

Twitter and the Press: an Ego-Centred Analysis

Chiara Boldrini

IIT-CNR

Pisa, Italy

chiara.boldrini@iit.cnr.it

Marco Conti

IIT-CNR

Pisa, Italy

marco.conti@iit.cnr.it

Mustafa Toprak

IIT-CNR

Pisa, Italy

mustafa.toprak@iit.cnr.it

Andrea Passarella

IIT-CNR

Pisa, Italy

andrea.passarella@iit.cnr.it

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, 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. 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. Moreover, the ego networks of journalists are much stabler than those of politicians and generic users, and the ego-alter ties are often information-driven.

CCS CONCEPTS

• **Networks** → **Online social networks**; • **Human-centered computing** → **Social network analysis**; • **Mathematics of computing** → *Graph theory*;

KEYWORDS

online social networks, ego networks, Twitter, journalists

ACM Reference Format:

Chiara Boldrini, Mustafa Toprak, Marco Conti, and Andrea Passarella. 2018. Twitter and the Press: an Ego-Centred Analysis. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23-27, 2018, Lyon, France*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3184558.3191596>

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23-27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191596>

1 INTRODUCTION

Online Social Networks are one of the most prominent examples of cyber-physical convergence: social relationships that previously could exist only in the offline world are now transported into the virtual, online dimension and new interactions are enabled that were not possible before. OSN are thus at the same time very similar and very different from offline social networks, and for this reason several researchers have tried to understand if and to which extent they obey to the same common principles.

Ego networks are the graph-based abstraction that is typically used to study the social relations between an individual and its peers [8, 10, 12, 13]. 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 [16]. In an ego network, the individual, referred to as *ego*, is at the center of the graph, and the edges connect her to the peers (called *alters*) with which she interacts. The ego-alter tie strength is typically computed as a function of the frequency of interactions between the ego and the alter. Grouping these ties 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. According to the *social brain hypothesis* from evolutionary psychology [6], the existence and sizes of the groups are determined by the maximum cognitive capacity of the brain that humans can allocate to *meaningful* social relationships, i.e., beyond the mere level of acquaintances. 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 [10, 18]. Beyond the Dunbar number, relationships are just acquaintances and their maintenance has a negligible effect on the cognitive resources. Figure 1 illustrates the typical sizes of each layer (1.5, 5, 15, 50, 150).

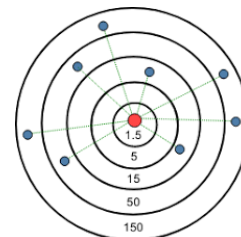


Figure 1: Layered structure of human ego networks

Quite interestingly, ego networks formed through many interaction means, including face-to-face contacts, letters, phone calls, emails, and, remarkably, also OSN, are well aligned with this model [18]. Specifically, Dunbar et al. [7] have found very similar properties also in Facebook and Twitter ego networks. 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 [9] 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 [14]. 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 [1].

Thanks to its public API and its impressive number of users (330 million active users as of October 2017¹), Twitter is one of the most studied OSN in the related literature on social networking. Generic users on Twitter have received a lot of attention, where the aim was to characterise how they engage with the platform [11], interact with each other, and how real life affects what we observe on Twitter and vice versa [17]. However, users on Twitter are not created equal: their behaviour on the platform is intertwined to their social status and their job. For this reason, researchers have recently focused their attention to special classes of Twitter users that, for one reason or another, can be singled out. For example, Arnaboldi et al. [3] have studied how politicians engage with other people on Twitter, highlighting that, while the same ego network structure appears, there is a significant difference with respect to generic users in terms of ego network variability and amount of information-driven relationships.

Another interesting category of Twitter users is that of journalists. Previous research has highlighted how journalists use Twitter as a platform for personal branding [4] and to promote content from their news websites [15]. In this sense, journalists share with politicians many of the same motivations for using Twitter. However, no previous work has studied how journalists create and maintain social relationships on Twitter, and whether this resembles what we observe for similar categories of users or for generic users. Thus, in this paper, exploiting the ego network abstraction described above, we investigate the activity of journalists on Twitter in terms of how they engage with other users.

We found that journalists are significantly more active than the average generic users and that they have a marked preference for social communications involving retweets. From the ego network’s standpoint, journalists feature a layered structure that is very similar to the general ego network model of Figure 1, sometimes much more than what has been observed for other types of Twitter users. Specifically, the size of the active network is much closer to the theoretical value of 150, and very similar to the one observed in offline ego networks before the diffusion of OSN. This is a surprising result, that may indicate that journalists invest a large portion,

compared to other types of users, of their social capacity on Twitter. The ego networks of journalists are also very stable over time, and this is very different from the findings about politicians and generic users on Twitter [3]. This may be due to the fact that journalists use Twitter less as a leverage to win people over and more as a social communication tool. Finally, we show that many ego-alter relationships are information-driven: they are often activated by a hashtag, their future interactions also involve hashtags, and this holds true approximately for all the layers with the same intensity.

2 THE DATASET

We downloaded the timeline for the journalists belonging to the Italian journalists list at https://twitter.com/stampa_tweet/lists/giornalisti. This list comprises 492 journalists that are the most popular in Italy. Due to the Twitter API limitations, only the 3200 most recent tweets could be downloaded for each journalist. Of these 492 members, we discard those with a protected timeline and with zero tweets, ending up with 486 journalists. All of them are long-time Twitter users, since their most recent registration dates back to 2012. However, as we will discuss later on, not all of them are equally engaged with the Twitter platform.

2.1 Observability

The 3200 tweets limitation imposed by the Twitter API affects journalists’ timelines non uniformly. Specifically, around 69% of users (corresponding to 335 out of the 486 journalists) have posted more tweets than what we are able to download. We refer to these users as *partially observed*. This is a significant difference with respect to the dataset analysed in [2], where, for 98% of users, 3200 tweets were enough to cover the whole Twitter activity. 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 is 0.6 for the fully observed users and 6.0 for the partially observed ones (which, anyway, is still a feasible value for truly human Twitter activity).

We refer to the portion of timeline that we are able to observe as *observed timeline*. When the 3200 tweets limitation does not kick in, the observed timeline overlaps with the *active timeline*, which we estimate as the Twitter activity since the user registration on Twitter². While the distinction between partially observed and fully observed users is binary, the actual coverage of the observed timeline with respect to the active timeline is more nuanced. Hence, the percentage of active timeline observed typically changes on a per user basis. The direct consequence of the uneven time coverage of the observed timelines is that the number of tweeting journalists in our dataset constantly increases over time starting from 2007, because the most active users are present only in the most recent intervals. Figure 2 shows how this impacts on the daily volume of tweets generated over time. We plot in gray the time series of total observed tweets per day, while in blue we plot the number of users with active observed timeline (i.e., for which the coverage of the 3200 tweets has already started). While the growth in the number of active users is approximately linear, the increase in

¹Source: Wikipedia

²The real active timeline would be the time between the first and last tweet. However, in this case the time of the first tweet is unknown due to the API limitations.

the observed tweets is exponential. This further highlights one of the characteristics of this dataset: the most active users are only observed for a short amount of time but they are the ones that impact the most on the number of generated tweets.

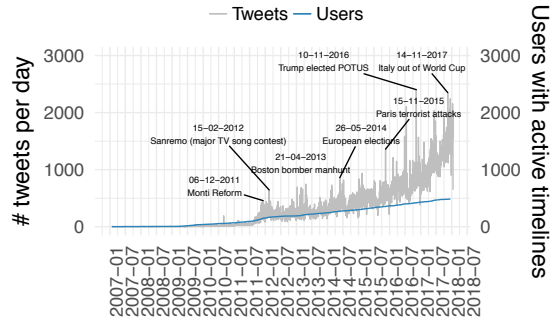


Figure 2: Tweeting volume VS active users over time

2.2 Data cleaning and filtering

Not all the downloaded timelines are suitable for either a general analysis of the behaviour of journalists on Twitter or for the ego network analysis. In fact, some journalists may have left Twitter long ago, others may use it only sporadically, others may feature an unstable ego network (we will discuss what this means in a later section). In the following, we discuss how to identify the Twitter users that should be considered for our analysis.

2.2.1 Users that have left Twitter. First, we focus on identifying users that are not actively engaged with the platform anymore. The *inactive life* of a user, defined as the time period between the last observed tweet and the time of downloading, is an indirect measure of inactivity. In fact, long inactive periods are a telltale sign of decreased engagement.

In previous works [2], a threshold-based approach mimicking Twitter's 6-months inactivity criterion is employed: a user was considered to have abandoned Twitter if her inactive life was longer than six months. However, here we argue that six months of inactivity may be perfectly in line with the historic behaviour of a user, without being suggestive of her having abandoned the platform. Therefore, we use a new classifier based on the concept of *intertweet time*, defined as the time interval between two consecutive tweets. For each journalist i , we observe a certain characteristic distribution of intertweet times (ITT_i). If the inactive lifespan $T_i^{inactive}$ of user i is significantly larger than the maximum intertweet time previously observed for the user, then it is likely that the user has indeed abandoned the platform. Specifically, we allow for a six-months grace period (in the spirit of the considerations in [2]) and we assume a user i is still active if $T_i^{inactive} < \max\{ITT_i\} + 6$ months. Using this classifier, 11 users are marked as having abandoned Twitter at the time we downloaded the dataset.

2.2.2 Regular users. While the previous classifier was concentrating on the inactive life of a user, we now focus on the observed life, and we investigate the level of regularity in Twitter activity.

Like in [3], we assume that a Twitter user is a regular user if she posts at least one tweet every 3 days (as an average frequency within each month) for at least 50% of the total number of months of their activity. If she does not, the user is classified as sporadic. In Figure 3 we combine the results of the abandonment classifier and the regularity classifier. As expected, partially observed users are generally more active and regular than fully observed ones. It is interesting to note that there are two users marked as regulars that have then stopped using Twitter. We manually checked these two users to confirm the classification and we found out that indeed one of them had died and one had changed its Twitter handle. Occasional users are not interesting for our analysis, since their engagement with the platform is low and so is their cognitive involvement. For this reason, in the following we will focus on regular and active users only (corresponding to the green bar in Figure 3), which amount to 387 journalists in total.

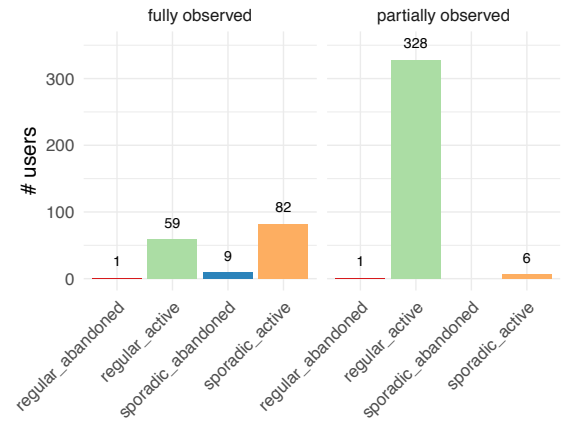


Figure 3: User classification

2.2.3 Stationarity. In our analysis, we are not only interested in active and regular users, but we also need to make sure that their Twitter activity has already stabilised. In fact, it is a frequent finding in the related literature that users tend to be more active during the initial interactions with the platform, reaching a steady state later on, when the engagement is somewhat consolidated. In order to verify whether this is the case also in our dataset, we compute for each user the average number of tweets generated each week. For this analysis, we align all observed timelines to start at the same time (i.e., time 0 for each given user is the time of the first observed tweet of the user). Then, we average these values across all users. Please note that we restrict our analysis to the first 84 weeks, since the number of active users abruptly decreases afterwards. For the sake of a fair comparison between users, we have rescaled the weekly frequencies of each user using the mean normalization technique. This procedure yields values in the range $[-1, 1]$, with values around zero being close to the average frequency. The results are shown in Figure 4, only for the 59 active and regular users that are fully observed³. It is clear that

³Please note that for partially observed users we are missing the initial portion of the timeline, so this analysis cannot be replicated.

there is not a significant variation in the activity between first and last portion of the timelines. Hence, we do not discard a transient period for the users in our dataset.

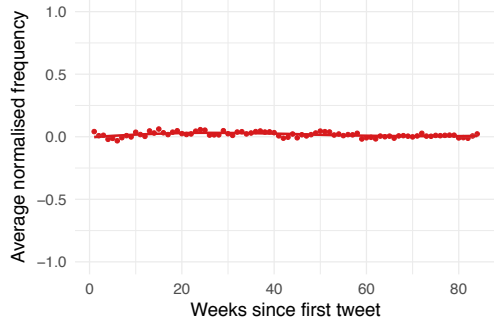


Figure 4: Stationarity analysis

2.3 Dataset overview

The main properties of the dataset resulting from the filtering performed in the previous section are summarised in Table 1. With an average active life of about 8 years, all these journalists are long-time Twitter users. However, the number of tweets that they have generated during their entire life varies significantly between users. Among the tweets that we observe, an average of 68% are social, i.e., following the definition in [2], involve forms of direct communications such as replies, mentions or retweets. Among these, retweets⁴ are by far the most popular mode for interaction, taking up half of the social communication space. Retweets were popular also for the politicians analysed in [3], but their percentage was around 40%, with mentions being used, by and large, with roughly the same frequency. The popularity of retweets is confirmed when looking at the most popular communication mode per user (Figure 5, where *indirect* means tweets that are neither mentions, nor replies, nor retweets): retweets are the go-to mode for more than 50% of the users. Figure 6 shows the breakdown across the three types of social tweets, grouping users by their dominant type of tweet. It shows that retweets rank always as first or second most popular social mode, regardless of the predominant mode.

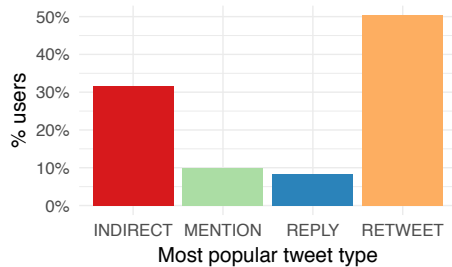


Figure 5: Most popular tweeting activity

Table 1 also reveals a great variability in the number of tweets per day, whose distribution is shown in Figure 7. Figure 7 shows

⁴Please note that, for the purpose of this analysis, quote tweets are treated as simple retweets.

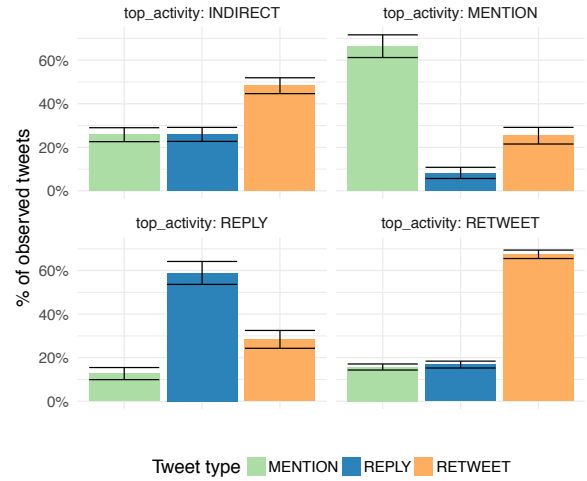


Figure 6: Tweets per top activity

that while most journalists post around 10-15 tweets per day, some users are clear outliers, with daily tweets well above 30. These users are marked as outliers when using standard clustering techniques such as DBSCAN on the tweet frequency. These outliers could have a significant impact on the ego network statistics, due to their intense Twitter activity, and should be studied separately. However, as shown in Figure 8, the observed timeline for these very active users is too short to satisfy the stability requirement (at least one year between the first and last observed tweets) for ego networks that has been typically used in the related literature [2]. For this reason, these users will be excluded from the analysis in Section 3. The average tweet frequency when excluding outliers is 4.17 tweets per day, which is still much higher than the one observed⁵ in politicians (between 2.6 and 3.2 depending on the dataset [3]).

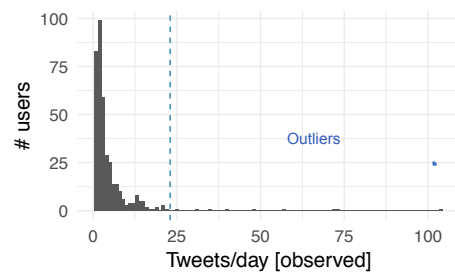
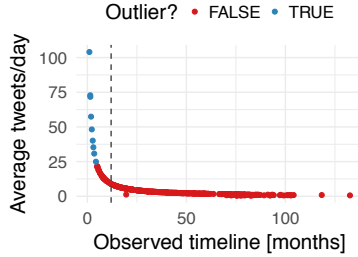


Figure 7: Histogram of tweet frequency

⁵Please note that the comparison between our results about journalists and those in [3] about politicians is a fair one. In fact, the filtering based on user regularity and activity intensity described in Section 2.2 was also performed very similarly in [3]. More specifically, the filtering described in Section 2.2.2 is exactly the same as the one used in [3]. Our filter described in Section 2.2.1 is slightly more conservative than the one used in [3], hence it potentially retains more users that are less active. Despite this, as discussed later in the paper, we find that journalists are generally more active than politicians and generic users.

Table 1: Summary statistics

	active life [years]	tot tweets	observed tweets	tweets/day	% social	% replies	% retweets	% mentions
mean	7.88	16,319	3061.34	5.33	68.26	22.33	54.06	23.62
sd	1.33	22,199	435.06	9.27	21.11	18.35	22.02	20.03

**Figure 8: Observed timeline length VS average tweet frequency**

3 EGO NETWORK ANALYSIS

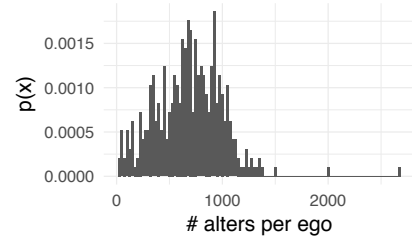
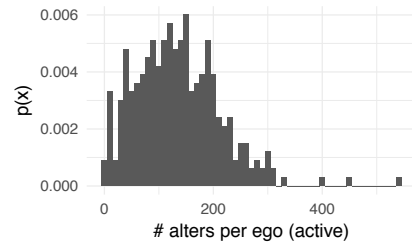
We now focus our attention on the ego networks of the journalists that have been selected according to the discussion in Section 2. For each of these journalists, the strength of ego-alter ties are inferred from the frequency of *direct tweets* (mentions, reply, or retweets) between the ego and the alters, which is normally considered as a proxy for their emotional closeness [2, 7]. Similarly to the related literature, we define as active tie one for which the ego and the alter exchange at least one message per year.

3.1 Static ego networks

The static view of an ego network is 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.

Figure 9 shows the distribution of the number of alters per ego, i.e., the total number of social ties, including both active and inactive ones. When compared with what was observed in [7] for generic Twitter users, we notice that, for journalists, the number of alters per ego is about twice as large as that of regular users. This implies that journalists tend to contact significantly more alters. This is in line with journalists being prominent users that stand out among the Twitter crowd. However, as discussed before, only active relationships consume cognitive resources and thus are affected by the human cognitive constraints. For this reason, in Figure 10, we consider the active network, i.e., the ego network encompassing only relationships with contact frequency higher than one tweet per year. While the size of the active network is substantially different from that of the complete ego network, the proportional difference is maintained when comparing what we observe for journalists against the results shown in [7] for generic Twitter users⁶.

⁶While the initial filtering used in [7] is different from the one discussed in this paper (Section 2.2), the comparison between the two sets of results is still meaningful. In fact,

**Figure 9: Histogram of number of alters per ego****Figure 10: Histogram of active networks size**

Each of these active networks can be partitioned into a set of layers (also called *circles*), grouping the tie strength between the ego and its alters. Offline social networks typically feature four layers [18], while five layers are a common finding for OSN [7]. The optimal number of layers can be obtained through cluster analysis. Specifically, here we use the Mean Shift algorithm [5]. As shown in Figure 11, similarly to other OSN, the optimal number of circles for the egos in our datasets is five.

When looking at the composition of circles (Figure 12), we obtain average values that are very close to the sizes of social groups (1.5, 5, 15, 50, and 150) predicted by Dunbar for offline social networks. While the layer sizes for OSN were found to be typically smaller than those in offline social networks [7], the journalists group seems instead to mimic very closely the offline dynamics. Specifically, the sizes of the two layers are slightly larger than in the model. Note, however, that this might depend on the specific clustering algorithm used. Similar deviations across clustering techniques have been observed also before [7]. On the other hand, the sizes of the two outermost layers are significantly larger for journalists than for generic users [7] and politicians [3]. Remarkably, the average size of the entire ego network is very close to the first empirical observation (~ 130) that validated the social brain hypothesis and provided evidence for the reference size of 150 alters [10].

filtering in [7] (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.

Figure 13 shows the scaling ratios between adjacent circles, confirming the remarkable match with the characteristic value of 3 commonly found in the related literature.

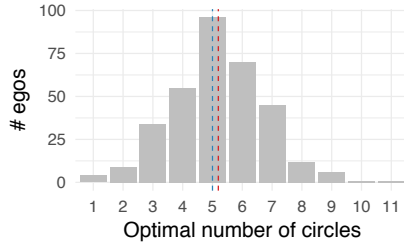


Figure 11: Optimal number of circles per ego (red line: average, blue line: median)

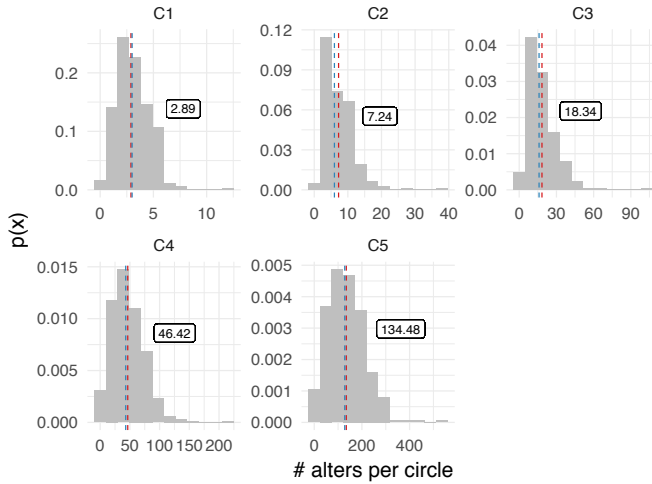


Figure 12: Distribution of circle sizes

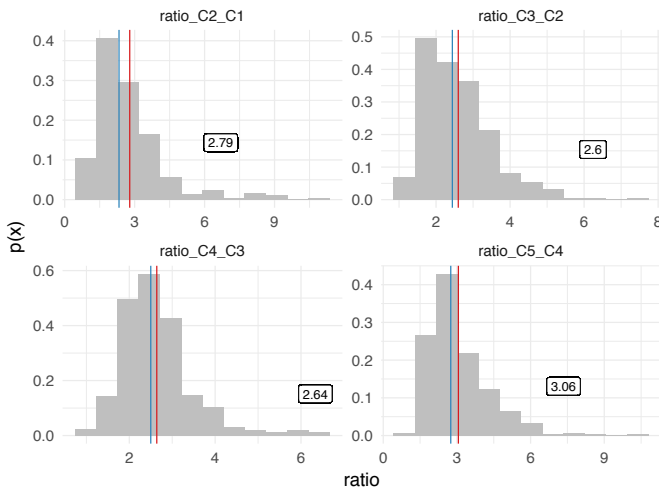


Figure 13: Distribution of scaling ratios

3.2 Dynamic ego networks

The dynamic analysis of ego networks focuses on their evolution over time. Specifically, snapshots of one year are considered⁷, 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.

The overlapping between the members of the same ring across different snapshots is measured using the Jaccard 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.

Figures 14 and 15 show the values of these indices in each of the five rings. They reveal very stable layers, in which the composition does not change much and in which jumps between rings are very few. This is in stark contrast with the results obtained for the politicians in [3]. In fact, politicians features low stability in all rings (but especially in the inner most ones), and their jump index is high. We conjecture that journalists may use Twitter less as a leverage to win people over and more as a social communication tool. Further analyses are required to substantiate this claim.

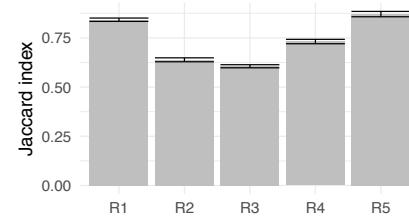


Figure 14: Average Jaccard index per ring

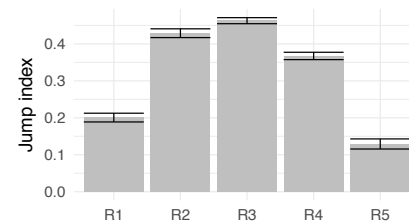


Figure 15: Average jump index per ring

3.3 Social tweets and hashtags

It is reasonable to expect the social tweets generated by journalists to be frequently topic-driven. In order to investigate whether this is the case, in this section we study the relation between alters and hashtags. In Figure 16, we show the histogram of the percentage of relationships activated by hashtag per ego. A relationship is considered activated by a hashtag if the first interaction between the ego and the alter contains a hashtag. The average in the journalists

⁷Hence, egos for which we observe less than two years of tweets are filtered out

dataset is 36%, while for generic Twitter users is around 6% and for politicians is around 15% [3]. Hence, the Twitter relationships of journalists are very much information-driven.

Figure 17 shows how the activation by hashtag impacts on the different rings. Ring R1 is the most affected, and this was also true for the politicians dataset studied in [3]. However, while the impact of activation-by-hashtag decreases progressively in the outer rings of politicians, it remains stable for journalists. This is again a confirmation that journalists engage on Twitter in an information-driven way, although, as shown by Figures 14 and 15, the set of alters with which they communicate is very stable, contrarily to politicians and generic users. Figure 18 shows that there is not a significant difference in contact frequency between alters whose relationship is activated by hashtags and those for which it is not. However, relations that are activated by hashtags tend to use hashtags significantly more often also in future interactions (Figure 19).

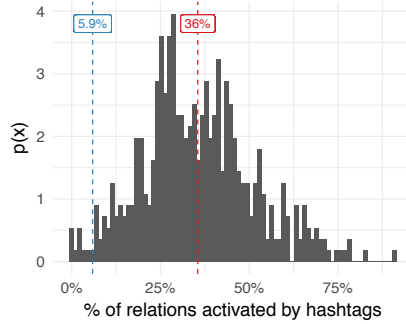


Figure 16: Histogram of the percentage of relations activated by hashtags (in red: mean, in blue: mean for generic Twitter users)

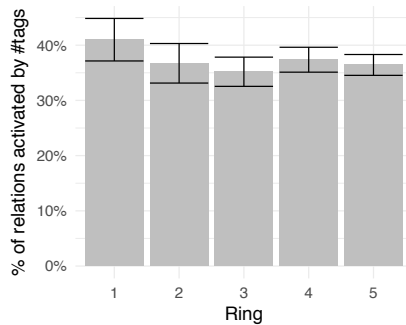


Figure 17: Average percentage of relations activated by hashtags per ring with confidence intervals

4 CONCLUSIONS

In this work we have studied the properties of ego networks for a set of the most popular Italian journalists. We have found that the structural patterns of their ego networks mirror those of offline social networks, even more closely than those of generic Twitter users. In particular, both the circle sizes and the circle scaling ratios are very close to the findings from the offline social network

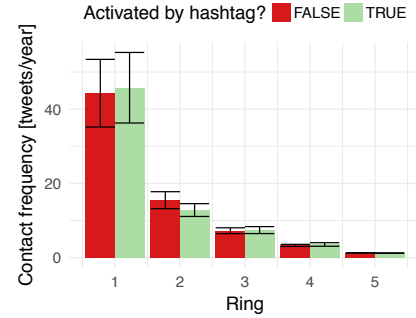


Figure 18: Contact frequency per ring

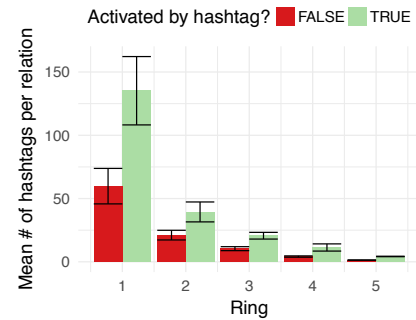


Figure 19: Mean number of hashtags per alter in ring

literature. Differently from what has been observed for politicians on Twitter, the composition of the circles is quite stable. Journalists stand out in terms of relationships activated by hashtags and number of hashtags per alter, both with respect to generic users and also with respect to politicians. This suggests that journalists' engagement on Twitter is mostly information-driven.

ACKNOWLEDGMENTS

This work was funded by the SoBigData project. The SoBigData project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024.

REFERENCES

- [1] Sinan Aral and Marshall Van Alstyne. 2011. The diversity-bandwidth trade-off. *Amer. J. Sociology* 117, 1 (2011), 90–171.
- [2] Valerio Arnaboldi, Marco Conti, Andrea Passarella, and Robin Dunbar. 2013. 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, 15–26.
- [3] Valerio Arnaboldi, Andrea Passarella, Marco Conti, and Robin Dunbar. 2017. Structure of Ego-Alter Relationships of Politicians in Twitter. *Journal of Computer-Mediated Communication* 22, 5 (2017), 231–247.
- [4] Lily Canter. 2015. Personalised Tweeting. *Digital Journalism* 3, 6 (2015), 888–907.
- [5] Dorin Comaniciu and Peter Meer. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence* 24, 5 (2002), 603–619.
- [6] R.I. Dunbar. 1998. The social brain hypothesis. *Evolutionary Anthropology* 9, 10 (1998), 178–190.
- [7] R.I.M. Dunbar, Valerio Arnaboldi, Marco Conti, and Andrea Passarella. 2015. The structure of online social networks mirrors those in the offline world. *Social Networks* 43 (2015), 39–47.

- [8] Martin Everett and Stephen P Borgatti. 2005. Ego network betweenness. *Social networks* 27, 1 (2005), 31–38.
- [9] Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. 2011. Modeling users' activity on twitter networks: Validation of dunbar's number. *PloS one* 6, 8 (2011), e22656.
- [10] Russell A Hill and Robin IM Dunbar. 2003. Social network size in humans. *Human nature* 14, 1 (2003), 53–72.
- [11] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 56–65.
- [12] Nan Lin, Karen S Cook, and Ronald S Burt. 2001. *Social capital: Theory and research*. Transaction Publishers.
- [13] Christopher McCarty. 2002. Structure in personal networks. *Journal of social structure* 3, 1 (2002), 20.
- [14] Daniele Quercia, Licia Capra, and Jon Crowcroft. 2012. The Social World of Twitter: Topics, Geography, and Emotions. *ICWSM* 12 (2012), 298–305.
- [15] Frank Michael Russell, Marina A Hendricks, Heesook Choi, and Elizabeth Conner Stephens. 2015. Who Sets the News Agenda on Twitter? Journalists' posts during the 2013 US government shutdown. *Digital Journalism* 3, 6 (2015), 925–943.
- [16] Alistair Sutcliffe, Robin Dunbar, Jens Binder, and Holly Arrow. 2012. Relationships and the social brain: integrating psychological and evolutionary perspectives. *British journal of psychology* 103, 2 (2012), 149–168.
- [17] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welp. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsn* 10, 1 (2010), 178–185.
- [18] W-X Zhou, D Sornette, R a Hill, and R I M Dunbar. 2005. Discrete hierarchical organization of social group sizes. *Proceedings. Biological sciences / The Royal Society* 272, 1561 (2005), 439–444.