2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

# Characterization of Cross-posting Activity for Professional Users Across Major OSNs

Reza Farahbakhsh*, Ángel Cuevas*† and Noël Crespi*
*Institut Mines-Télécom, Télécom SudParis, {reza.farahbakhsh, noel.crespi}@it-sudparis.eu
†Universidad Carlos III de Madrid, acrumin@it.uc3m.es

***Abstract*—Online Social Networks (OSNs) are being intensively used by professional users (e.g., companies, politician, athletes, celebrities, etc) in order to interact with a huge amount of regular OSN users with different purposes (marketing campaigns, customer feedback, public reputation, etc). Hence, due to the large catalog of existing OSNs, professional users usually count with OSN accounts in different systems. In this context an interesting question is whether professional users publish the same information across their OSN accounts, or actually they use different OSNs in a different manner. We define as cross-posting activity the action of publishing the same information in two or more OSNs. In this paper we aim at characterizing the cross-posting activity of professional OSN users across three major OSNs, Facebook, Twitter and Google+. To achieve this goal we perform a large-scale measurement-based analysis across more than 2M posts collected from 616 professional users with active accounts in the three referred OSNs.**

## I. Introduction

Online Social Networks (OSNs) have become one of the most popular services in the Internet attracting billions of subscribers and millions of daily active users. We can find three dominant OSNs according to their number of subscribers: Facebook (FB), Twitter (TW) and Google+ (G+). While these systems have been demonstrated to be very attractive to regular users that perform a wide variety of social interactions on them, they also present a golden opportunity to professional players (i.e. brands, politicians, celebrities, etc.) to interact with a huge amount of potential customers/voters/fans to increase their reputation and popularity, to run marketing campaigns, to attract voters, etc.

Most professional users do not limit their activity to a single OSN, but usually they have accounts in multiple OSNs, including the most popular ones such as FB, TW and G+. Then an interesting question is whether professional players use all OSNs in the same way, or actually they use each OSN for different purposes. In other words, when a professional user wants to advertise or notify some update, does she publish that information in several OSNs?, or contrary, she publishes it in a single OSN depending on the type of information (e.g., if it is a personal update she publishes a post in one OSN, but in case it is a commercial update she selects another OSN). We refer to the information that a professional player publishes in multiple OSNs as *cross-posting activity*. Therefore, if a professional user publishes a post in FB and a post TW that contain the same information we consider them as a cross-post.

To the best of our knowledge, although there are other works that have analyzed the behaviour of regular users across two OSNs [1], [2], this paper presents the first large scale study on cross-posting activity of professional users across the three major OSNs, i.e., FB, TW and G+. We analyze the activity of 616 (popular) professional users with active accounts in the three referred OSNs. Among these users we can find big companies, politicians, athletes, artists, celebrities, etc. To perform the study we have analyzed more than 2M posts distributed across the 616 users in TW, FB and G+.

The first contribution of this paper is a simple yet efficient methodology that is able to precisely determine whether two posts contain the same information, and thus classify them as a cross-post. The validation of our methodology shows an accuracy of 99% for the classification of cross-posts. Based on this methodology, the first goal of the paper is to characterize the cross-posting activity of professional OSN users across FB, TW and G+. In order to achieve this objective we perform a data analysis that allows us to shed light to three key aspects of the cross-posting activity. $(i)$ An immediate question is whether the cross-posting phenomenon actually exists, and if it exists what fraction of the activities from a professional user is associated to cross-posting. $(ii)$ In case the cross-posting activity is relevant, we aim at understanding between which OSNs it is more frequent. This means, can we find more cross-posts between FB-TW, FB-G+, or TW-G+? $(iii)$ Finally, we measure what is the benefit, if any, that professional users obtain from the cross posting activity in terms of engagement.

Following, we list the main findings of our research:
(1) Cross-posting is a frequent practice across professional users. In median a professional user shares in other OSN 25% of the posts published in FB and G+, and only 3% of the tweets. However, we must note that professional users are much more active in TW than FB and G+, hence, in absolute terms, the TW account of users generate a larger volume of cross-posts than G+ accounts and similar volume to FB.
(2) The cross-posting phenomenon mainly happens between FB and TW, but it is also relevant between FB and G+. However, it is surprising that is more likely to find a cross-post published in FB, TW and G+, than only in TW and G+.
(3) Professional users obtain a substantial benefit in FB and TW by cross-posting since they attract 30% and 100% more engagement as compared to non-cross-post. However, in the case of G+ non-cross-posts attract 2× more engagement than cross-posts.
(4) Among the 616 analyzed users 50% prefer FB as most frequent option to initially upload their cross-posts, 45% prefer

ASONAM '15, August 25-28, 2015, Paris, France
 ISBN 978-1-4503-3854-7/15/08 $15.00
DOI: http://dx.doi.org/10.1145/2808797.2809382

TABLE I
DATASET DESCRIPTION

| OSN | total posts | avg. posts per user |
|---|---|---|
| FB | 422 K | 685 |
| G+ | 173 K | 280 |
| TW | 1.64 M | 2664 |

TW, and only 5% give priority to G+.
(5) Professional users with a strong preference for TW publish cross-posts that: $(i)$ are very similar across the different OSNs, $(ii)$ mostly includes textual content, and $(iii)$ mostly include links to websites different than OSNs sites.
(6) Professional users with a strong preference for FB publish cross-posts that: $(i)$ mostly includes audiovisual content, and $(ii)$ mostly include links to content stored in major OSNs sites.

## II. METHODOLOGY

This section briefly explains the implemented methodologies of the data collection and cross-posts identification.

### A. Data Collection

Our first challenge was to identify a numerous group of relevant professional users having active and popular accounts across FB, TW and G+. To this end, we rely on a large dataset that includes thousands of professional and regular users with an account in the three OSNs collected for a previous work [3]. From these users we were interested in those ones that meet two requirements: (i) have an active account in FB, TW and G+; (ii) present a high popularity in at least two of the systems. We found 616 professional users that satisfy the popularity requirement. We validated that the selected users were actually relevant in all the three OSNs by means of an external source [4] ranks professional users in each system in terms of popularity. Next we implemented three separate crawlers for FB, G+ and TW which receive a user ID (or username) as input and uses the API of each social network to collect the posts published by the user in her accounts. For more details on these implemented crawlers we refer the reader to our previous work [3], [5].

Table I summarizes the dataset used in this paper. In total, we analyze more than 2M posts published across 616 professional publishers in FB, TW and G+. It must be noted that the collection campaign finished on May 2013, thus our dataset may not include novel features released by any of the analyzed OSNs after that period.

### B. Methodology to Identify Cross-posts

In order to being able to compare cross-posting activity of professional users we need to have an accurate mechanism that detects when two posts are actually containing the same information. Hence, we have implemented a hierarchical classification algorithm that determines whether two posts can be considered as cross-posts in two steps. Then, given the description (i.e. the text associated to a post) of two posts, $P_1$ retrieved from the account of user $U$ in $OSN_A$ and $P_2$ published by $U$ in her account of $OSN_B$, our algorithm proceeds as follows:

(1) it compares $P_1$ and $P_2$ using NTLK Fuzzy Match [6] that provides a binary decision based on the similarity of the compared texts. NTLK Fuzzy Match generates a positive answer (i.e., same text) when both texts are very similar and only differ in some few characters. In summary, all the pairs of posts receiving a positive classification are labelled as cross-posts while the remaining pairs need to go through the second step of our algorithm.
(2) We compare $P_1$ and $P_2$ using two similarity metrics: cosine similarity [7] and string similarity [8]. These two metrics provide as output a value ranging between 0 and 1, so that the closer is the output to 1 the more similar $P_1$ and $P_2$ are. Based on the obtained results, we classify $P_1$ and $P_2$ as cross-post if both metrics, cosine similarity and string similarity, are $\geq 0.5$. Later in this section we validate our methodology and demonstrate why we have selected the 0.5 threshold.

The previous algorithm serves to classify any pair of posts as cross or non-cross. In addition, we must note that our algorithm is not bound to any particular alphabet, so it can be applied in multiple languages. However, the use of the hierarchical algorithm is not enough for the purpose of this research. Following we describe two more elements we had to integrate in our methodology to ensure the accuracy of the results obtained in the paper.

First, we had to define which pairs of posts should be compared together. In order to be accurate and efficient and skip repetitive patterns by users over time, we applied the following methodology. Given a post $P_{FB}$ published by a user $U$ in her FB account at the timestamp $t_{FB}$, we compare $P_{FB}$ with all the posts that user $U$ published in her TW and G+ accounts in a time window starting one week before and finishing one week after $t_{FB}$. In other words, we compare each post in a time window of two weeks around the date that post was published.

Second, TW API limits the number of retrieved posts for any user to the last 3,200 posts she published, while FB and G+ do not have that limitation and provide all the posts published by the user since she registered in the system. Hence, in order to perform an accurate study, we have restricted our cross-post analysis to the time window imposed by the limitation of TW API for each user in our dataset. It must be noted that the number of posts depicted in table I already consider this limitation.

*We applied the described methodology to the selected 616 OSN professional users and more than 2M posts and we found 176K cross-posts across their OSNs accounts.*

### C. Methodology Validation

In order to ensure the accuracy of the proposed methodology, 3 persons manually classified 12.8K random posts as cross-posts or non-cross-posts. In order to have a meaningful validation set we ensured that half of the posts had been labelled as cross-post and half as non-cross-posts by our classification tool. Then, given two posts published by a user in two different OSNs we classify them as a cross-post if at least 2 out of the 3 persons performing the manual inspection

TABLE II
METHODOLOGY VALIDATION, FALSE POSITIVE (FP) AND FALSE NEGATIVE (FN) RATES OF DIFFERENT SIMILARITY THRESHOLD (ST) IN OUR CROSS-POSTING IDENTIFICATION METHODOLOGY.

| ST>0.3 similarity | | ST>0.5 similarity | | ST>0.7 similarity | |
|---|---|---|---|---|---|
| FP | FN | FP | FN | FP | FN |
| 15.006 | 0.194 | 0.140 | 1.117 | 0.016 | 4.593 |

indicate that both posts contain the same information. This allows us to obtain a ground truth set to determine the false positive and false negative rate of our methodology. A false positive occurs when our tool classifies as cross-post two posts (published by the same user in two different OSNs) that are actually referring to a different piece of information. A false negative happens when our tool classifies as non-cross-post two posts that actually contains the same information.

Based on the ground truth set we compute the false negative and false positive rate for our methodology using three different thresholds for the second step of the algorithm: 0.3, 0.5 and 0.7. Basically, a lower threshold requires less similarity between the compared posts to classify them as cross-post. Table II shows the false positive and false negative rate for our algorithm when it uses each of the evaluated thresholds. The results clearly determine that 0.5 is a very good threshold since it presents a very low rate for false positives (0.14%) and false negatives (1.11%).

## III. CROSS-POSTING CHARACTERIZATION

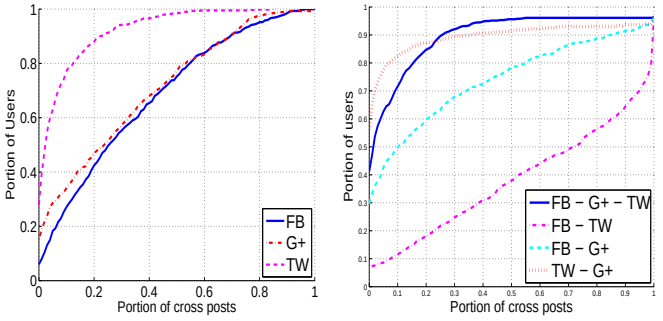
The first question we aim to answer in this section is whether the cross-posting phenomenon exists in the activity of professional users, and what is its weight in FB, TW and G+. To this end, we quantify what is the volume of cross-posting happening between FB-G+, FB-TW, TW-G+ and FB-TW-G+, in order to determine what pair of OSNs is actually sharing more common information. Finally, we also want to characterize the impact of cross-posting in the attracted engagement measured in terms of likes comments, and shares.

### A. Quantification of Cross-Posting Activity

The goal is to quantify the cross-posting phenomenon for professional users in FB, TW and G+. Towards this end, we compute for each user and each OSN the portion of cross posts with respect to all the posts each user has published. For instance, given a user $U$ and her FB account we compute how many posts published in that account also appear in TW, G+ or both. We quantify the same parameter for the TW and G+ accounts of user $U$[1].

Figure 1(a) shows the CDF for the portion of cross posts across the 616 users analyzed in the three OSNs. The x axis refers to the portion of posts and the y axis to the portion of users. For instance, the point {x=0.2, y=0.4} in the line associated to FB indicates that 40% of the users have $\leq$ 20% of cross-posts in their FB accounts. The first immediate conclusion extracted from the graph is that most of the professional users have published some cross-post. In

[1]It must be noted that for this analysis we do not take into account where the post appears first, but only consider whether it is unique in an OSN or it appears in 2 or 3 of them.




(a) (b)

Fig. 1. (a) CDF for the portion of cross-posts per user in FB, G+ and TW. (b) CDF for the portion of cross-posts and in each possible cross-posting pattern (FB-TW, FB-G+, TW-G+ or FB-G+-TW).

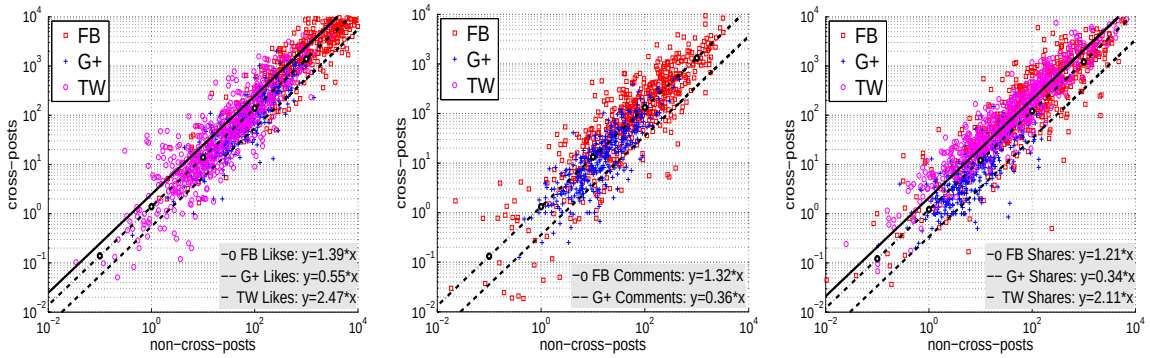
particular, when we consider FB accounts we find that only 6% of the users do not have any cross-post, which means for those users the information published in FB cannot be found neither in TW nor in G+. This number grows up to 15% and 28% for G+ and TW, respectively. Therefore, a vast majority of professional users published some cross-post at some point. Hence, the first conclusion is that in general professional users find some value on the cross-posting activity.

If we compare the results obtained for the three OSNs, we clearly observe that, in relative terms, the cross-posting activity is more frequent for those posts published in FB and G+ than in TW. The results for TW show that most of the tweets are not replicated neither in FB nor in G+. The median value, which shows the typical portion of cross-posts for a user in each OSN, shows that for a typical professional user around 1/4 of the posts that appear in FB and 1/4 of the posts that appear in G+ are also available in at least one more OSN. However, in the case of TW, out of 100 tweets only 3 of them are replicated in other OSNs. Finally, we can find quite a lot professional users with an intensive cross posting activity. In particular, 25%, 23% and 1.5% of the analyzed users, in FB, G+ and TW, respectively, have published more cross posts (i.e., $\geq$ 50%) than posts appearing exclusively in a single OSN. We refer to these posts as *non-cross-posts*.

The previous analysis refers to the cross-posting activity in relative terms. However, it is important to notice that, according to the overall activity of the professional users in our dataset, the publishing rate of professional users in TW is 4$\times$ higher than in FB and G+.

### B. Inter-OSN Cross-Posting

Once we have demonstrated that cross-posting is a common practice among professional users in FB, TW and G+, we analyze how cross-posting happens among them. Then, our goal is to find whether professional users prefer to share things in FB and TW, or rather it is more frequent finding common posts in FB and G+, or actually there are lots of cross-posts published in TW and G+. In order to perform this analysis we proceed as follows. For a given user $U$ we get all her cross-posts in FB (independently whether the first appearance happened in that OSN or another one) and compute which




(a) Likes (b) Comments (c) Shares

Fig. 2. Users' average attracted engagement per post, for cross posts initiated in each OSNs vs. non-cross posts.

portion of them also appears in TW, which portion in G+ and which portion in both TW and G+. We repeat the same process for the TW and G+ accounts of user $U$. Therefore, for each user we know what is the cross-posting level for the following relations: FB-TW, FB-G+, TW-G+ and FB-TW-G+.

Figure 1(b) shows the CDF for the portion of cross-posts that occurs for the four referred relations across the users in our dataset. Again in this figure the x axis refers to portion of posts and the y axis shows the portion of users. Then for instance the point x=0.4, y=0.3 in the FB-TW line indicates that 30% of the users publish $\leq$40% of their cross posts in FB and TW. The results in the figure demonstrate that professional users perform much more cross-posting between FB and TW than in any other combination of OSNs. This claim is supported by the fact that in median a professional user publishes 70% of their cross-posts in FB and TW. In addition, we can only find 8% of the users that never shared a post between their FB and TW accounts, while this value grows to 30% between FB and G+, to 40% for the case in which the three OSNs are involved, and to 55% for TW and G+. Therefore, this last result surprisingly states that is more likely that a user publishes the same posts in the three OSNs than only in TW and G+.

### C. Engagement Analysis

A plausible reason of why professional OSN users publish the same information across different OSNs is to try to increase the coverage in order to engage as many end-user as possible within their accounts. Therefore, in this subsection we want to conclude whether cross-posts achieve more engagement than non-cross-posts in FB, TW and G+. In order to measure the engagement we use standard reaction mechanisms available for end users in OSNs: likes, comments and shares[2]. It should be acknowledged that our TW crawler could not retrieve comments.

In order to measure the efficiency of cross-posts to attract engagement in one OSN we measure, for a given user $U$, the average engagement for $U's$ non-cross-posts and $U's$ cross-posts initiated in that OSN in terms of likes, comments and shares. Figure 2 shows a scatter plot for FB, G+ and TW for each of the engagement type: likes (Figure 2(a)), comments (Figure 2(b)) and shares (Figure 2(c)). Each point in the graphs represents a user with an x coordinate referring to the average engagement for non-cross-posts and y coordinate referring to the average engagement for cross-posts initiated by that user in that OSN. In addition, all the figures include three lines (one per OSN) showing the linear regression for the cloud of points represented by an equation[3] of type $y = ax$. When the slope of the linear regression, represented by the value of $a$, is greater than 1, it means that for that OSN cross-posting is worthy since cross-posts attract more engagement than non-cross-posts in average.

The results demonstrate that cross-posts in FB and TW allows professional users to attract more engagement than non-cross-posts. However in the case of G+, cross-posts receive considerably less attention than non-cross-posts. In more detail, a FB user attracts 39% more likes, 32% more comments and 21% more shares in FB when she uses cross-posts instead of non-cross-posts. In the case of TW cross-posting provides even more benefit. This is, a cross-post initiated in the TW of a professional user attracts 2.47$\times$ and 2.1$\times$ more likes (i.e., favourites) and shares (i.e., retweets) than a non-cross-posts. Finally, in the case of G+ a cross-post roughly achieves 1/2 of the likes (i.e, +1's), 1/3 of the comments and 1/3 of the shares compared to non-cross-posts. Therefore, cross-posting seems to be a bad strategy if the goal of a professional user is to attract as many reactions as possible in G+.

*In summary, cross-posting exists and it is a frequent practice across professional users in FB, TW and G+. It mostly happens between the FB and TW accounts of professional users, and it very rarely occurs between TW and G+.* Finally, in terms of attracted engagement, cross-posting is beneficial in FB and TW, but not in G+.

## IV. Preference of Professional Publishers

In this section we tackle two interesting questions. First, we want to know in overall which OSN is used more frequently as first option to publish fresh information that later will be republished in other OSNs. Second, we want to understand what is the OSN that professional users prefer to publish first

[2]This is the nomenclature employed in FB. A like is associated to a +1 in G+ and to a favourite in TW. A share is associated to reshare in G+ and a retweet in TW.

[3]Usually a linear regression is represented as y=ax+b, but in the figure we just use y=ax, since we are interested in the slope, but not in the offset.

TABLE III
PORTION OF CROSS-POSTS PUBLISHED FOR FIRST TIME IN FB, TW OR G+ FOR DIFFERENT CROSS-POSTING PATTERNS. THE TABLE ALSO INCLUDES THE PORTION OF POSTS THAT ARE PUBLISHED IN AT LEAST TWO OSNS AT THE SAME TIME (I.E., EXACT TIMESTAMPS)

| Patterns | #Posts | %Posts | %FB(1st) | %G+(1st) | %TW(1st) | %Exact-T |
|---|---|---|---|---|---|---|
| FB-TW-G+ | 18619 | 10.56 | 34.93 | 12.32 | 49.80 | 2.95 |
| FB-G+ | 34337 | 19.48 | 73.68 | 24.07 | - | 2.26 |
| FB-TW | 117276 | 66.52 | 36.28 | - | 56.80 | 6.92 |
| G+-TW | 6073 | 3.44 | - | 23.79 | 75.96 | 0.25 |

TABLE IV
CROSS-POSTS INITIATED IN GB, G+, TW

| OSN | #Posts | %Posts |
|---|---|---|
| FB | 74355 | 42.17 |
| G+ | 12002 | 6.81 |
| TW | 80497 | 45.66 |
| other | 9451 | 5.36 |

TABLE V
PREFERRED OSN PER USER

| OSN | #Users | %Users |
|---|---|---|
| FB | 307 | 50 |
| G+ | 30 | 5 |
| TW | 275 | 45 |

TABLE VI
USERS CLASSIFICATION BASED ON OSN PREFERENCE CRITERIA

| Criteria | #User | #FB | %FB | #G+ | %G+ | #TW | %TW |
|---|---|---|---|---|---|---|---|
| 100% | 32 | 11 | 1.79 | 2 | 0.32 | 19 | 3.08 |
| $\geq$80% | 182 | 75 | 12.18 | 5 | 0.81 | 102 | 16.56 |
| $<$50% | 95 | - | - | - | - | - | - |

the information. Answering the first question will determine which OSN is used more times as source of cross-OSN information, while the response to the second question will roughly determine what is the OSN that professional users value more to publish first their fresh updates.

### A. OSN-based Analysis

Table IV shows the number and portion of cross-posts in our dataset that were initiated in FB, TW and G+. The results demonstrate that TW appears as initial source of information for 45% of the cross-posts closely followed by FB with 42%, while G+ is rarely chosen as first option. Finally, we find a very interesting result associated to the category *"other"* that represents those cross-posts that could not be assigned to a particular OSN since they were published exactly at the same time (i.e., same timestamp) in at least two OSNs. It is surprising that almost 10K cross-posts, which represent 5.3% of all the cross-posts in our dataset, experienced this parallel publication. This reflects the use of automatic publishing tools that upload in parallel some information to two or more OSNs.

As we determined in the previous section, most of the posts are not published in all the three OSNs, but just two of them. Therefore, it is interesting to analyze for each particular publishing pattern which OSN appears more frequently as initial source of information. Table III shows the results for all the possible cross-post patterns: FB-TW-G+, FB-TW, FB-G+ and TW-G+. First of all, the results confirm the conclusion obtained in the previous section since 2/3 of the cross-posts appear exclusively in FB and TW, 1/5 belong to the category FB-G+, and as we already stated it is more likely finding cross-posts across the three OSNs (10%) than only across G+ and TW (3.4%). In the most popular category, i.e., FB-TW, TW appears as first option for 57% of the posts while FB is chosen in first place only 36% of the times. When G+ competes individually either with FB or TW, it is source of information only 1/4 of the times. For those posts published in the three OSNs, 1/2 of them appear first in TW, 1/3 in FB and 1/10 in G+. Finally, we want to highlight that all the categories include some portion of posts that where published in parallel at the same exact time in two OSNs. This phenomenon is especially relevant for cross-posts between FB-TW.

In summary, the OSN-based analysis demonstrates that Twitter is the OSN selected as initial source of information more frequently. FB appears as the second option close to Twitter. Finally, G+ is the least preferred option.

### B. User-based Analysis

The OSN-based analysis revealed that Twitter is chosen as first option for a larger number of cross-posts. However, we cannot extract from that analysis that TW is the preferred OSN for most of the users, since it may happen that very active users contributing a large number of posts prefer TW but less active users prefer FB or G+. Therefore, in this section we analyze which is the preferred OSN for professional users. For a given user its preferred OSN is the one she selected in first place for a major number of posts. Table V shows the number and portion of users in our dataset that prefer each OSN. The results reveal that half of the professional users prefer FB, closely followed by 45% of the users that prefer TW, while only 5% of the users chooses G+ as initial OSN for publishing their post. Therefore, FB and TW has exchanged their positions as compared to the OSN-based results. As we indicated above, the difference between the post-based and user-based results comes from the fact that users tend to be more active in TW.

Once we have classified professional users' preference, a subsequent question is, can we find users that shows a strong preference for a particular OSN? In other words, are there users that utilize as source of information one single OSN for most of their cross-posts?

Table VI shows the number and portion of professional users in our dataset that choose either FB, TW or G+ to initiate 100% or 80% of their cross-posts showing a clear strong preference. In addition, we also quantify the number and portion of users that publish in first place less than 50% of their posts in all three OSNs and thus do not show any strong preference. We can find 19, 11 and 2 users that always choose TW, FB and G+ as initial source for their cross-posting activity, respectively. If we move down the threshold to 80% the number of users showing a clear evidence of which OSN they prefer grows a lot for FB and TW, but not for G+ that only accounts for 5 users. There are 75 (12.18%) users with a preference for FB and 102 (16.56%) with a noticeable preference for TW. In contrast to these users showing a clear

OSN preference, we can find 95 (15.4%) users that are not biased towards any OSN, even though they make use of cross-posts.

*In summary, professional users are (more or less) equally divided into those that prefer TW and those that prefer FB, and very few cases that show a preference for G+.*

## V. Related Work

There exist several works that have studied the graph and connectivity properties of Facebook, [9], [10], Twitter [11], [12], and Google+ [13], [14]. In addition, there are other works in the literature that compare two or more OSNs based on their graph properties [15], [16]. However, these works do not consider the same users in the different OSNs for their analysis since their goal is to characterize OSNs at a macroscopic level. There are only few works that try to characterize the behaviour of the same user or group of users across different OSNs. The main reason is that it is not an easy task to identify and collect the information of the same users across different system and, in addition, it requires to have one data collection tool for each system. Authors in [2] compare 195 users from the archival community and study their activity pattern in TW and FB. This is a small-scale study based on 2,926 links to external documents. In [17], we find again a comparative analysis for users having accounts in FB and TW. This work studies the behaviour of 300 users from a psychological perspective and the results reveal a correlation between end-users personality and their use of FB and TW. Finally, the most similar work to our paper is a very recent study [1] that compares the behaviour of 30,000 regular users across TW and Pinterest. Although this study similar in spirit to our work, we differ from [1] since we are focusing in professional OSN players instead of regular users, and we are comparing TW, FB and G+ instead of TW and Pinterest.

## VI. Conclusions

This paper presents the first large-scale measurement-based characterization of the cross-posting activity for OSN professional users across FB, TW and G+. We have used a simple yet efficient methodology that is able to determine with an accuracy of 99% whether two posts, even from different OSNs, contains the same information, and if so classify them as cross-post. We have used that methodology to classify more than 2M posts published for 616 professional publishers with active accounts in FB, TW and G+. Following we list the main outcomes of the paper.

First, we have demonstrated that professional users frequently publish the same information in at least two OSNs, especially in the case of FB and G+. Although professional users in TW present a low portion of cross-posts, the fact that they are very active implies that in absolute terms we can find quite a lot cross-posts in their TW accounts. Second, a professional user publishes (in median) 70% of her cross-posts exclusively in FB and TW, and around 15% in FB and G+. Furthermore, we demonstrated that the cross-posting activity between TW and G+ is negligible. Third, professional users benefit of cross-posting in their TW and FB accounts since they attract 2× and 30% more engagement with cross-posts than non-cross-posts, respectively. However, cross-posts in G+ leads to halve the engagement as compared to non-cross-posts. Finally professional users equally prefer FB and TW as initial source of information, but they rarely choose G+.

## VII. Acknowledgments

This work is partially supported by the European Celtic-Plus project CONVINcE, eCOUSIN (EU-FP7-318398) and ITEA3 CAP. This work is also funded by the Ministerio de Economa y Competitividad of SPAIN through the project BigDatAAM (FIS2013-47532-C3-3-P). We would like thank Reza Motamedi, Reza Rejaie, Roberto Gonzlez and Ruben Cuevas for providing Twitter and Google+ dataset to be used in this study.

## References

[1] R. Ottoni, D. Casas, J. Pesce, W. Meira, C. Wilson, A. Mislove, and V. Almeida, "Of Pins and Tweets: Investigating How Users Behave Across Image- and Text-Based Social Networks," in *ICWSM*, 2014.

[2] A. Crymble, "An Analysis of Twitter and Facebook use by the Archival Community," in *Journal of Association of Canadian Archivists*, 2010.

[3] R. Motamedi, R. Gonzalez, R. Farahbakhsh, A. Cuevas, R. Cuevas, and R. Rejaie, "Characterizing group-level user behavior in major online social networks," Technical Report available at: http://mirage.cs.uoregon.edu/pub/CIS-TR-2013-09.pdf, 2014.

[4] Socialbakers, *Socialbakers Portal*, 2014, http://www.socialbakers.com/.

[5] R. Gonzalez, R. Cuevas, R. Motamedi, R. Rejaie, and A. Cuevas, "Google+ or google-?: dissecting the evolution of the new osn in its first year," in *WWW*, 2013.

[6] *NLTK madules for similarity*, http://www.nltk.org/api/nltk.metrics.html.

[7] A. Singhal, "Modern Information Retrieval: A Brief Overview," *IEEE Data(base) Engineering Bulletin*, vol. 24, pp. 35–43, 2001.

[8] Q. X. Yang, S. Y. Sung, L. Chun, L. Zhao, and S. Peng, "Faster algorithm of string comparison," *Pattern Analysis and Applications*, 2003.

[9] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The Anatomy of the Facebook Social Graph," *CoRR*, vol. abs/1111.4503, 2011.

[10] M. Gjoka, M. Kurant, and A. Markopoulou, "Walking in facebook: A case study of unbiased sampling of osns," in *IEEE INFOCOM*, 2010.

[11] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?" in *WWW*, 2010.

[12] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *ICWSM*, 2010.

[13] G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida, "New kid on the block: Exploring the google+ social graph," in *ACM IMC*, 2012.

[14] D. Schiöberg, F. Schneider, H. Schiöberg, S. Schmid, S. Uhlig, and A. Feldmann, "Tracing the birth of an osn: Social graph and profile analysis in google," in *ACM WebSci*, 2012.

[15] R. Rejaie, M. Torkjazi, M. Valafar, and W. Willinger, "Sizing Up Online Social Networks," *IEEE Network*, 2010.

[16] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *ACM IMC*, 2007.

[17] D. J. Hughesa, M. Rowea, M. Bateya, and A. Leea, "A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage," in *Journal of Association of Canadian Archivists*, 2011.