Towards Understanding Political Interactions on Instagram

Martino Trevisan, Luca Vassio, Idilio Drago, Marco Mellia
Politecnico di Torino

ABSTRACT
Online Social Networks (OSNs) allow personalities and companies to communicate directly with the public, bypassing filters of traditional media. As people rely on OSNs to stay up-to-date, the political debate has moved online too. We present a preliminary study of interactions in a popular OSN, namely Instagram. We take Italy as a case study in the period before the 2019 European Elections. We observe the activity of top Italian Instagram profiles in different categories: politics, music, sport and show. We record their posts for more than two months, tracking “likes” and comments from users. Results suggest that profiles of politicians attract markedly different interactions than other categories. People tend to comment more, with longer comments, debating for longer time, with a large number of replies, most of which are not explicitly solicited. Moreover, comments tend to come from a small group of very active users. Finally, we witness substantial differences when comparing profiles of different parties.

KEYWORDS
Instagram; User behaviour; Online social networks; Politics

1 INTRODUCTION
Online Social Networks (OSNs) have become a space of paramount importance for the exchange of content and dissemination of information. As more and more people rely on OSNs to stay up-to-date, the political debate has naturally moved online too. Politicians and political associations communicate directly with the public via OSNs, bypassing filters of journalists and traditional media. While this lack of mediation has opened unprecedented forms of interaction, we witness the sudden explosion of harsh political debates and the dissemination of rumours in OSNs. Identifying such behaviour requires a deep understanding on how people interact via OSNs during political debates. We present a preliminary study of interactions in a popular OSN, namely Instagram. We take Italy as a case study in the period before the 2019 European Elections. We observe the activity of top Italian Instagram profiles in different categories: politics, music, sport and show. We record their posts for more than two months, tracking “likes” and comments from users. Results suggest that profiles of politicians attract markedly different interactions than other categories. People tend to comment more, with longer comments, debating for longer time, with a large number of replies, most of which are not explicitly solicited. Moreover, comments tend to come from a small group of very active users. Finally, we witness substantial differences when comparing profiles of different parties.

2 METHODOLOGY
This section describes how we collect Instagram data and provides some definitions of entities used to characterise interactions on this social network. 2.1 Data collection
We focus on data collected from Instagram public profiles. We use a custom crawler to download and store data and meta-data regarding the profiles and the corresponding posts and comments.

Our crawler collects the activity of the set of monitored profiles in real-time. Periodically, it downloads the meta-data of the profiles and all their new generated content, i.e., their posts. For these posts, our crawler downloads all the comments written by any user in

The research leading to these results has been funded by the foundation Compagnia di San Paolo through the 2019 internationalisation project program between Politecnico di Torino (Italy) and Universidade Federal de Minas Gerais (Brazil), by the Brazilian research agencies CNPq, CAPES and FAPEMIG and by the SmartData@Polito center.
the first 24 hours after the posting time. We remove all sensitive information from the data, e.g., any account identification of the users, and store the remaining information on a Hadoop-based cluster for further processing.

We are interested in the top public figures that publish information via Instagram. As such, we set an arbitrary threshold to include a profile in the crawling: Only profiles with at least 10,000 followers are considered. Since here we evaluate Italian profiles only, we further restrict the data capture to profiles whose posts are composed by at least 40% of Italian words.\(^1\)

The list of monitored profiles is not fixed and has grown during the crawling campaign. We search for new profiles using the hashtags present in posts, and whenever a profile explicitly mentions another one. We started the crawler on late December 2018, with a list of 50 popular Italian public figures. The profile list started growing very fast, and after two weeks it already included more than 10,000 entries. After four months of collection (April 2019), the crawler includes 19,156 Italian public profiles that generated 1,367,949 posts associated to 57,617,533 comments. To avoid possible artefacts due to the fact that some profiles have been added to the crawler after than others, for the analysis that follows we consider only profiles present in the crawler on February 1st, 2019. Moreover, we evaluate only posts and comments from February 1st to April 10th, 2019.

Instagram provides little information about a profile: a name, a biography and a profile picture. We thus resort to external sources to map each profile to a category and each politician’s profile to a political party. First, we use HypeAuditor,\(^2\) an online analytics platform, to get the list of top Italian influencers. HypeAuditor’s public list is restricted to the top-1000 profiles per country, from which we obtain 150 profiles belonging to the categories of our interest (i.e., sport, music and show) and passing the filters described above (e.g., posting mostly in Italian in the evaluated time period). Second, to find politicians, we search the set of Italian public profiles that generated 1,367,949 posts associated to 57,617,533 comments. To avoid possible artefacts due to the fact that some profiles have been added to the crawler after than others, for the analysis that follows we consider only profiles present in the crawler on February 1st, 2019. Moreover, we evaluate only posts and comments from February 1st to April 10th, 2019.

We use a Hunspell dictionary to filter Italian words. It is available at: https://cgit.freedesktop.org/libreoffice/dictionaries/tree/

We are interested in evaluating the effectiveness of explicitly calling someone else to the debate of a post. In this case, we consider that user A is calling user B to the debate. These two cases are illustrated in Figure 1 (i) and (ii).

Second, whenever we find a comment from user A mentioning user B, who has not commented on the post yet, we classify this first mention found in A’s comment as:\(^3\)

- *Answered first mention*: If B comments at least once on the post after A’s comment. Note that we consider the mention in A’s comment as answered if B comments afterwards, regardless of whether B’s comment contains other mentions;
- *Unanswered first mention*: If B never comments on the post after the mention of user A.

Here we are interested in evaluating the effectiveness of explicitly calling someone else to the debate of a post. In this case, we consider that user A is calling user B to the debate. These two cases are illustrated in Figure 1 (i) and (ii).

Second, whenever we find a comment of a user B mentioning user A, who has already commented on the post, we classify the mention found in B’s comment as:

- *Solicited reply*: If A has mentioned B before in the post;
- *Unsolicited reply*: if A has not mentioned B before in the post.

In these cases we evaluate whether people engage in conversations and whether people reply to comments without being invited to the debate. These cases are illustrated in Figure 1 (iii)–(v). Note that a mention classified as “solicited reply” may appear after another

\(^1\)Note that we may miscount some of these two cases due to border effects, since we stop monitoring the comments of a post after 24 hours.

---

**Table 1: Dataset overview. The last three rows are subclasses of politics.**

| Class       | Profiles | Posts   | Comments   | Commenters |
|-------------|----------|---------|------------|------------|
| Music       | 50       | 2,846   | 1,565,141  | 653,683    |
| Sport       | 68       | 4,393   | 1,773,835  | 724,959    |
| Show        | 32       | 2,338   | 1,183,989  | 565,126    |
| Politics    | 85       | 7,617   | 1,621,857  | 306,541    |
| Lega + FI + FdI | 32 | 4,107   | 1,225,597  | 234,745    |
| M5S         | 33       | 2,037   | 276,593    | 70,075     |
| PD          | 20       | 1,473   | 119,667    | 33,647     |

**Figure 1: Examples of mentions: Users’ comments are represented as circles, mentions as arrows. Time evolves top-down. Colours represent the way we categorise mentions.**

We classify the mentions by evaluating the comments of a post chronologically, from the oldest to the newest comment, and checking the mentioning/mentioned users. We classify all the mentions into four categories which are exemplified with colours in Figure 1. First, whenever we encounter a comment from user A mentioning user B, who has not commented on the post yet, we classify this first mention found in A’s comment as:\(^5\)

- *Answered first mention*: If B comments at least once on the post after A’s comment. Note that we consider the mention in A’s comment as answered if B comments afterwards, regardless of whether B’s comment contains other mentions;
- *Unanswered first mention*: If B never comments on the post after the mention of user A.

---

\(^2\)https://hypeauditor.com/

\(^3\)We use a Hunspell dictionary to filter Italian words. It is available at: https://cgit.freedesktop.org/libreoffice/dictionaries/tree/

\(^4\)https://en.wikipedia.org/wiki/List_of_political_parties_in_Italy

\(^5\)Nevertheless, these three parties run autonomously for the 2019 European Election.
We first quantify the number of comments received by influencers, relatively larger number of comments per follower. Politics instead

![CDF of number of comments per post normalised by 1 000 followers.](image1)

shows a remarkable difference: the normalised median number of comments per post is 3.2 times larger than for sport with 8% of posts collecting more than 4 comments per 1 000 follower. Figure 2b breaks down the analysis for political parties. Posts of PD receive more comments: its median is 1.1 comments every 1 000 followers while for other parties it is less than 0.7.

We next compare the number of comments with respect to the number of likes per post. Figure 3 shows the CDFs of the ratio between the number of comments and the number of likes. We observe that posts tend to collect many more likes than comments. This low ratio of comments per like is expected, and reflects the fact that most Instagram users limit their interactions to a like, with only few users actively commenting on posts. Yet, for politics, comments are sensibly more frequent than for other categories. Considering political parties, we observe that posts from centre-right parties are sensibly more frequent than for other categories. Considering political parties, we observe that posts from centre-right parties have a smaller ratio of comments per like when compared to other groups.

Let us focus on how concentrated the comments are with respect to the population of commenters. We use the Lorenz curve for this, depicted in Figure 4. Each curve shows the fraction of comments written by the bottom x% of the commenters (sorted by this quantity). The more the Lorenz curve leans towards the right bottom of the plot, the more inequalities are present. Consider the area between the line of perfect equality (the main diagonal) and the Lorenz curve. Consider then the area between the line of perfect equality and the line of perfect inequality (the x axis). The ratio between these two areas is the Gini Index (GI). The closer to 1 the index is, the more unequal the distribution is.

Results are striking: the inequality for commenters of politics is much larger than the ones in other categories (Gini Index is
3.2 Comment length and timing

We now analyse the length of comments and show in Figure 5a the CDFs of the number of characters per comment. Most comments are quite short, i.e., 75% are shorter than 50 characters for music, show and sport. However, politics collects much longer comments, with the 75-percentile reaching 83 characters, and the 90-percentile 4 times larger than for other categories. Indeed, only 7% of them are composed by a single emoji, while this happens in 8-11% of cases for other categories. Figure 5b shows a breakdown by political faction. We notice that posts of conservative parties tend to collect shorter comments than PD and M5S.

Let us now focus on the longevity of the interactions with a post. We measure it analysing the time passed from the moment a post is published to the time users post comments on it. This means that users reply to comments of previous users, without being explicitly consulted. Analysing in depth the content of such replies as well as the users generating them is left for future work. Overall, when we combine both results, it appears that political comments do not try to drag other users into the discussion (Figure 7), but instead they tend to reply to prior comments (Figure 8).

3.3 Mentions in comments

To further examine how users’ interactions are generated, we study mentions between them. Intuitively, mentions can be an effective way of raising the number of comments in a post.

In Figure 7 we report the number of first mentions per 1 000 comments. Recall that a first mention is a mention toward a user that have not yet commented on this post. We distinguish between first mentions that are answered (red bars) and those that are left unanswered (blue bars). Once again, politics show a remarkably different behaviour than other categories: Every 1 000 comments, there are only 24 first mentions, and only 5 (20%) get answered. For comparison, for show category, there are 188 mentions in 1 000 comments, from which about 43% get answered. In a sense, this could point out that most initial mentions in politics posts are not focused on attract new users to the discussion.

The behaviour of users is even more interesting when we look at replies. Recall that a reply is a comment that mentions a user that has also previously commented. Replies help us to understand how users commenting on a post interact with each other. Figure 8 shows the number of replies every 1 000 comments, for solicited (red bars) and unsolicited (blue bars) replies. Here, politics exhibits the largest number of replies 205, most of which are unsolicited. This means that users reply to comments of previous users, without being explicitly consulted. Analysing in depth the content of such replies as well as the users generating them is left for future work.

Overall, when we combine both results, it appears that political comments do not try to drag other users into the discussion (Figure 7), but instead they tend to reply to prior comments (Figure 8).

---

6A comment may contain more than one mention. Hence, the maximum number of comments including first mentions is 24 every 1 000 comments.
This may indicate that users start discussions after reading the opinion of others, possibly even without knowing each other. If we put these numbers in perspective with the fact that comments in politics are more concentrated around a small fraction of very active users (Section 3.1), these results may also be interpreted as an indication of the existence of a passionate group of users, actively reacting to comments, e.g., to influence the political debate.

4 CONCLUSION

We analysed how users interact with politicians and other influencers on Instagram. Based on a large dataset including hundreds of Italian public profiles and millions of comments, we find notable differences across categories. Comments to politicians are more frequent and longer in comparison to other categories. Political commenters tend to discuss more, but are less likely to drag users not yet engaged in the discussion. This work aims at fostering further researches in this field. We are currently monitoring the evolution in time of the interactions before and after the 2019 European Parliament election.7

REFERENCES

[1] Eddy Borges-Rey. 2015. News Images on Instagram: The paradox of authenticity in hyperreal photo reportage. *Digital Journalism* 3, 4 (2015), 571–593.

[2] Martin Gibbs, James Meese, Michael Arnold, Bjorn Nansen, and Marcus Carter. 2015. # Funeral and Instagram: death, social media, and platform vernacular. *Information, Communication & Society* 18, 3 (2015), 255–268.

[3] Tim Highfield and Tama Leaver. 2015. A methodology for mapping Instagram hashtags. *First Monday* 20, 1 (2015), 1–11.

[4] Philip N Howard and Bence Kollanyi. 2016. Bots,# StrongerIn, and# Brexit: computational propaganda during the UK-EU referendum. *Available at SSRN 2798311* (2016).

[5] Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati. 2014. What we instgram: A first analysis of instagram photo content and user types. In *Eighth International AAAI conference on weblogs and social media*.

[6] Jamie Mahoney, Tom Feltwell, Obinna Ajuruichi, and Shaun Lawson. 2016. Constructing the visual online political self: an analysis of Instagram Use by the Scottish Electorate. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3339–3351.

[7] Andrew G Reece and Christopher M Danforth. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 6, 1 (2017), 15.

[8] Gustavo Resende, Philipe Melo, Hugo Sousa, Johnnatan Messias, Marisa Vasconcelos, Jussara Almeida, and Fabricio Benevenuto. 2019. (Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures. In *To appear in Proceedings of the Web Conference 2019*.

[9] Jessica L Ridgway and Russell B Clayton. 2016. Instagram unfiltered: Exploring associations of body image satisfaction, Instagram# selfie posting, and negative romantic relationship outcomes. *Cyberpsychology, Behavior, and Social Networking* 19, 1 (2016), 2–7.

[10] Pavica Sheldon and Katherine Bryant. 2016. Instagram: Motives for its use and relationship to narcissism and contestual age. *Computers in Human Behavior* 58 (2016), 89–97.

[11] Martino Trevisan, Danilo Giordano, Idilio Drago, Marco Mellia, and Maurizio Munafò. 2018. Five Years at the Edge: Watching Internet from the ISP Network. In *Proceedings of the 14th International Conference on Emerging Networking Experiments and Technologies (CoNEXT ’18)*. 1–12.