0.36 ± 0.09 |
| Turkey    | 1.97 ± 0.14| 1.83 ± 0.11| 1.61 ± 0.09| 1.39 ± 0.08| 0.34 ± 0.04 |
| UK        | 1.87 ± 0.3 | 1.78 ± 0.25| 1.56 ± 0.21| 1.36 ± 0.18| 0.19 ± 0.07 |
| Denmark   | 2.39 ± 0.12| 2.25 ± 0.09| 1.94 ± 0.07| 1.69 ± 0.06| 0.3 ± 0.04  |
| Finland   | 2.33 ± 0.2 | 2.19 ± 0.16| 1.88 ± 0.13| 1.65 ± 0.11| 0.17 ± 0.05 |
| Norway    | 2.46 ± 0.23| 2.4 ± 0.17 | 2.04 ± 0.14| 1.69 ± 0.11| 0.26 ± 0.07 |
| Sweden    | 2.24 ± 0.2 | 2.2 ± 0.16 | 1.93 ± 0.12| 1.65 ± 0.1 | 0.37 ± 0.07 |
| Greece    | 2.33 ± 0.25| 2.14 ± 0.2 | 1.86 ± 0.16| 1.51 ± 0.13| 0.41 ± 0.07 |
| Italy     | 2.15 ± 0.26| 2.02 ± 0.21| 1.75 ± 0.17| 1.55 ± 0.14| 0.34 ± 0.08 |
| Spain     | 2.01 ± 0.28| 1.87 ± 0.22| 1.63 ± 0.18| 1.44 ± 0.16| 0.41 ± 0.09 |
| France    | 1.76 ± 0.23| 1.84 ± 0.2 | 1.62 ± 0.17| 1.45 ± 0.14| 0.32 ± 0.07 |
| Germany   | 2.18 ± 0.22| 2.12 ± 0.17| 1.89 ± 0.14| 1.72 ± 0.12| 0.27 ± 0.08 |
| Netherlands| 2.04 ± 0.08| 1.99 ± 0.07| 1.79 ± 0.06| 1.57 ± 0.05| 0.37 ± 0.03 |
| Australia | 2.15 ± 0.21| 2.04 ± 0.16| 1.77 ± 0.13| 1.55 ± 0.11| 0.38 ± 0.07 |
| All journalists | 1.84 ± 1.28| 1.78 ± 1.09| 1.57 ± 0.92| 1.40 ± 0.80| 0.28 ± 0.38 |

To understand the effect of the popularity of the alters in different ego network layers, we have carried out a correlation analysis between the popularity of egos and that of their alters. To this aim, we first labeled the alters of the journalists by using the same method discussed in the Section ?? (which entails leveraging the Google Knowledge graph together with the keywords in the Twitter bio). Then, the correlation scores between the ego’s popularity (its follower number) and the average alter popularity (mean of follower number of alters) are calculated. The correlation measures vary between −1 and +1, where +1 represents the perfect positive association, −1 represents the perfect negative association and 0 represents the independence of the variables. Note that the methods used for correlation analysis are descriptive statistical measurement, not causal explanation [32]. In this section, we will consider pearman’s rank-order/rho correlation (Spearman’s correlation) and Kendall’s tau correlation (Kendall’s correlation). Spearman’s and Kendall’s correlations capture how well a monotonic function can describe
the relationship, not essentially linear relationship unlike Pearson’s correlation, between two variables [33]. Since Spearman’s and Kendall’s correlations are a rank-based method instead of interval-based method, they are less sensitive to outliers/noise [33]. Kendall’s correlation is also less sensitive to outliers than Spearman’s correlation, and Spearman’s correlation coefficient is usually closer to Pearson’s correlation coefficient and generally has higher correlation values than Kendall’s [32].

In Tables 9 and 10, we provide the Spearman’s and Kendall’s correlations, respectively, between egos and their alters, which can be other journalists or other, unlabelled categories. Red rows represent the correlation between the popularity of the egos and all their alters independently of their profession. Blue and green rows represent the correlation between the egos and their journalist alters and non-journalists alter correspondingly. Empty cells corresponds to cases where the p-value is not low enough to draw reliable conclusions. For the correlation analysis, we align the innermost circle (1st circle) of all egos to calculate the correlation of the datasets. Therefore, there are different numbers of egos per ring since the circle number of egos varies as shown in Figure 10. We visualize the first 7 rings since the number of egos with more than 7 circles decreases drastically, and we don’t have enough samples to measure correlation. The last column of the table active shows the active ego network of egos where we include all circles of the egos.

Table 9 shows Spearman’s correlations between journalist egos and their
First of all, let’s focus on the correlation in the active network. In each dataset, we see that correlation of journalist alters is higher than non-journalist alters. Swedish journalists have much less difference than the other countries. We observe that all the correlation values for journalist alters are positive, although some of them are quite close to 0. For example, Turkish journalists have 0.09 for journalists alters while having −0.14 for non-journalists alters. The average correlation of all journalist egos in the dataset (last row of the table) is 0.453 for journalist alters and 0.116 for non-journalist alters. This shows that highly popular journalists tend to engage with other journalists of similar popularity, and vice versa. Instead, when interacting with generic users, their popularity doesn’t seem to play a major role. When considering individual circles, almost all of the datasets preserve the same characteristics of having a higher popularity correlation with journalist alters except Swedish journalists. The values are decreasing through outer rings until the active network. We can infer that journalist egos tend to keep their popular colleagues in the more intimate layers and allocate more cognitive resources for them. This may be an indication of having popular alters.
information sources in the inner layer either to keep information access or a way to promote their brands in general. The same results are verified by Table 10 where the patterns are almost the same with lower correlation values as mentioned before.

5. Conclusions

In this paper, we have studied the ego network structure of journalists from 17 different countries belonging to 8 different continental regions and 4 continents. We have demonstrated that how they allocate their cognitive resources among their relationships is mostly independent of the countries they are from. In general, the ego networks of journalists on Twitter all mirror the same offline human ego network structure where, according to the Dunbar’s model, our finite social cognitive capacity imposes limits to how many people we meaningfully interact with and to the intensity of these relationships. Differently from general users, though, journalists establish and maintain their relationships in a topic-oriented way. This suggests that journalists’ engagement on Twitter is mostly information-driven. Journalists also tend to have stable relationships and tend to maintain their relationships longer than politicians. As a final step, we have shown that journalists are not randomly allocating their cognitive resources among their colleagues. They establish more intimate relationships with their counterparts with similar popularity. However, popularity is not a parameter when they establish non-colleague relationships. An interesting research direction for this activity is related to leveraging the consistent properties of ego networks to detect community members that are malicious or bots. To this aim, we should first verify that the ego networks of, e.g., bots, do not mirror the ego networks of
real human users. If this intuition is confirmed, we can then leverage the lack of a regular social structure to detect bot presence within a community.
Table 9: Spearman coefficient of the rings. (*)=p<0.001, (**)=p<0.05, (***)=p<0.1.

| Dataset  | Alter Type | 1       | 2       | 3       | 4       | 5       | 6       | 7       | active |
|----------|------------|---------|---------|---------|---------|---------|---------|---------|--------|
| US       | all        | 0.12*   | -       | -       | -       | -       | -       | -0.113**| -      |
|          | journalist | 0.312*  | 0.33*   | 0.282*  | 0.308*  | 0.269*  | 0.105** | 0.164** | 0.274* |
|          | unlabelled | 0.117   | -       | -       | -       | -       | -       | -0.113**| -      |
| Canada   | all        | 0.16*   | 0.117** | 0.128** | -       | -       | -       | -       | 0.089**|
|          | journalist | 0.353*  | 0.292*  | 0.293*  | 0.204*  | 0.215*  | -       | -       | 0.228* |
|          | unlabelled | 0.147*  | 0.190** | 0.114*  | -       | -       | -       | -       | 0.095**|
| Brazil   | all        | 0.12**  | 0.141** | 0.095*** | -       | -       | -       | -       | -0.331*|
|          | journalist | 0.366*  | 0.365*  | 0.352*  | 0.213*  | 0.257*  | 0.215** | -       | -      |
|          | unlabelled | 0.088***| 0.108** | -       | -       | -       | -       | -       | -      |
| Japan    | all        | 0.102** | 0.123** | -       | -       | -       | -       | -       | -0.199***|
|          | journalist | 0.472*  | -       | 0.313** | 0.212** | -       | -       | -       | -      |
|          | unlabelled | -       | 0.128** | -       | -       | -       | -       | -       | 0.16** |
| Denmark  | all        | 0.235*  | 0.144*  | -       | -       | -       | -0.078** | -0.095***| -      |
|          | journalist | 0.372*  | 0.283*  | 0.205*  | 0.141*  | 0.097** | -       | -       | 0.162* |
|          | unlabelled | 0.146*  | 0.097** | -       | -       | -       | -0.181** | -0.107***| -      |
| Finland  | all        | 0.182*  | -       | -       | -       | -       | -0.199* | -0.168* | -0.135***| -0.14* |
|          | journalist | 0.361*  | 0.314*  | 0.257*  | 0.092***| -       | -       | -       | 0.098* |
|          | unlabelled | 0.142** | -       | -0.129**| -0.199* | -       | -       | -       | -0.149*|
| Norway   | all        | -       | -       | -       | -       | -       | -       | -0.124**| -0.149**| -0.21* |
|          | journalist | 0.27**  | 0.26**  | 0.366*  | -       | -       | -       | -       | -0.166**|
|          | unlabelled | -       | 0.114** | -0.128**| -0.142**| -       | -       | -       | -0.207*|
| Sweden   | all        | 0.371*  | 0.272*  | 0.265*  | -       | 0.144** | 0.2**   | -       | 0.152**|
|          | journalist | 0.39*   | 0.192** | 0.215*  | 0.167** | 0.144** | 0.203** | -       | 0.175**|
|          | unlabelled | 0.342*  | 0.273*  | 0.248*  | -       | 0.129** | 0.185** | -       | 0.149**|
| Greece   | all        | -0.09***| -       | -0.121**| -0.154**| -0.242* | -0.261**| -       | -0.284*|
|          | journalist | -       | -0.123**| 0.138** | -       | -       | -       | -       | 0.119**|
|          | unlabelled | -0.091**| -0.096**| -0.105**| -0.159**| -0.266**| -0.268**| -       | -0.218*|
| Italian Journalists | all | -       | -       | -       | -       | -       | -0.13** | -       | -0.196***|
|          | journalist | 0.391*  | 0.298** | 0.205** | 0.239*  | 0.252** | 0.277** | -       | 0.267* |
|          | unlabelled | -       | -       | -0.145**| -       | -       | -       | -       | -0.121**|
| Spain    | all        | 0.388*  | 0.260*  | 0.178** | 0.128** | -       | -       | -       | 0.34*  |
|          | journalist | 0.547*  | 0.532*  | 0.374*  | 0.304*  | 0.269*  | -       | -       | -      |
|          | unlabelled | 0.367*  | 0.239*  | 0.155** | 0.169***| -       | -       | -       | -      |
| France   | all        | -       | -       | -       | -       | -0.131**| -0.121**| -       | -      |
|          | journalist | 0.446*  | 0.271*  | 0.192** | -       | -       | -       | -       | -      |
|          | unlabelled | -       | -       | -       | -0.131**| -0.121**| -       | -       | -      |
| Germany  | all        | 0.156** | -       | 0.131** | -       | 0.171** | -       | 0.141** | 0.149**|
|          | journalist | 0.456*  | 0.28**  | 0.182** | 0.255** | -       | -       | 0.171** | 0.296* |
|          | unlabelled | 0.14**  | -       | 0.124** | 0.17**  | -       | -       | 0.141** | 0.158**|
| Netherlands | all | 0.201*  | 0.098*  | -       | -       | -       | -       | -       | -0.992***| -0.136**|
|          | journalist | 0.36*   | 0.333*  | 0.314*  | 0.205*  | 0.138*  | 0.128*  | -       | 0.231* |
|          | unlabelled | 0.179*  | 0.062** | -       | -       | -       | -       | -       | -      |
| Australia| all        | 0.077** | 0.148** | 0.229*  | 0.144** | 0.193*  | 0.153** | -       | 0.216* |
|          | journalist | 0.224** | 0.187** | 0.333** | 0.112*  | 0.09*   | 0.085*  | 0.014***| 0.116**|
|          | unlabelled | -       | -       | -       | -0.088**| -0.135**| -       | -       | -      |
| dataset       | alter type | ring index | 1     | 2     | 3     | 4     | 5     | 6     | 7     | active |
|---------------|------------|------------|-------|-------|-------|-------|-------|-------|-------|--------|
| US            | all        | 0.081*     | -     |       |       |       |       |       |       |        |
|               | journalist | 0.214*     | 0.226*| 0.197*| 0.217*| 0.187*| 0.072**| 0.116**| 0.189*|        |
|               | unlabelled | 0.079*     |       |       |       |       |       |       |       | -      |
|               | all        | 0.107*     | 0.082**| 0.085**| -     |       |       |       |       | 0.041**|
|               | journalist | 0.253*     | 0.202*| 0.207*| 0.143*| 0.148*| -     |       |       | 0.164* |
|               | unlabelled | 0.099*     | 0.072**| 0.075**| -     |       |       |       |       | 0.055**|
| Brazil        | all        | 0.083**    | 0.096**| 0.066**| -     |       |       |       |       |        |
|               | journalist | 0.259*     | 0.21*  | 0.243*| 0.149*| 0.174*| 0.143**|        |        | 0.227**|
|               | unlabelled | 0.058***   | 0.073**| -     |       |       |       |       |       |        |
| Japan         | all        | -          | 0.226*| -     |       |       |       |       |       |        |
|               | journalist | -          |       | 0.149**| -     |       |       |       |       | 0.139* |
|               | unlabelled | -          | -0.073**| -     |       |       |       |       |       |        |
|               | all        | -0.084**   | -     | -0.071**| -     |       |       |       |       | 0.037**|
|               | journalist | -0.067*    | -     |       |       |       |       |       |       |        |
|               | unlabelled | -0.084**   | -     |       |       |       |       |       |       |        |
| Turkey        | all        | 0.055**    |       | 0.107*| 0.125*| 0.115*| 0.07** | -     |       | -0.093*|
|               | journalist | 0.197*     |       | 0.115*| 0.07** | 0.07** |        |       |       | 0.089**|
|               | unlabelled | 0.035**    |       | -0.032**| -0.081*| -0.113*| -0.096**| -     |       | 0.009**|
| Denmark       | all        | 0.162*     | 0.101*|       |       |       |       |       |       | 0.119**|
|               | journalist | 0.257*     | 0.2*   | 0.143*| 0.099*| 0.066**| -     |       |       | 0.114* |
|               | unlabelled | 0.1*       | 0.067**| -     |       |       |       | -0.053**| -0.071**| 0.129* |
| Finland       | all        | 0.128*     |       | -0.084**| -0.071**| -     |       |       |       | -0.141**|
|               | journalist | 0.257*     | 0.217*| 0.178*| 0.062**| -     |       |       |       | 0.134**|
|               | unlabelled | 0.095**    | -0.084**| -0.071**| -     |       |       |       |        |
| Norway        | all        | 0.065***   | -     |       |       |       |       |       |       | 0.139* |
|               | journalist | 0.186**    | 0.177**| 0.241*| 0.082**| -     |       |       |       |        |
|               | unlabelled | -0.084**   | -     |       |       |       |       | -0.091**| -0.139**|        |
| Sweden        | all        | 0.257*     | 0.189*| 0.179*|       |       |       |       |       | 0.103**|
|               | journalist | 0.274*     | 0.135**| 0.149*| 0.118**| 0.095**| 0.14** | -     |       | 0.123**|
|               | unlabelled | 0.237*     | 0.187*| 0.168*|       |       |       |       |       | 0.15**  |
| Greece        | all        | -0.058***  | -     | -0.081**| -0.104**| -0.161*| -     |       | -0.136**|
|               | journalist | -0.061***  | -0.059***| -0.07**| -0.109**| -0.179*| -0.182**| -     |       | 0.145**|
|               | unlabelled | -0.061***  | -     |       |       |       |       |       |       |        |
| Italy         | all        | 0.061***   | -     |       |       |       | -0.086**| -0.098***| -0.174**| 0.074**|
|               | journalist | 0.279*     | 0.214*| 0.135**| 0.162*| 0.166*| 0.196**| 0.184**|       |        |
|               | unlabelled | -0.061***  | -     |       |       |       |       |       |       |        |
| Spain         | all        | 0.258*     | 0.184*| 0.121**| 0.083**| -     |       |       |       | 0.229* |
|               | journalist | 0.378*     | 0.373*| 0.258*| 0.212*| 0.184*| -     |       |       |        |
|               | unlabelled | 0.243*     | 0.165*| 0.102**| 0.071***| -     |       |       |       |        |
| France        | all        | -0.086**   | -     |       |       |       |       |       |       | 0.099**|
|               | journalist | 0.308*     | 0.189*| 0.13** | -     |       |       |       |       | 0.199**|
|               | unlabelled | -0.086**   | -     |       |       |       |       |       |       |        |
| Germany       | all        | 0.103**    | 0.087**|       | 0.115**| -     |       |       |       | 0.092**|
|               | journalist | 0.335*     | 0.19**| 0.12** | 0.174*| -     |       |       |       | 0.199**|
|               | unlabelled | 0.092**    | 0.034***| -     |       |       |       |       |       |        |
| Netherlands   | all        | 0.138*     | 0.062*|       |       |       |       |       |       | -0.162*|
|               | journalist | 0.248*     | 0.229*| 0.217*| 0.141*| 0.099*| 0.087*| -     |       | -0.162*|
|               | unlabelled | 0.122*     | 0.043**| -     |       |       |       |       |       |        |
| Australia     | all        | 0.051***   | 0.128**| 0.156*| 0.094**| 0.079**| 0.075*| 0.05**| 0.1*  |        |
|               | journalist | 0.154*     | 0.128**| 0.156*| 0.098**| 0.147**| 0.101**| 0.149**|       |        |
|               | unlabelled | -0.059***  | -0.086**| -     |       |       |       |       |       |        |
| All journalists| all        | 0.206*     | 0.156*| 0.129**| 0.094**| 0.079**| 0.075**| 0.05**| 0.1*  |        |
|               | journalist | 0.357*     | 0.34**| 0.235*| 0.284**| 0.264**| 0.234**| 0.223**| 0.315**|        |
|               | unlabelled | 0.177*     | 0.128*| 0.102**| 0.075**| 0.06*  | 0.056**| 0.033***| 0.077**|        |
Appendix A. The dataset: collection and pre-processing

Appendix A.1. Twitter journalist lists

Appendix A.2. Labelling users

The lists that are used to collect information about journalists are created by other Twitter users. The general aim of these lists are creating a pool for the people to access some news/information sources generally related with a specific country or region. Therefore, some news companies are also included in these lists. First of all, before labelling the Twitter accounts, we manually filtered the lists to remove news company accounts. However, there may be still accounts who are not journalist even though they are sources of information or they may be former journalists. Since our aim is to evaluate behavioral characteristics of journalists on Twitter, we have applied an automated labelling methodology to decrease noise in the datasets by detecting the accounts that are not journalist. To this aim, we have evaluated three approaches: keyword extraction, Google’s Knowledge Graph (GKG) and bot detection, and their combinations. First of all, we have labelled 513 British and 329 Italian accounts manually. While these lists are being created, they are a kind of prefiltered by the list creators. Therefore, most of the accounts are journalists. The number of accounts labelled as journalist and not-journalist can be seen in Table A.12.

Twitter API already provides us different information about the profile of users. One of them is account description where the user can supply brief description of themselves with a text that includes different information such as their profession, which football team they support, which video games they play, and so on. Considering that the journalists use Twitter as a platform for personal branding [2] and to promote content from their news websites [3], they are more likely to mention their professions and the companies they work for. Therefore, one of the tools that we can use for automatizing the labelling process is extracting the keywords from profiles to match some journalism-related keywords. However, we should consider that there are journalists from different countries, so different languages used in the descriptions of the accounts. Instead of finding different APIs or approaches for each language, we used language detection and machine translation methods to create a generalized method for keyword extraction. Considering the number of tools and accessibility, we decided to generalize the method by translating descriptions into English. As a first step, we should be able to detect the language of the description to decrease the usage of translation
API which has a limitation. We used two offline Python libraries to detect the language of the description: pyplot\(^7\) and langdetect\(^8\). If the libraries detect the same language, the language is assigned to the profile description. Otherwise, the language is accepted as undetectable. After assigning the languages, English descriptions are directly used in keyword extraction while the others are translated into English first. The descriptions of whose language was undetectable or not English are translated into English by using a Python library googletrans\(^9\) which uses Google Translate Ajax API with the feature to detect the source language automatically. After translating the descriptions, description text are tokenized and stop words are removed by using the library spacy\(^10\). As a final step, a lemmatization has been applied on the remaining words to find related lexical roots / dictionary forms of the words by using WordNet\(^11\) lexical database for English via Python library nltk\(^12\). Then, the keywords that represent the professions, in which the aim is to produce news, are selected. Selected keywords are: critic, columnist, correspondent, editor, journalist and reporter besides bigrams staff writer and senior writer. Since there are approximately 2000 keywords for each manually labelled dataset, here we visualise it with word clouds instead of tables. Extracted keywords are shown in Figure A.15.

The term knowledge graph was introduced in 2012 by Google with the motto ‘things, not strings’\(^13\). It has been announced as a first step for transition from information graph to knowledge graph where the real-world objects are not just strings but entities with inter-relationships. The power of this representation may be leveraged by searching the real-world object via Google Knowledge Graph Search API\(^14\). The API returns related top n entities as a result. These entities could be different type of entities such as persons, sport teams, cities, and so on. For labelling the journalist by using GKG, we search users’ both Twitter display name and screen name. Among

\(^7\)https://polyglot.readthedocs.io/en/latest/index.html
\(^8\)https://pypi.org/project/langdetect/
\(^9\)https://pypi.org/project/googletrans/
\(^10\)https://spacy.io/usage/linguistic-features
\(^11\)https://wordnet.princeton.edu/
\(^12\)https://www.nltk.org/
\(^13\)https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html
\(^14\)https://developers.google.com/knowledge-graph
the returned entities, the first person entity is accepted as the target user’s identity, and its profession is extracted from the entity details. If there is no person as an entity within top 10 result, the profession of the user is accepted as not journalist.

A social bot is an automated social media account that interacts with other users of the social platform according to predefined algorithms [34]. The literature about bot detection is a well studied domain. The details of the bot detection algorithms are not within the scope of this study. Therefore, the readers who are interested in details can refer to the article about social bots [35]. In the scope of this study, we used a bot detection tool named as botometer (formerly BotOrNot) \(^{15}\). Botometer is a publicly-available service to label Twitter accounts as either bot or not by using more than one thousand features [34]. Determining a Twitter account as bot or not may help to distinguish users in our dataset as journalist or not. To this aim, we leveraged the tool to label them as bot or not. There are different feature sets defined in the scope of botometer tool that can be utilized to predict bots. We are using all features to determine if an account is a bot or not with one feature set exception which is language features. Language feature set is defined only for English. Therefore, if an account’s language is assigned as English, with the method mentioned in the previous paragraph about keyword extraction, we also include language features, otherwise we use language-independent

\(^{15}\text{https://botometer.iuni.iu.edu/#!/}\)
features. As a result of bot detection algorithm, we get two score: first one is bot score which is the result of prediction algorithm, and the second one is CAP (Complete Automation Probability) score which is calculated based on Bayes’ theorem that also considers an estimate of the overall prevalence of bots. Both of the score are in the interval $[0, 1]$. In the scope of this study, we accept an account as bot, therefore not-journalist account if both the bot score and CAP score is greater than 0.5.

All three methods mentioned above which are keyword extraction, knowledge graph and bot detection methods have been applied on manually labelled 513 British and 329 Italian journalists’ Twitter accounts to see predictive power of these tools and their combinations. Prediction statistics are shown in Table A.13.

After building a classification model, one of the first question is how to evaluate the results and compare them with each other. Different statistical metrics, which are applied on confusion matrix, exist to evaluate the performance of classification methods. Precision, recall, accuracy and F1-score are some of these defined statistical calculations commonly used in the literature. However, selecting a suitable and ‘fair’ statistical function not only depends on the selected learning method but also data itself. If the data used is imbalanced, then the metric that is selected to compare learning methods must be selected carefully. Most of the popular metrics are overoptimistic in favor of major class of the data and depends on which class is accepted as positive or negative.

Accuracy is calculated as ratio between correctly classified samples and all samples as can be seen in Formula A.1. Assigning all test samples to the majority class will lead us to a high accuracy value even though we would be labelling all minor class samples wrong.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}. \quad (A.1)$$

Precision is ratio between correctly predicted positive samples and all samples predicted as positive (Formula A.2). Recall is ratio between correctly predicted positive samples and all positive samples (Formula A.3).

$$\text{precision} = \frac{TP}{TP + FP}. \quad (A.2)$$

$$\text{recall} = \frac{TP}{TP + FN}. \quad (A.3)$$

38
F1-score is harmonic mean of precision and recall.

\[
F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}.
\]  

(A.4)

As can be seen from Formulae A.2, A.3 and A.4, correctly predicted negative samples (TN) are not considered during the calculation of these metrics. In the case of imbalanced dataset, if we swap the positive and negative class definitions to have positive class as the one which has more samples, we will end up with higher metric values. One suggested solution to overcome this problem is calculating metrics twice by swapping the class and getting the average score. However, this method is claimed to be biased in the study [36].

The area under the curve of Precision-Recall (PR AUC) and the area under the curve of the Receiver Operating Characteristic (ROC AUC). These curves are constructed by changing the confusion matrix threshold starting from lowest to highest by an increment value. However, since our methods based on three methods (keyword extraction/matching, Google’s knowledge graph and bot detection) where either there is no confusion matrix threshold or it is expensive in terms of time consumption, AUC metric is not feasible for our evaluation method. In keyword extraction/matching approach, we have no confusion matrix threshold because extracted keywords either match or not with defined keywords that define a journalist in the scope of our study. In Google’s knowledge graph approach, we change the number of top ranked results returned which is expensive because of API rate limitation. The only approach that has a confusion matrix threshold which we can easily use is bot detection. However, since we are also considering the combination of these methods to label the users, we didn’t prefer to use PR AUC and ROC AUC metrics.

Considering our dataset is imbalanced, we prefer to use an evaluation metrics which considers the importance of labelling both positive and negative samples, and has a class swap property. To this aim, we select Matthews’ Correlation Coefficient (MCC) that was introduced in the study [26] for evaluation/comparison of secondary structure prediction of proteins. MCC is a specific case of \( \phi \) coefficient [36]. It is a Pearson product-moment correlation coefficient between actual and predicted classes as shown in Formula A.5. MCC values are in the interval \([-1, +1]\) where –1 is perfect misclassification and +1 is perfect classification. MCC scores of manually labelled datasets can be seen in Figure A.16 in which \( g \) represents google knowledge graph, \( k \)
represents keyword extraction, \(b\) represents boot detection methods while \& represents logical—\textit{and} and \(\mid\) represent logical—\textit{or} operation. In Italian journalists dataset and all journalists dataset (British and Italian journalists are merged), \(g\mid k\), \(b\& (g\mid k)\), \(g\mid (b\& g)\), and \(k\mid (b\& g)\) have highest MCC scores while in British journalists dataset, \(g\mid k\) and \(k\mid (b\& g)\) have highest scores. Since \(g\mid k\) achieves highest score by using less features, and it is a faster method since it doesn’t use the bot detection API, we selected \(g\mid k\) as the labelling method.

\[
MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}
\]  

(A.5)
Appendix A.3. Tweeting volume all dataset

Appendix A.4. DBScan outlier all datasets

Appendix A.5. Twitting activity all datasets

Appendix A.6. Stationarity

The last check we perform is about the stationarity of the Twitter activity of the users in our dataset. As previously shown in the literature ([23], [18], [29]), newly registered social network users add relationships into their network and interact with others at a higher rate compared to long-term users. After a while, their activity stabilizes. Newly registered users are thus outliers with respect to the general population of users, and they should be discarded from the analysis. To this aim, we calculate the average weekly tweet number of all users. We align their first week of the observed timeline (starting from the first observed tweet) and get the average of the dataset. To be fair and keep this stability calculation at the individual user level, we use mean normalization for each user. As a result, values are located in the range $[-1, 1]$ where zero values are close to the mean value. In Figure A.20, we show the first 80 weeks of averaged normalized tweeting frequency per week since the number of active users decreases drastically after 80th week. For the stability analysis, we only show fully observed users since we cannot observe the full timeline of the partially observed ones. As can be seen from the figure, we do not have a drastic change in tweeting activity for none of the datasets. Therefore, we do not discard any part of the timeline as a transient period to stabilize tweeting activity.

Appendix B. Alter number distribution all datasets

Appendix C. Active network size all datasets

Appendix D. Optimal circle numbers all datasets

Appendix E. Relations activated by hashtags per ring

Appendix F. Mean number of hashtags per alter in ring

Appendix G. Contact frequency per ring
| Continent | Region | Country | Twitter List(s) |
|-----------|--------|---------|-----------------|
| Americas  | Northern America | United States of America | https://twitter.com/r0eland/lists/usa-journalist-one https://twitter.com/r0eland/lists/usa-journalist-two https://twitter.com/r0eland/lists/usa-journalist-three https://twitter.com/r0eland/lists/usa-journalist-four https://twitter.com/r0eland/lists/usa-journalist-five |
|          | Canada  | https://twitter.com/canadatrail/lists/canada-journalism https://twitter.com/Clem_Labasse/lists/journalistes-canada https://twitter.com/NewsLive/lists/canada-news https://twitter.com/swarajmann/lists/reporters-qp https://twitter.com/DeeDub8/lists/cdnjournalists |
|          | Southern America | Brazil | https://twitter.com/josemurilo/lists/mediabr https://twitter.com/danielbastos/lists/jornalistas-brasil https://twitter.com/jornalistas/lists/jornalistas-do-brasil https://twitter.com/pauliobazaar/lists/jornalistas-tv |
|          | Asia     | Eastern Asia | Japan | https://twitter.com/morimori_naha/lists/%E3%82%B8%E3%83%A3%E3%83%BC%E3%83%AA%E3%82%B9%E3%83%88%EF%BC%9Ajournalist https://twitter.com/news_paper_lists/%E3%82%B8%E3%83%A3%E3%83%BC%E3%83%AA%E3%82%B9%E3%83%88%EF%BC%9Ajournalist https://twitter.com/tumblr_lists/journalists https://twitter.com/politicometrogr/lists/gr-journalists https://twitter.com/sakis37/lists/gr-jornalists |
|          | Western Asia | Turkey | https://twitter.com/LeventResul/lists/turk-basin-turkish-press https://twitter.com/BitcoinHaber/lists/uzman-gazetecilik https://twitter.com/suyunbige/lists/yurtici-journalist https://twitter.com/zeydnep/lists/gazeteci-yorumcu |
|          | European Northern Europe | United Kingdom of Great Britain and Northern Ireland | https://twitter.com/journalismnews/lists/uk-national-journalists https://twitter.com/journalismnews/lists/uk-national-journalists-2 |
|          | Denmark  | https://twitter.com/MorgTorben/lists/journalistik https://twitter.com/SNordby/lists/danske-journalister https://twitter.com/ernstpoulsen/lists/danske-journalister https://twitter.com/denmarktrail/lists/denmark-journalism1 |
|          | Finland  | https://twitter.com/Astatenhunen/lists/toimittajat https://twitter.com/ulkoministerio/lists/toimittajat https://twitter.com/anttihirvonen/lists/suomalaiset-toimittajat https://twitter.com/Ragamuf_/lists/toimittajat |
|          | Norway   | https://twitter.com/ernslars/lists/norske-journalister https://twitter.com/akaarbo/lists/journalister https://twitter.com/P_Deshayes/lists/journalistes-norv-egeiens https://twitter.com/alk1986/lists/norske-journalister |
|          | Sweden   | https://twitter.com/anton_andersson/lists/journalister https://twitter.com/makthavare/lists/journalister https://twitter.com/fredrikhardt/lists/journalister https://twitter.com/mymlan/lists/journalister https://twitter.com/ceciliaaxel/lists/journalister |
|          | Southern Europe | Greece | https://twitter.com/nekosensitive/lists/greek-journalists https://twitter.com/Gregorios_lists/greek-journalists https://twitter.com/kritikoubrik/lists/greek-journalists https://twitter.com/kleonitis_lists/greek-journalists https://twitter.com/kostas_lists/greek-journalists https://twitter.com/myrtos_lists/greek-journalists https://twitter.com/kafkas_lists/greek-journalists https://twitter.com/kostas_lists/greek-journalists https://twitter.com/thellis_lists/greek-journalists |
|          | Italy    | https://twitter.com/Stampa_Tweet/lists/giornalisti https://twitter.com/Stampa_Tweet/lists/giornalisti https://twitter.com/Stampa_Tweet/lists/giornalisti https://twitter.com/Stampa_Tweet/lists/giornalisti |
|          | Spain    | https://twitter.com/saxom/lists/periodistas https://twitter.com/saxom/lists/periodistas https://twitter.com/periodistas/lists/periodistas https://twitter.com/periodistas/lists/periodistas https://twitter.com/josemaria_lists/periodistas |
|          | Western Europe | France | https://twitter.com/ymobactus/lists/journalistes https://twitter.com/davidbauer/lists/journalistes https://twitter.com/NL_lists/journalistes https://twitter.com/nl_lists/journalistes |
|          |          | Germany | https://twitter.com/davidbauer/lists/journalisten https://twitter.com/nl_lists/journalisten https://twitter.com/nl_lists/journalisten |
|          | Oceania  | Australia and New Zealand | Australia | https://twitter.com/caniumhoover/lists/australian-journalists |
Table A.12: Manually labelled journalists

| dataset            | journalist | not journalist | imbalance ratio |
|--------------------|------------|----------------|-----------------|
| British Journalists| 434        | 79             | 0.846           |
| Italian Journalists| 316        | 13             | 0.960           |
| All                | 750        | 92             | 0.890           |
Table A.13: Prediction statistics.

| dataset       | method name | TP  | FP  | TN  | FN  | precision | recall | accuracy | F1   |
|---------------|-------------|-----|-----|-----|-----|-----------|--------|----------|------|
| British Journalists | b           | 428 | 77  | 2   | 6   | 0.84      | 0.98   | 0.93     | 0.91 |
| g             | 59          | 3   | 76  | 375 | 95  | 0.13      | 0.26   | 0.23     | 0.23 |
| k             | 342         | 54  | 92  | 93  | 78  | 0.78      | 0.77   | 0.85     |      |
| b & g         | 59          | 3   | 76  | 375 | 95  | 0.13      | 0.26   | 0.23     |      |
| b & k         | 347         | 55  | 97  | 93  | 77  | 0.77      | 0.76   | 0.84     |      |
| g & k         | 44          | 1   | 78  | 390 | 97  | 0.1       | 0.23   | 0.18     |      |
| b | g          | 428         | 77  | 2   | 6   | 0.84      | 0.98   | 0.93     | 0.91 |
| b | k          | 433         | 78  | 1   | 1   | 0.84      | 0.99   | 0.94     | 0.91 |
| g | k          | 357         | 27  | 52  | 77  | 92  | 0.82      | 0.79   | 0.87     |      |
| b & (g | k)      | 352         | 26  | 53  | 82  | 93  | 0.81      | 0.78   | 0.86     |      |
| g & (b | k)      | 357         | 27  | 52  | 77  | 92  | 0.82      | 0.79   | 0.87     |      |
| all | b         | 44          | 1   | 78  | 390 | 97  | 0.1       | 0.23   | 0.18     |      |
| all | g         | 352         | 26  | 53  | 82  | 93  | 0.81      | 0.78   | 0.86     |      |
| all | k         | 357         | 27  | 52  | 77  | 92  | 0.82      | 0.79   | 0.87     |      |
| Italian Journalists | b           | 316 | 13  | 0   | 0   | 0.96      | 1      | 0.96     | 0.97 |
| g             | 70          | 1   | 12  | 246 | 98  | 0.22      | 0.24   | 0.36     |      |
| k             | 215         | 7   | 6   | 101 | 96  | 0.68      | 0.67   | 0.79     |      |
| b & g         | 70          | 1   | 12  | 246 | 98  | 0.22      | 0.24   | 0.36     |      |
| b & k         | 215         | 7   | 6   | 101 | 96  | 0.68      | 0.67   | 0.79     |      |
| g & k         | 37          | 1   | 12  | 279 | 97  | 0.11      | 0.14   | 0.2      |      |
| b | g          | 316         | 13  | 0   | 0   | 0.96      | 1      | 0.96     | 0.97 |
| b | k          | 316         | 13  | 0   | 0   | 0.96      | 1      | 0.96     | 0.97 |
| g | k          | 248         | 7   | 6   | 68  | 97  | 0.78      | 0.77   | 0.86     |      |
| b & (g | k)      | 248         | 7   | 6   | 68  | 97  | 0.78      | 0.77   | 0.86     |      |
| g & (b | k)      | 215         | 7   | 6   | 101 | 96  | 0.68      | 0.67   | 0.79     |      |
| all | b         | 316         | 13  | 0   | 0   | 0.96      | 1      | 0.96     | 0.97 |
| all | g         | 248         | 7   | 6   | 68  | 97  | 0.78      | 0.77   | 0.86     |      |
| all | k         | 248         | 7   | 6   | 68  | 97  | 0.78      | 0.77   | 0.86     |      |
| All          | b           | 744         | 90  | 2   | 6   | 0.89      | 0.99   | 0.98     | 0.91 |
| g             | 129         | 4   | 88  | 621 | 96  | 0.17      | 0.25   | 0.29     |      |
| k             | 557         | 32  | 60  | 193 | 94  | 0.74      | 0.73   | 0.82     |      |
| b & g         | 129         | 4   | 88  | 621 | 96  | 0.17      | 0.25   | 0.29     |      |
| b & k         | 552         | 31  | 61  | 198 | 94  | 0.73      | 0.72   | 0.82     |      |
| g & k         | 81          | 2   | 90  | 669 | 97  | 0.1       | 0.2    | 0.19     |      |
| b | g          | 744         | 90  | 2   | 6   | 0.89      | 0.99   | 0.98     | 0.93 |
| b | k          | 749         | 91  | 1   | 1   | 0.89      | 0.99   | 0.98     | 0.94 |
| g | k          | 605         | 34  | 58  | 145 | 94  | 0.8       | 0.78   | 0.87     |      |
| b & (g | k)      | 600         | 33  | 59  | 150 | 94  | 0.8       | 0.78   | 0.86     |      |
| g & (b | k)      | 552         | 31  | 61  | 198 | 94  | 0.73      | 0.72   | 0.82     |      |
| b | (g & k)    | 744         | 90  | 2   | 6   | 0.89      | 0.99   | 0.98     | 0.93 |
| k | (b & g)    | 605         | 34  | 58  | 145 | 94  | 0.8       | 0.78   | 0.87     |      |
| all | b         | 81          | 2   | 90  | 669 | 97  | 0.1       | 0.2    | 0.19     |      |
| all | g         | 749         | 91  | 1   | 1   | 0.89      | 0.99   | 0.98     | 0.94 |
Figure A.17: Tweeting volume VS active users over time
Figure A.18: Histogram of tweet frequency
Figure A.19: Observed timeline length VS average tweet frequency
Figure A.20: Stationarity analysis
Figure B.21: Histogram of number of alters per ego
Figure C.22: Histogram of active networks size
Figure D.23: Optimal number of circles per ego (red line: average, blue line: median)
Figure E.24: Average percentage of relations activated by hashtags per ring with confidence intervals
Figure F.25: Mean number of hashtags per alter in ring
Figure G.26: Contact frequency per ring
References

[1] R. Dunbar, V. Arnaboldi, M. Conti, A. Passarella, The structure of online social networks mirrors those in the offline world, Social Networks 43 (2015) 39–47.

[2] L. Canter, Personalised tweeting, Digital Journalism 3 (6) (2015) 888–907.

[3] F. M. Russell, M. A. Hendricks, H. Choi, E. C. Stephens, Who sets the news agenda on twitter? journalists’ posts during the 2013 us government shutdown, Digital Journalism 3 (6) (2015) 925–943.

[4] C. Brems, M. Temmerman, T. Graham, M. Broersma, Personal branding on twitter: How employed and freelance journalists stage themselves on social media, Digital Journalism 5 (4) (2017) 443–459.

[5] N. Y. Lee, Y. Kim, Y. Sang, How do journalists leverage twitter? expressive and consumptive use of twitter, The Social Science Journal 54 (2) (2017) 139–147.

[6] P. MacGregor, A. Balcytiene, L. Fortunati, V. Nuust, J. O’Sullivan, N. Roussou, R. Salaverría, M. Sarrica, A cross-regional comparison of selected european newspaper journalists and their evolving attitudes towards the internet-including a single-country focus on the uk, Journalism 12 (5) (2011) 627–646.

[7] A. S. Weiss, The digital and social media journalist: A comparative analysis of journalists in argentina, brazil, colombia, mexico, and peru, International Communication Gazette 77 (1) (2015) 74–101.

[8] F. Hanusch, T. Hanitzsch, Comparing journalistic cultures across nations: What we can learn from the worlds of journalism study (2017).

[9] M. Everett, S. P. Borgatti, Ego network betweenness, Social networks 27 (1) (2005) 31–38.

[10] N. Lin, K. S. Cook, R. S. Burt, Social capital: Theory and research, Transaction Publishers, 2001.

[11] C. McCarty, Structure in personal networks, Journal of social structure 3 (1) (2002) 20.
[12] R. A. Hill, R. I. Dunbar, Social network size in humans, Human nature 14 (1) (2003) 53–72.

[13] W.-X. Zhou, D. Sornette, R. a. Hill, R. I. M. Dunbar, Discrete hierarchical organization of social group sizes., Proceedings. Biological sciences / The Royal Society 272 (1561) (2005) 439–444.

[14] A. Sutcliffe, R. Dunbar, J. Binder, H. Arrow, Relationships and the social brain: integrating psychological and evolutionary perspectives, British journal of psychology 103 (2) (2012) 149–168.

[15] V. Arnaboldi, M. Conti, A. Passarella, R. Dunbar, Dynamics of personal social relationships in online social networks, in: Proceedings of the first ACM conference on Online social networks - COSN '13, ACM Press, New York, New York, USA, 2013, pp. 15–26.

[16] R. Dunbar, The social brain hypothesis, Evolutionary Anthropology 9 (10) (1998) 178–190.

[17] S. G. Roberts, R. I. Dunbar, T. V. Pollet, T. Kuppens, Exploring variation in active network size: Constraints and ego characteristics, Social Networks 31 (2) (2009) 138–146.

[18] G. Miritello, R. Lara, M. Cebrian, E. Moro, Limited communication capacity unveils strategies for human interaction, Scientific reports 3 (1) (2013) 1–7.

[19] V. Arnaboldi, A. Passarella, M. Conti, R. Dunbar, Structure of Ego-Alter Relationships of Politicians in Twitter, Journal of Computer-Mediated Communication 22 (5) (2017) 231–247.

[20] B. Gonçalves, N. Perra, A. Vespignani, Modeling users’ activity on twitter networks: Validation of dunbar’s number, PloS one 6 (8) (2011) e22656.

[21] D. Quercia, L. Capra, J. Crowcroft, The social world of twitter: Topics, geography, and emotions., ICWSM 12 (2012) 298–305.

[22] S. Aral, M. Van Alstyne, The diversity-bandwidth trade-off, American Journal of Sociology 117 (1) (2011) 90–171.
[23] V. Arnaboldi, M. Conti, A. Passarella, R. I. Dunbar, Online social networks and information diffusion: The role of ego networks, Online Social Networks and Media 1 (2017) 44–55.

[24] C. Quadri, M. Zignani, S. Gaito, G. P. Rossi, Feature-rich ego-network circles in mobile phone graphs: tie multiplexity and the role of alters, in: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2018, pp. 1280–1285.

[25] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on pattern analysis and machine intelligence 24 (5) (2002) 603–619.

[26] B. Matthews, Comparison of the predicted and observed secondary structure of t4 phage lysozyme, Biochimica et biophysica acta 405 (2) (1975) 442—451. doi:10.1016/0005-2795(75)90109-9. URL https://doi.org/10.1016/0005-2795(75)90109-9

[27] C. Boldrini, M. Toprak, M. Conti, A. Passarella, Twitter and the press: An ego-centred analysis, in: Companion Proceedings of the The Web Conference 2018, WWW ’18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018, p. 1471–1478. doi:10.1145/3184558.3191596. URL https://doi.org/10.1145/3184558.3191596

[28] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, Vol. 96, 1996, pp. 226–231.

[29] B. Viswanath, A. Mislove, M. Cha, K. P. Gummadi, On the evolution of user interaction in facebook, in: Proceedings of the 2nd ACM workshop on Online social networks, 2009, pp. 37–42.

[30] V. Arnaboldi, M. Conti, A. Passarella, F. Pezzoni, Ego networks in twitter: an experimental analysis, in: 2013 Proceedings IEEE INFOCOM, IEEE, 2013, pp. 3459–3464.

[31] L. Molyneux, What journalists retweet: Opinion, humor, and brand development on twitter, Journalism 16 (7) (2015) 920–935.
[32] N. S. Chok, Pearson’s versus spearman’s and kendall’s correlation coefficients for continuous data, Ph.D. thesis, University of Pittsburgh (2010).

[33] J. Hauke, T. Kossowski, Comparison of values of pearson’s and spearman’s correlation coefficients on the same sets of data, Quaestiones geographicae 30 (2) (2011) 87–93.

[34] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, Botornot: A system to evaluate social bots, in: Proceedings of the 25th International Conference Companion on World Wide Web, WWW ’16 Companion, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2016, p. 273–274. doi:10.1145/2872518.2889302.
URL https://doi.org/10.1145/2872518.2889302

[35] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (7) (2016) 96–104. doi:10.1145/2818717.
URL https://doi.org/10.1145/2818717

[36] D. Chicco, G. Jurman, The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation, BMC Genomics 21 (2020).