Characterization of Professional Users Strategies in major OSNs

Reza Farahbakhsh
Institut Mines-Telecom, Telecom SudParis
reza.farahbakhsh@it-sudparis.eu

Ángel Cuevas
Universidad Carlos III de Madrid & Telecom SudParis
acrumin@it.uc3m.es

Rubén Cuevas
Universidad Carlos III de Madrid
rcuevas@it.uc3m.es

Noel Crespi
Institut Mines-Telecom, Telecom SudParis
noel.crespi@mines-telecom.fr

ABSTRACT

On-line Social Networks (OSNs) are being used intensively by countless professional players (e.g., large companies, politicians, athletes, celebrities, etc.) as a means of interacting with a huge amount of regular OSN users. This has led to an increasing research interest that aims at understanding what are the strategies of professional users in OSNs. In this paper we study the global strategy of professional users by sector (e.g., Cars companies, Clothing companies, Politician, etc.). To perform that analysis we have to first validate that users belonging to the same sector/category present a similar strategy in their use of OSNs. To find whether there are some sectors fulfilling that requirement, we use a dataset of 616 professional users with active accounts in the three most popular OSNs: Facebook (FB), Twitter (TW) and Google+ (G+). We find 8 categories whose users present similar behavioural elements: Athletes, Cars, Media News, Movie, Musician-Band, News Website, Politician, and Sports Teams. We describe the behaviour for these categories across FB, TW and G+ highlighting those elements that differentiate each strategy. Finally, we present a simple methodology that allows us to estimate the success of each strategy based on the number of reactions per post that a category is able to attract.

Keywords

OSN, Behaviour, Cross-Posting, Professional users, Strategy, Success.

1. 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. This tremendous success has created a golden opportunity to professional players (i.e. big industry brands, politicians, celebrities, etc.) in order to: interact with a huge amount of potential customers/voters/fans, improve their reputation and popularity, run marketing campaigns, etc. The presence and interest of professional users in OSNs as well as their concern to engage more people [5] with their OSNs accounts is becoming so relevant that we can even find an award ceremony to best professionals users in social media [1].

In this context there is an increasing research interest, especially in the area of management and marketing, to study what are the strategies[1] that professional users apply in their use of OSNs [12, 13, 7]. It seems that understanding the factors that allow professional users to engage more people with their OSN activity will have a tremendous value in the future for marketing purposes. To the best of our knowledge most of the studies available in the literature only focus on a limited number of users and extract very particular conclusions for those users that cannot be generalized. Furthermore, all previous studies are either based on manual inspection of OSNs accounts [27] or interviews [31] that cover very few aspects that again lead to not generalizable conclusions. Therefore, we believe that a large-scale data-driven approach based on the actual activity of a large number of professional users across major OSNs will help to shed light into the challenging problem of devising the way professional users utilize OSNs. Towards this end in this paper we rely on a dataset formed by 616 very popular users with active accounts in FB, TW and G+. For each user we capture her activity (i.e., published posts) in the three systems over a long-term time window that overall generates a corpus of 2M posts.

In contrast to previous studies we do not aim at studying the strategy of individual users. Instead, our main goal is to make a global analysis to characterize the strategy of a particular sector/category (e.g., Cars Industry, Politician, Athletes, News Media, etc) in OSNs. This analysis can be only conducted for those sectors that fulfil the following hypothesis: *professional users that belong to a particular sector present a similar strategy in OSNs.* Therefore, the first objective of this paper is to determine whether this hypothesis is true for some sector. For this we classify the 616 users in our dataset into 62 categories according to the sector reflected by their FB account. Out of these 62 groups only 16 had enough users to perform a meaningful validation of the hypothesis. We apply the methodology proposed

[1]In this paper, we will use indistinguishably the terms strategy and behaviour to refer the way a professional utilizes an OSN.

in [15] that determines whether the behaviour of the users within a category is significantly similar and, in addition, differs from the behaviour of the users outside that category. The results reveal 8 categories whose users present a common behaviour. These categories are: Athletes, Cars, Media News, Movie, Musician-Band, News Website, Politician, and Sports Teams. After discovering 8 sector fulfilling the baseline hypothesis, we devote our effort to derive the behavioural elements that characterize their use of OSNs.

We base our analysis in a set of meaningful behavioural elements that allow us to discriminate the strategy of each sector. These elements include: activity rate, preference among FB, TW and G+, popularity and type of content published. Using these behavioural elements we are able to describe the strategy and highlight the differential characteristics of each category. There is a last element that, to the best of our knowledge, has never been used to analyze the strategy of professional users across multiple OSNs, which is referred to as *cross-posting activity*. This element captures the volume of common information that a user publishes in more than one OSN. This means, when a professional user wants to post some information she can decide to publish it in a single OSN, or in multiple OSNs. Even more, when she decides to post it in multiple OSNs, there are several combinations of OSNs she could use (e.g., FB-TW or FB-G+ or TW-G+, or the three OSN in our work). Hence, we believe that the *cross-posting activity* of a user is an important behavioural element that for instance reveals whether a user utilizes each OSN for different purposes or not. In this paper we dedicate a full section to characterize the *cross-posting phenomenon* across professional users.

Finally, to conclude this research we address the very challenging question of whether the strategies implemented by each category are successful or not. To the best of our knowledge there is no standard mechanism in the literature that allows measuring the success of a strategy in OSNs. Therefore, in this paper we propose a simple methodology to quantitatively measure such success. The rationale of this methodology is to estimate the number of reactions per post a category should attract based on its popularity, and compare that estimation to the actual number of reactions received by the category. We provide an estimation of the success of each category for eight types of reaction: FB Likes, FB comments, FB shares, G+ +1s, G+ reshares, TW favourite and TW retweets.

The main findings of our research can be summarized as follows:

(1) Cross-posting is a frequent practice across professional users. In addition, the cross-posting phenomenon mainly happens between FB and TW, but it is also relevant between FB and G+. However, professional users rarely publish the same information in their TW and G+ accounts.

(2) We demonstrated that for some sectors professional users present a common behaviour. The sectors we found in the paper that fulfil this statement are: Athletes, Cars, Media News, Movie, Musician-Band, News Website, Politician, and Sports Teams.

(3) Each of the categories listed above present differential elements in their use of OSNs. For instance, Athletes activity and preference is biased to TW; categories related to news are extremely active in the three OSNs; Cars is the category with major interest in G+, and Movie shows a low activity and a clear preference for FB.

(4) The categories listed above can be further clustered into three significant groups based on the similarities in their strategies: *individual users* (Athletes, Musician-Band, and Politician), *commercial brands* (Cars and Sport Teams) and *news* (Media News and News Website).

(5) We demonstrate that the level of engagement of a professional user is linearly correlated to her popularity, which allows us to define a model that estimates the number of reactions per post a category should obtain according to its popularity.

(6) The only categories with a successful strategy in FB are Movie (which is successful in all OSNs) and Politician, which is the only category that do not cover the engagement expectation in G+. Similarly, the only two categories that fail in attracting the expected number of reactions in TW are Media News and News Website.

The remainder of this paper is as follow. Section 2 introduces the data collection process and professional users selection, and Section 3 describes our methodology to identify cross-posts and characterizes the cross-posting phenomenon. In Section 4 we verify our baseline hypothesis for 16 categories. Section 5 describes the strategies of those categories that fulfil the mentioned hypothesis and Section 6 studies whether those strategies are successful or not. Finally, Section 7 discusses related work, and Section 8 concludes the paper.

2. DATA COLLECTION METHODOLOGY

In this section we explain the selection of professional OSNs, describe our crawlers to collect data from those users, and introduce the way we classify the users into categories.

The first concept we need to define is what we refer to as OSNs *professional user*. It corresponds to a social profile behind a private company, public body or very popular individuals that usually have presence in most of the major OSNs and pursuit different goals than regular OSN users. This is, professional users utilize OSNs to increase their visibility, improve their popularity, enhance their reputation, etc. Some examples of professional users are: companies, celebrities, politicians, etc.

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 collected for a previous work [24] that includes thousands of very popular professional and regular users with an account in at least one of these OSNs. 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 have an active account in the three systems and satisfy the popularity requirement. We validated that the selected users were actually very relevant in at least two of the three considered OSNs by means of an external source [4] that ranks professional users in each system in terms of popularity. It must be noted that in many cases the selected users appear in relatively high positions in the three rankings.

In order to define the strategies of these users we need to collect the activity of these users as well as information associated to each activity (i.e., post) like: timestamp, type of content, number of reactions, description of the post, etc. Following we briefly introduce the crawlers developed to retrieve the activity of professional users from each OSN. For

Table 1: Dataset description

OSN	#posts	avg(posts)	%cross posts	#like	#comments	#shares
FB	423K	695	33.63	2.9B	98M	235M
G+	175K	304	29.36	27M	5M	3M
TW	1.7M	2648	7.17	274M	-	491M

a more detailed information of these crawlers we refer the reader to [24, 17]:

FB crawler.

The crawler receives a user ID (or username) as input and uses the FB API to collect the posts published by the user in her FB account. We note that the 616 users are FB pages (instead of regular accounts) that is the FB type of account suitable for professional users. The API provides quite a lot information from a post from which the most relevant for our work is: (*i*) the description of the post that refers to the text included by the user in that post, (*ii*) the timestamp associated to the exact publication time of the post, (*iii*) the type of content associated to the post, which we classify as[2]: *photo*, *video*, *link* (when the post includes an url) or *text* (that refers to the post that only include text), and (*iv*) the number of reactions associated to the post. There are three types of reactions in FB: likes, comments and shares. It must be noted that FB API imposes a maximum threshold of 600 queries every 10 minutes. Hence, in order to speed up our data collection process, we used multiple instances of the crawler working in parallel.

TW crawler.

The crawler receives as input a user identifier that can be either the user's id or the user's screen name and queries the Twitter API to obtain the user's profile attributes, the total number of published tweets, and the last 3,200 tweets posted by the user along with the number of reactions associated with each one of the user's tweets: favourites and retweets. Unfortunately the API did not allow to collect the responses (i.e., comments) for a tweet. We also classify tweets into to different types: links (when the tweet includes an url) and text. At the time of our data collection TW did not allow to include photos and videos within a tweet. Twitter imposes a limit of 150 requests per hour per IP address. To overcome this limitation, we use PlanetLab [9] infrastructure to parallelize our data collection process.

G+ Crawler.

This crawler is composed by two modules. The first one collects the public profile information as well as the connectivity information of all the users in the largest connected component (LCC) of G+. This module is a web-crawler that parses the web page of G+ users to collect the previous information. The second module uses the G+ API to collect all the public posts as well as their type, and their associated reactions. G+ posts type and reactions are the same than in FB, but the reactions receive different names:+1s, comments and reshares, respectively. Google limits the number of queries to the G+ API to 10K per hour per access token. In order to overcome this limitation we have created several hundred accounts with their correspondent access tokens.

Following we highlight three relevant elements related to the data collection process and the implications they have for our research. (*i*) Our crawlers only collect public posts. However, for this particular research this is not a limiting factor since most professional publishers posts are public. (*ii*) We had to convert the timestamp associated to the collected posts to a common time zone taking into account seasonal time changes. We decided to use GMT. (*iii*) In order to properly study the strategy of a user across FB, TW and G+ we need to use the same temporal window in the three systems. TW only allows to retrieve the last 3,200 tweets of a user that imposes a temporal limitation that should be extrapolated to FB and G+. Then, the time window employed in each user ranges between the last collection day, 24 Aug. 2013[3] (which is the same for all users), and the date from which we can retrieve the oldest tweet (which varies form user to user). This guarantees an analysis of the activity for each user in the three systems during the same period.

Table 2: Categories in the Dataset with more than 10 users.

#	category	#user	#	category	#user
1	Musician_band	134	9	Food_beverages	18
2	Tv_show	40	10	Website	16
3	Public_figure	32	11	Cars	15
4	Media_news_publishing	29	12	Clothing	13
5	Actor_director	28	13	Movie	12
6	Athlete	24	14	News_media_website	12
7	Sports_team	23	15	Tv_network	12
8	Product_service	20	16	Politician	6

Table 1 summarizes the datasets used in this paper. In total, we analyze more than 2M posts published by 616 professional publishers in FB, TW and G+.

Finally, in order to address the main goal of the paper we need to assign the 616 users to the categories they are representing. Towards this end we have used a straightforward approach based on the category each professional user selects when they register their FB page. Therefore we assign each user to the category they have selected in FB. Overall, the 616 users are classified into 62 different categories. The goal of this paper is to find whether users in some category present a common behaviour on their utilization of OSNs, describe the strategy in that category and determine its degree of success in FB, TW and G+. We can only perform that analysis for those categories in our dataset that includes enough users. Then, we have decided to study categories represented by at least 10 users in our dataset. Table 2 shows the number of users associated to the 15 categories[4] that meet that requirement[5]. We have made an exception for the category *Politician*, which is formed by only 6 users. Although we acknowledge that 6 users must not be enough to generalize the strategy of politicians, we believe it is worthy to study such an interesting category. We believe the 16 categories we are going to analyze present a quite interesting heterogeneity of sectors (e.g., popular individuals, big industrial companies, news agencies, TV or the Internet) that address different audiences.

[2]There are some other categories of posts but they are very marginal in our dataset. This statement also holds for G+.

[3]It must be noted that our dataset does not include novel features released by any of the analyzed OSNs after that period.

[4]We use News Website instead of News_media_website and Media News instead of Media_news_publishing from now on in the paper.

[5]The reader can find the name of the users in each category in section 10.

3. CROSS-POSTING

This paper studies the behaviour of users across different OSNs rather than the isolated behaviour in each system. Hence, a key element of this analysis is to understand whether a user publishes different information in each OSN, or contrary she publishes the same information in several OSNs. Basically, when a user wants to publish some information she can decide to publish it only in one OSN, in FB and TW, in FB and G+, in TW and G+, or in the three systems. Therefore, the portion of information that a user publishes within each option is an important element of the publication strategy of that user. We refer to the action of publishing the same information in several OSNs as *cross-posting*. Then, if we find two posts in the FB and TW (or any other combination of two or three OSNs) accounts of a given user that contains the same information we refer to them as *cross-posts*. Contrary, we will refer as *regular-posts* to those posts that appear exclusively in a single OSN. In this Section, we describe and validate our methodology to identify cross-posts, and characterize the cross-posting phenomenon across FB, TW and G+ using the 616 professional users in our dataset.

3.1 Methodology to Identify Cross-Posts

In order to compare the 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 with a post) of the two posts[6], P_1 retrieved from the account of user U in OSN_A and P_2 published by U in her account in OSN_B, our algorithm proceeds as follows:

(1) We compare P_1 and P_2 using NTLK Fuzzy Match [2] which provides a binary decision based on the similarity of the compared texts. NTLK Fuzzy Match generates a positive answer (i.e., the same text) when both texts are very similar and only differ in some few characters. Therefore, in the context of cross-posting analysis if NTLK Fuzzy Match determines that $P1$ and $P2$ are similar, we can safely classify them as cross-post. However, in the case in which the output is negative we cannot guarantee that P_1 and P_2 are not referring to the same information, thus we cannot classify them as regular posts. 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 [28] and string similarity [32]. These two metrics provide as output a value ranging between 0 and 1, so that the closer the output is to 1, the more similar P_1 is to P_2. Based on the obtained results, we classify P_1 and P_2 as cross-post if both metrics, cosine similarity and string similarity, are ≥ 0.5. Later in this section we validate our methodology and demonstrate why we have selected the 0.5 threshold.

It must be noted that P_1 is compared to P_2 in case P_2 was published in a period ranging between one week before and one week after P_1 was published. In addition, we highlight that our algorithm is not bound to any particular alphabet so it can be applied in multiple languages.

[6] 98% of the posts in our dataset include a description

Table 3: Validation of the cross-posts identification methodology. The table shows the false positive (FP) and false negative (FN) ratio for different similarity thresholds (ST) in percentage.

ST>0.3 similarity		ST>0.5 similarity		ST>0.7 similarity	
FP	FN	FP	FN	FP	FN
15%	0.19%	0.14%	1.12%	0.02%	4.6%

In order to ensure the accuracy of the proposed methodology 3 people manually classified 13K random posts as cross-posts or regular-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 regular-post 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 individuals performing the manual inspection indicate that both posts contain 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. Table 3 shows the false positive and false negative rate for our algorithm for each of the selected 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%).

We applied the described methodology to the selected 616 OSN professional users and found 176K cross-posts across their OSNs accounts.

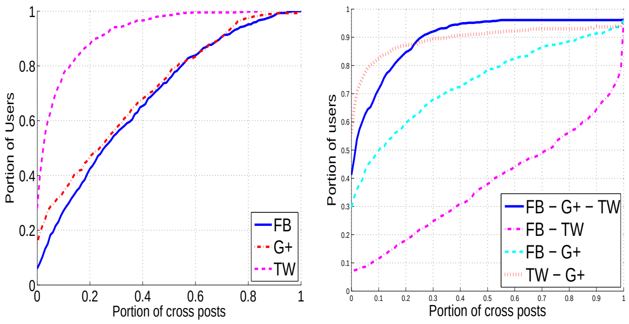
3.2 Cross-Posting Characterization

The first question we aim to answer is whether the cross-posting phenomenon exists in the activity of professional users, and what is its weight in FB, TW and G+. We then look at how this cross-posting occurs among the three OSNs under analysis. To this end, we quantify the fraction of cross-posting between FB-G+, FB-TW, TW-G+ and FB-TW-G+, in order to determine what set of OSNs is actually used more frequently by users to publish the same information. Finally, we also look at the preference of the users in our dataset for FB, TW and G+. We borrow the concept of preference from [25]. The authors define preference for an OSN as the bias of a user to choose more frequently that OSN as initial source of information when she aims at posting a given information in several OSNs.

3.2.1 *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[7]. Figure 1(a) shows the CDF for the portion of cross posts across the users 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 ≤20% of cross-posts in their FB accounts.

[7] 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)

Figure 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-TW-G+).

The first immediate conclusion extracted from the graph is that most of the professional users have published some cross-posts. Only 6%, 15% and 28% of the users in FB, G+ and TW, respectively, did not present any cross-post. Hence, the first conclusion is that in general professional users find some value in cross-posting.

If we compare the results obtained for the three OSNs, we clearly observe that, in relative terms, 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 indicates 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 other 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 large portion of professional users with intensive cross posting activity. In particular, 25%, 23% and 1.5% of the analyzed users, in FB, G+ and TW, respectively, published more cross-posts than regular-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× higher than in FB and G+. Hence, although TW presents a much lower cross-posting activity in relative terms, it actually has a larger number of cross-posts than G+, and it is much closer to FB in the absolute number of cross-posts. In median, a professional user presents 114, 85 and 20 cross-posts in FB, TW and G+, respectively.

3.2.2 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. Our goal is quantifying whether professional users prefer to share information in FB and TW, or rather it is more frequent finding common posts in FB and G+, or if they have more 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 of whether the first appearance was in that OSN or another one) and compute which portion of them also appears in TW, which portion in G+ and which portion in both TW and G+. We repeat

Table 4: Preferred OSN per user

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

the same process for user U's TW and G+ accounts. Therefore, for each user we know 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 616 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. For instance the point x=0.4, y=0.3 in the $FB - TW$ line indicates that 30% of the users publish ≤40% of their cross posts in FB and TW. The results reveal 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 on FB and TW. In addition, we find that only 8% of the users never shared a post between their FB and TW accounts, while this value grows to 30% between FB and G+, to 40% when the three OSNs are involved, and goes to 55% when we consider TW and G+. Therefore, this last result surprisingly states that is more likely that a user publishes a given information in the three OSNs than just in TW and G+.

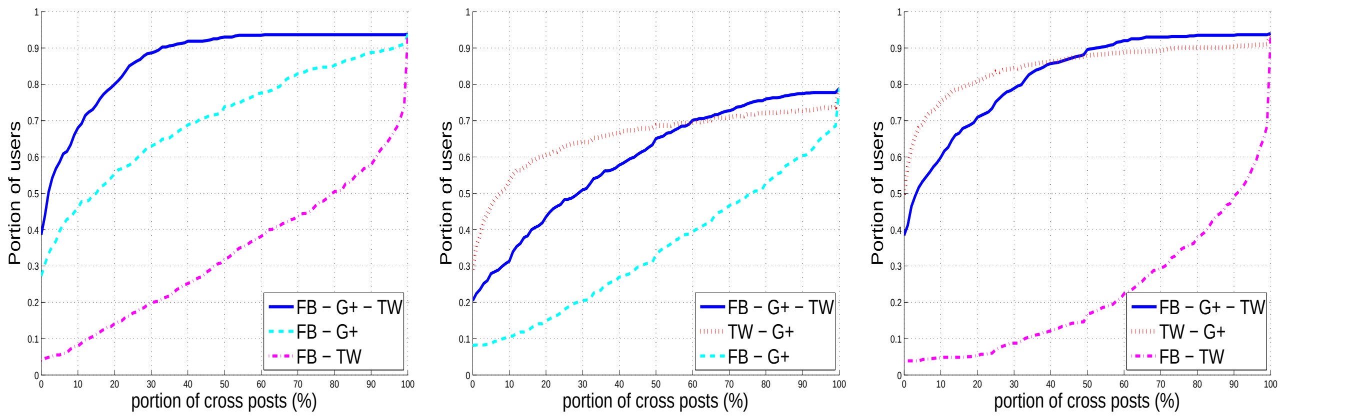
In order to complete this analysis we repeated the experiment for each OSN by only considering those cross-posts that were first published in each OSN, and we obtained the same conclusions. Figure 2 shows the result of this study.

3.2.3 Preference of professional publishers

We want to understand what is the OSN that professional users prefer to publish first the information. Answering this question will roughly determine what is the OSN that professional users value most for publishing an information that they plan to post in two or more OSNs. We define the preferred OSN of a user as the one she selected in first place for most of her cross-posts [25]. For instance, if a user has generated 20 cross-posts from which 10 were first published on FB, 6 on G+ and 4 on TW, we define FB as the preferred OSN for that user. Table 4 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 their initial OSN for publishing their post. Furthermore, we compute the number of users that select first a particular OSN for more than 80% of their cross-posts, which shows a strong preference. There are 102 (16.56%), 75 (12.18%), and 5 (0.8%) users with a strong preference for TW, FB and G+, respectively. 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+.

4. DETECTION OF COMMON STRATEGIES BY SECTORS

The goal of this section is to verify the baseline hypothesis of whether the users of a particular sector present a similar behaviour in their use of OSNs. Then we first introduce the behavioural metrics used to describe the strategy of a



(a) FB (b) G+ (c) TW

Figure 2: CDF for portion of cross posts in each possible cross-posting pattern for the cross-posts initiated in FB (a), G+ (b) and TW (c)

user, and later apply the methodology proposed in [15] to discriminate which categories follows our hypothesis.

4.1 Metrics to Capture the Behaviour

The strategy of a user is defined by the decisions that she takes when posting information across several OSNs. Therefore, the elements we use to define the activity are behavioural metrics directly related to those decisions. Each behavioural metric is captured with one (or more) values in each OSN as it is detailed below. Overall each user is represented with a behavioural vector of 33 values that defines her strategy across FB, TW and G+. We wanted to provide the same weight to all the parameters, hence all the values range between 0 and 1 in the vector. This has led us to normalize one of the metrics, the activity rate. We have performed the normalization using the 90th-percentile[8] of that parameter considering all the users in our dataset. All the users with a value above the 90th-percentile was assigned a value equal to 1 in the normalization. Note that we perform the normalization individually for each OSN.

Activity rate: We measure the average posts/day published by the user. As it is reported in [24], OSN users are intrinsically much more active in TW than in FB and G+. Therefore, we are interested on knowing how active is a user in a particular OSN with respect to the activity of other users in that OSN. With the proposed normalization for this metric we achieve that goal. This metric generates 3 values in the behavioural vector, one per OSN.

Fraction of Cross-Posting: We use as metric the portion of cross-posts in each OSN per user (3 values in the vector).

Cross-Posting pattern: We use as metric the portion of cross-posts happening in each possible OSN combination, i.e., FB-TW, FB-G+, TW-G+ or FB-TW-G+ (4 values in the vector).

Preference: This element is measured using the portion of cross-posts initiated in each OSN. This metric allows us to establish what is the preference of a user among the evaluated OSNs (3 values in the vector).

Type of content in regular-posts: This metric measures the portion of posts assigned to different type of content from the regular posts published by the user. In the case of FB and G+ the options are: photos, videos, links and text. In the case of TW only text or link. This metric generates 4 values in the vector for FB, one per type of content, 4 values in G+ and 2 Values in TW (10 values in total in the vector).

Type of content in cross-posts: This metric is similar to the previous one but in this case it only considers cross-posts (10 values in the vector).

[8] We did not use the maximum since we have checked that usually for the parameters we had to normalize the maximum value was actually an outlier.

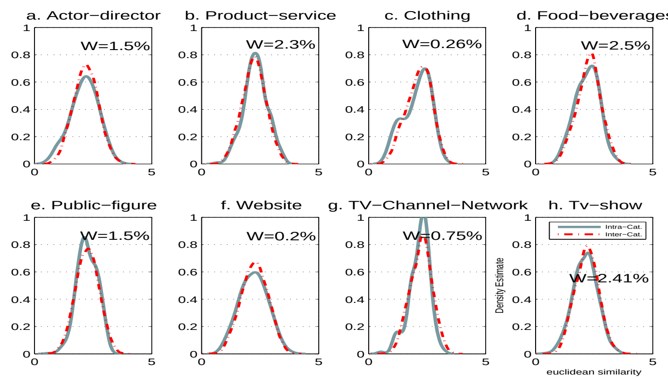


Figure 3: Kernel Density Estimation of the intra-category and inter-category euclidean distance for those categories whose users do not present a common strategy.

4.2 Identifying Categories Whose Users Present a Similar Strategy

We compare the similarity in the strategy of two different users by computing the euclidean distance between their vectors. Hence, the lower the euclidean distance the closer the strategies of the two users are. We can apply this process to compute what we refer to as intra-category and inter-category similarity. The former refers to the euclidean distance between each pair of users within the category, while the latter is represented by the euclidean distance of each user in the category to all the user outside that category.

We now apply the methodology proposed in [15] to find what are the categories whose users present a similar strategy across FB, TW and G+. First, we measure the intra-category and inter-category cohesion of each category using a Kernel Density Estimation (KDE) [18] method, where cohesion is measured based on the euclidean distance. In addition, for each category, we run the Wilcoxon rank-sum

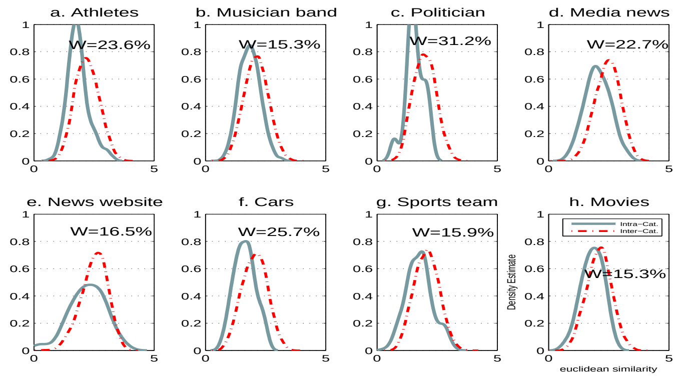


Figure 4: Kernel Density Estimation of the intra-category and inter-category euclidean distance for those categories whose users present a common strategy.

test [30] on the distributions of the intra-category and inter-category euclidean distance. This is a non-parametric test of the null hypothesis that two populations are the same. The Wilcoxon test also provides the parameter W that measures the distance between the median of both distributions. In our analysis $W = Median\_inter - Median\_intra$, thus the larger W is the stronger is the intra-category cohesion. We note that we compute the parameter W as the difference of the medians in percentage (instead of absolute term) that provides more clear insights.

Figure 3 shows the KDE results for those categories in which the euclidean distance among the users inside the category is very similar to the euclidean distance with external users. This can be easily observed since the distributions are overlapped. Aligned to this result, the Wilcoxon test validates the null-hypothesis in all the cases (i.e., the distributions are the same), and W is below 2.5% in all the cases. Therefore, we conclude that the users in those eight categories do not present a common behaviour.

Contrary, Figure 4 depicts the KDE for those categories with a major intra-category cohesion. In this case, the Wilcoxon test rejects the null-hypothesis in all cases. This means that the intra-category and inter-category distributions are statistically different (p-$value$<0.001) for these eight categories. This statement is supported by the fact that for these categories W ranges between 15% and 30%. Therefore, these results uncover eight categories whose members present common behavioural elements (i.e., strategy) that globally differs from the strategy of the users outside that category. These eight categories are: Athletes, Cars, Media News, Movie, Musician-Band, News Website, Politician and Sport Team.

We note that from now on in the paper the strategy of each category will be represented by the centroid[9] of the category.

4.3 Similarity Between Categories' Behaviour

We have demonstrated that there are 8 categories whose users present a similar use of OSNs. However, the previous analysis neither says how close are the strategies of these categories nor defines the main elements of each strategy. In this subsection we address the first point, while the second question is covered in the next section.

[9]Each of the 33 values characterizing the centroid corresponds to the median of each metric across the users in the category.

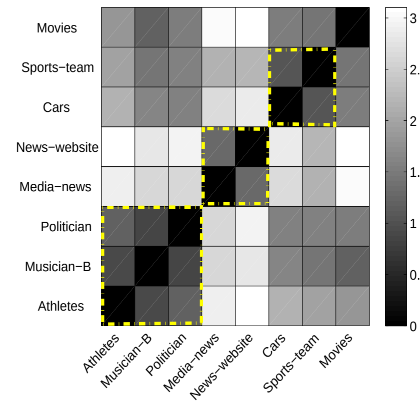


Figure 5: Colormap that represents the euclidean distance between the behaviour of the eight categories with a similar strategy. The closer the strategy of two categories is the darker the cell representing their euclidean distance. We find three relevant clusters among the analyzed users that are highlighted using a yellow dotted line.

To compare the strategies between two categories we calculate the euclidean distance between their centroids. Figure 5[10] shows a colormap in which each cell unveils the euclidean distance between the centroids of two categories. Visually, the closer the strategy of two categories is the darker the cell is[11].

The results reveals three interesting clusters. First, Media News and News Website have very different strategies to any other category, while they present some commonalties in their use of OSNs. Second, the categories that represent individual users, i.e., Athletes, Music-Band and Politician, present a more similar strategy among them than to other categories. Third, Cars and Sport Teams, the two categories representing companies, present a major similarity to each other than to any other category. Finally, Movie present a strategy that is neither far away nor close to any other category except the two categories referring to news.

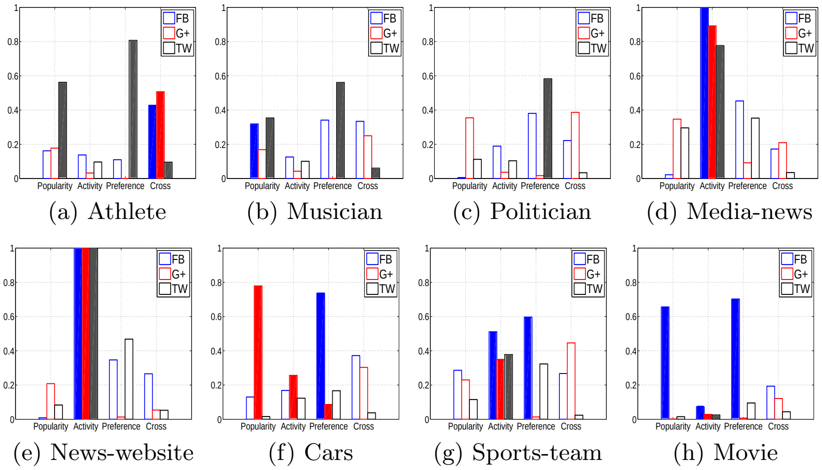
It is important to highlight that the fact that two categories present a higher similarity in their strategy does not mean they present exactly the same behaviour (i.e., the same values in the metrics). Instead, the correct interpretation is that those two categories will present some commonalities in some behavioural elements that make their strategies closer with respect to other categories.

5. UNVEILING STRATEGIES

In this Section we reveal and discuss what are the most significant elements in the strategy of the 8 categories under analysis. Towards this end we use all the behavioural elements introduced in Section 4 except *Cross-Posting pattern* because it is only relevant in the strategy of Cars. The other categories closely follow the general results reported in

[10]We advise the reader to visualize all the figures from this point in the computer to get a better color resolution.

[11]Note that in Figure 5 the black diagonal act as a mirror. The results are the same in the upper and lower part of the diagonal since the euclidean distance between two categories is a bidirectional parameter.



(a) Athlete (b) Musician (c) Politician (d) Media-news

(e) News-website (f) Cars (g) Sports-team (h) Movie

Figure 6: Bar plot that shows the value of the following metric for each category and OSN: *popularity*, *activity rate*, *preference* and *fraction of cross-posts*.

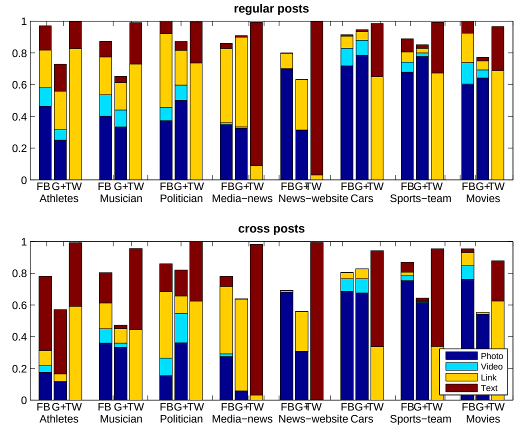


Figure 7: Bar plot that shows the type of content published in each category per OSN.

Section 3.2.2 for this metric. In addition to the behavioural parameters, we use the popularity (i.e., number of followers) of each category in each OSN in the analysis. The reason is that although the popularity is not a behavioural element itself, it can influence the decisions of a user. As we did for the activity rate, we have normalized the popularity using the 90th-percentile in each OSN.

Figure 6 shows one bar plot per category in which each bar shows the value of *popularity*, *activity rate*, *preference* and *fraction of cross-posts* in each OSN, respectively. We have highlighted in full color the bars that represent the most significant elements of the behaviour of each category. In addition, Figure 7 shows *type of content in regular-posts* and *type of content in cross-posts* for each category and OSN, respectively. Following, we describe the strategy of each category:

Athlete: It is the category with the strongest preference for FB and with the most intense cross-posting activity in the three OSNs. It presents a low activity in all OSNs compared to other categories. Regular posts are mostly photos and links in FB and G+, however cross-posts are dominated by text in these two OSNs. This is explained because most of the cross-posts are initiated by TW (as shown by the strong TW preference) and replicated in FB and G+ as text. Finally, it is the most popular category in TW, which may explain its strong preference for this OSN.

Musician-Band: This category presents a clear preference for TW and an important level of cross-posting in this OSN (only surpassed by Athletes). The posts published in FB and G+ are mostly audiovisual content, both in cross-posts and regular-posts. The activity rate is low in the three OSNs. Finally, in terms of popularity, Musician-Band is the second most popular category in FB and TW behind Movie and Athlete, respectively.

Politician: Similar to Athlete and Musician-Band this category presents a preference for TW as well as a low activity in all 3 OSNs. The most interesting behavioural element of Politician is that it uses different content in FB and G+. Politician publishes more links in FB than in G+, where it mostly publishes audiovisual content. They also opt for using links in most of the tweets.

Media News: The differential strategy of this category is clearly a very high activity rate in the three OSNs. Actually, this seems reasonable since the users in this category are news agencies, portals, etc that are continuously publishing recent news. In addition, a second particularity of Media News is that the most common type of content in FB and G+ is link. However, it very rarely uses links in TW. In addition, together with News Website, is the category with a more balanced preference between FB and TW.

News Website: As the previous category, the differential behavioural element of News Website is its extraordinary high activity rate in all OSNs. In addition, News Media Website also shows a quite balanced preference between FB and TW. Contrary to Media News, in this case posts in FB are mostly photos, while in G+ they are balanced between photos and links.

Cars: Cars is the category with a major interest in G+, which may be due to its high popularity in that OSN. The behavioural elements that shows that interest are: (*i*) it is the only category in which the selection of G+ as initial source of information is relevant (it happens in almost 10% of the cross-posts), (*ii*) Cars is the only category in which its (relative) activity rate is higher in G+ than in TW and FB, and (*iii*) Cars is the only category in which the cross-posting activity between TW and G+ is not negligible since this pattern appears in 15% of the cross-posts. Apart from its interest in G+, Cars is clearly biased to FB in terms of preference and mostly uses audiovisual content in its posts. This seems reasonable since the business of Cars companies has a lot to do with presenting an attractive view of their cars and this requires the use of audiovisual material.

Sports Team: There are three elements that denote the behaviour of Sport Teams. First, a clear preference for FB. Second, an intense use of photos in its posts. Three, a considerably high activity in the three OSNs compared to the other categories (with the exception of the two categories related to news).

Movie: The behaviour of this category is defined by a strong preference of FB, the use of photos in most of its FB and G+ posts, and the lowest activity rate in the three OSN among the categories under analysis. This happens because the OSN accounts associated to movies are only active in a short period of time around their release and later they just keep a residual activity. Finally, there is a big contrast in its popularity since it is the most popular category in FB, but the least popular in TW and G+.

We conclude our analysis by enumerating the common behavioural aspects for the three clusters identified in Section 4.3. (1) All the *individual* users present a preference for TW and a relatively low activity in all OSNs compared to other categories. (2) Cars and Sports Teams, which represent *commercial companies*, shows a clear preference for FB and mostly post audiovisual content in FB and G+. (3) The categories related to news reporting coincides in having a very high activity rate.

6. EVALUATION OF STRATEGIES SUCCESS

To conclude this paper we want to assess the success of the strategies adopted by the analyzed categories. To the best of our knowledge it does not exist any standard metric or methodology to evaluate the success of an strategy in OSNs. Our approach is based on the conviction that the number of reactions that a user attracts in her posts is the only objective available metric to capture the interest/engagement of end-users in the activity of a professional user. Therefore, in this paper we propose to measure the success of the strategy of a category as a function of the average number of reactions that the category attracts per post. We believe that the proposed methodology is a useful tool to estimate the success of a particular strategy in the context of this paper. However, we do not pretend to present it as a reference methodology to globally evaluate success in OSNs. Following, we first introduce our methodology and later we discuss the results extracted from applying it.

Table 5: Pearson coefficient, p-value, and Regression Coefficient of the correlation between popularity and reactions.

Reaction	PPMC	p-value	Regression Coefficient
FB likes	0.97	6e-5	1.78e-3
FB comments	0.94	4e-4	4.92e-5
FB shares	0.94	4e-4	1.14e-4
G+ +1s	0.76	0.03	7.02e-5
G+ comments	0.14	0.73	-
G+ reshares	0.94	5e-4	8.11e-6
TW favourite	0.78	0.02	2.07e-5
TW retweet	0.71	0.049	5.04e-5

6.1 Methodology to Measure the Success Degree of Strategies

Our methodology proposes to compute the success of the strategy of a category as the difference between the expected number of reactions per post that category should receive and the actual number of reactions it receives. Therefore, our goal is to propose a model that estimates the expected volume of reactions per post for the eight categories under discussion.

Our intuition is that the number of reactions that a user attracts in a post in an OSN is strongly correlated to her popularity in that OSN. Therefore, our first step is to validate this hypothesis that would allow us to formulate the expected number of reactions as a function of the popularity.

We calculate the Pearson Product-Moment Correlation Coefficient (PPMCC) between the popularity and all the

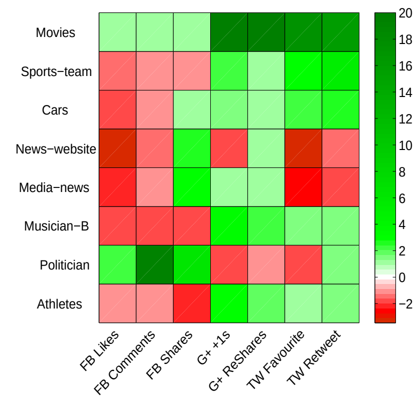


Figure 8: Colormap that represents the success of the strategy of each category across different types of reaction. The green color represents success and the red color represents failure.

reaction types separately. The PPMCC measures the degree of linear dependence between two variables, which becomes higher as the PPMCC moves to 1. Table 5 shows the PPMCC and p-value for the correlation associated to each reaction type. The results reveal a very strong linear positive correlation between popularity and volume of reactions per post for all type of reaction in all OSNs (PPMCC>0.7 and p-value<0.05). There is only one exception, G+ comments (p-value>0.05), which are omitted from our analysis in the rest of the section.

Based on these results, we propose a simple linear model that estimates the number of reactions a category should receive based on its popularity. Hence, we perform a linear regression to obtain the regression coefficient, listed in Table 5, associated to each type of reaction. In a nutshell, we estimate the number of reactions per post for a particular type of reaction in a category multiplying the popularity of that category by the regression coefficient for that reaction type.

Once we have the model to estimate the expected number of reactions we are able to evaluate the success of the different strategies. Figure 8 shows a colormap that represents the level of success of each category for each type of reaction. The colormap shows a positive (associated to green color) and negative (associated to red color) scale. For instance, a value of +2 implies that the category under analysis is obtaining 2× more reaction per post than what our model suggests. In contrast, a value of -2 indicates that the category is attracting 1/2 of the expected reactions per post. Note that the darker is the green color in a cell the higher is the success. Similarly, the darker is the red color in a cell the less efficient the strategy is. Each row corresponds to one category and presents a visual overview of the success of its strategy across the different OSNs and types of reactions.

6.2 Discussion of Strategies' Success

Movies is the only category with a successful strategy in all OSNs according to the volume of reactions it receives per post. This is an indicator that the adopted strategy is well adapted to the requirements of its audience in each OSN.

Athletes and Musician-Band are successful in TW and G+, but they fail in FB. Based on their clear preference for TW, it seems its strategy is adequate to cover their main objective, however they should modify their behaviour in FB in order to increase the engagement of end-users.

Politician has a successful strategy in FB, especially on attracting comments, but it fails in G+. In the case of TW it manages to get more retweets than expected, but does not cover the expectation in number of favourites. Its strategy is fair enough in FB to cover the expected reactions. In the case of TW, if its major interest focuses on spreading tweets its strategy is also adequate.

It seems that the interest of Cars in G+ is obtaining its reward since it manages to attract more reactions than the estimation of our model. In contrast, it seems Cars should revise their behaviour in FB since it only succeeds on the number of shares, even though it has a strong preference for this OSN.

Sports Team fails in FB, but is successful in TW and G+. Therefore, it should change some behavioural aspects to increase their engagement in FB.

Finally, Media News and News Website categories present a quite similar success pattern with the exception of G+ likes. We believe the most important type of reaction for news agencies and portals is share, reshare and retweet, since their goal is to spread the reported news as much as possible. For these reactions they present an almost identical result that reflects a success in FB and G+, but a failure in TW. This is a quite negative outcome since TW is considered a very relevant communication channel to disseminate news nowadays.

7. RELATED WORK

There is two different type of works related to the contributions of this paper: users' behaviour analysis and strategy of professional users in social media.

User Behavior Analysis in OSNs.

There is some baseline works that study basic properties of OSNs such as connectivity, users' activity, users' profile, etc, in each of the three OSNs considered in our work: Facebook, [29, 16], Twitter [19, 8], and Google+ [21, 17]. In addition, there is some works that compare properties of users' across two or more OSNs [26, 23]. However, none of these works consider the same group of users across different OSNs since their goal is to characterize OSNs at a macroscopic level. There are some Internet portals [4, 3] that provide some basic information (e.g., number of followers, aggregated engagement or popularity trend) of very popular users in different social systems. However, this information is very limited to address complex issues like the ones we cover in this paper. More aligned to our work, we find some few studies that have analyzed the behaviour of the same users across different OSNs from different perspectives. Authors in [10] compare 195 users from the archival community and study their activity in FB and TW. However, it is a small-scale study that is based on 3K links to external documents. Hughesa et al.[11] demonstrate that there is a correlation between the personality of a person and how she uses FB and TW based on an analysis of 300 users. Finally, a recent work [25] studies the cross-posting phenomenon between TW and PinInterest using 30K regular users. To the best of our knowledge this is the only work that analyzes the cross-posting phenomenon at large scale apart from our study. However, there is significant differences since we focus on professional users and analyze the cross-posting in FB, TW and G+.

Strategy of Professional Users in Social Media.

There are a number of books [7, 13] and reports [22] that propose general guidelines to companies to enhance their marketing strategies in social media. However, most of these guidelines are based on qualitative elements rather than quantitative metrics. Following this line, authors in [20] manually look to the publishing activity of 11 brands from 6 different categories, and provide some general guidelines for the manager of those brands on how to enhance the engagement of their followers in social media. Another study [6] aims at studying the importance of brands' Fans and the Fans' friend as a key factor in the strategy of three Facebook accounts. However, their study is limited to just three brands and they only considered one metric, the number of fans for each brand. Therefore, the last two references only derive ad-hoc conclusion for very few users that cannot be generalized. We found some larger scale works like [27] where authors manually look to the type of activity of 275 non-profit organization profiles in FB. However, they just look at two elements: how the users disseminate their messages and what type of posts they are considering in their strategies. This work differentiates from our paper in three main aspects: they only look at FB, they do not look into professional users, and they only use type of content to evaluate the behaviour of the FB users. In addition, authors in [12] explore the strategic use of social media for 250 of U.S. based companies on Facebook, Twitter, and YouTube. Although this work is more similar to our study due to the analysis of multiple OSNs it present high differences in the methodology and the analyzed behavioural elements. First of all the authors in this paper rely on manual inspection of the accounts that is a much more subjective method than a data-dirven approach. In addition, they use a number of social metrics (adoption, integration, code of conduct, human voice, dialogic loop, activity and stakeholder willingness) that are not linked to the actual activity of a user in OSN, and again are subjective. In contrast to these previous works relying on manual inspection, we have found a number of works that uses surveys or interview community managers to analyze the strategy of some few brands. The most relevant work is [31] in which the authors interview nine community managers of NBA teams. This study just focus in a single sector and perform a qualitative analysis based on the replies of the community managers. Finally, we also find a couple of studies that attach the success of a social media brand to the popularity [6] and to the number of reactions [20]. However, none of them take into account that both parameters are related and that considering success using them isolated may lead to wrong conclusions.

In summary, the main novelties of our work compared to the previous studies are: (*i*) it is the first data-driven approach over a large number of professional users. (*ii*) It aims at understanding the strategies from a global point of view per sectors. (*iii*) It is a longitudinal study across the three major OSNs: FB, TW and G+. (*iv*) It proposes a quantitative estimation of the success of OSNs strategies.

8. CONCLUSIONS

This paper advances the state of the regarding the strategy used by professional users in OSNs in three main elements: (*i*) To the best of our knowledge this is the first study that follows a data-driven approach to analyze the strategy of professional users in OSNs. (*ii*) We evaluate the global strategy of some professional sectors in the three major OSNs, namely FB, TW and G+. In contrast, most previous work focuses in the analysis of individual users and obtain adhoc conclusions. (*iii*) To the best of our knowledge, this paper is the first one that proposes a quantitative estimation of the success of a strategy. In order to be able to make an analysis per sector, our first step has been to demonstrate that there are sectors whose users present similar behavioural elements that define a common strategy in OSNs. In particular, we have found eight sectors with a common strategy: Athletes, Cars, Media News, Movie, Musician-Band, News Website, Politician, and Sports Teams. The more interesting findings for the analyzed sectors are: (*i*) the two categories related to news show an extremely intense activity in the three OSNs; (*ii*) Athlete shows a strong preference for TW that directly impacts the information published in FB and G+; (*iii*) Cars gives a high value to G+ where they have a much stronger presence than any other category, and, (*iv*) Movie is very active around the release of the film but later the activity becomes residual. Finally, we estimate the success of each strategy. The success is measured as the difference between the actual volume of engagement (i.e., reactions per post) and the expected volume of engagement based on the popularity of the category. Movie is the only category that overpasses the engagement expectation in all OSNs. Politician is the only category, in addition to Movie, with a clear success in FB, but it is the only category that does not reach the expectation in G+. Finally, the news-related categories are the only ones that do not reach the expected engagement in TW, neither in retweets nor in favourites. In addition to all the previous findings, this work presents an aside contribution that characterizes the cross-posting phenomenon for professional users across FB, TW and G+. We have demonstrated that this phenomenon exists and is relevant. The dominant cross-posting pattern is FB-TW, while it is very rare finding information shared between TW and G+.

9. REFERENCES

[1] *Annual Shorty Awards Winners.* http://shortyawards.com/.

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

[3] *Pagedata Portal.* http://www.pagedatapro.com/.

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

[5] *WWE top 100 million facebook fans at wrestlemania.* http://corporate.wwe.com/news/2013/2013_04_10.jsp.

[6] Andrew Lipsman, G. M., and Rich, M. The power of like: How brands reach and influence fans through social media marketing. *Journal of advertising research* (2012).

[7] Barlow, M., and Thomas, D. B. In *The Executive's Guide to Enterprise Social Media Strategy: How Social Networks Are Radically Transforming Your Business.* John Wiley and Sons, 2010.

[8] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. Measuring user influence in twitter: The million follower fallacy. In *AAAI ICWSM* (2010).

[9] Chun, B., Culler, D., Roscoe, T., Bavier, A., Peterson, L., Wawrzoniak, M., and Bowman, M. Planetlab: an overlay testbed for broad-coverage services. *ACM SIGCOMM Computer Communication Review* (2003).

[10] Crymble, A. An Analysis of Twitter and Facebook Use by the Archival Community. In *The Journal of the Association of Canadian Archivists* (2010).

[11] David John Hughesa, Moss Rowea, M. B., and Leea, A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. In *The Journal of the Association of Canadian Archivists* (2011).

[12] DiStaso, M. W., and McCorkindale, T. A benchmark analysis of the strategic use of social media for fortuneâĂŹs most admired u.s. companies on facebook, twitter, and youtube. *Public Relations Journal, 7(1), 1-33* (2013).

[13] Evans, L. In *Social Media Marketing: Strategies for Engaging in Facebook, Twitter & Other Social Media.* Pearson Education, 2010.

[14] Farahbakhsh, R., Cuevas, A., Cuevas, R., and Crespi, N. Characterization of professional users strategies in major osns. Technical report available at: http://www.it.uc3m.es/acrumin/papers/WWW_TR_strategy.pdf, 2014.

[15] Ferrara, E., Varol, O., Menczer, F., and Flammini, A. Traveling trends: Social butterflies or frequent fliers? COSN '13, ACM (2013).

[16] Gjoka, M., Kurant, M., Butts, C., and Markopoulou, A. Walking in facebook: A case study of unbiased sampling of osns. In *IEEE INFOCOM* (2010).

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

[18] Hastie, T., Tibshirani, R., and Friedman, J. *The elements of statistical learning.* Springer, 2006.

[19] Kwak, H., Lee, C., Park, H., and Moon, S. What is Twitter, a Social Network or a News Media? In *WWW* (2010).

[20] Lisette de Vriesa, S. G., and Leeflang, P. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing, 2012, Vol.26(2), pp.83-91.*

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

[22] Martin, A., and van Bavel, R. Assessing the benefits of social networks for organizations. Tech. rep., 2013. {http://ftp.jrc.es/EURdoc/JRC78641.pdf}.

[23] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. Measurement and Analysis of Online Social Networks. In *ACM IMC* (2007).

[24] Motamedi, R., Gonzalez, R., Farahbakhsh, R., Cuevas, A., Cuevas, R., and Rejaie, R. 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.

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

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

[27] Richard D. Watersa, Emily Burnettb, A. L., and Lucasb, J. Engaging stakeholders through social networking: How nonprofit organizations are using facebook. *Public Relations Review, Elsevier* (2009).

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

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

[30] Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics bulletin* (1945), 80–83.

[31] Wysocki, M. The Role of Social Media in Sports Communication: An Analysis of NBA Teams Strategy. Tech. rep., 2012. {http://www.american.edu/soc/communication/upload/Capstone-Wysocki.pdf}.

[32] Yang, Q. X., Sung, S. Y., Chun, L., Zhao, L., and Peng, S. Faster algorithm of string comparison. *Pattern Analysis and Applications 6* (2003), 122–133.

10. APENDIX 1 - LIST OF USERS IN EACH CATEGORY

Table 6: User List 1

#	TW_id	fb_category	#	TW_id	fb_category	#	TW_id	fb_category
1	selenagomez	Actor_director	61	SergioRamos	Athlete	121	ChickfilA	Community
2	EyeOfJackieChan	Actor_director	62	DwightHoward	Athlete	122	GeekandSundry	Community
3	ashleytisdale	Actor_director	63	paulpierce34	Athlete	123	officialtulisa	Community
4	aplusk	Actor_director	64	TigerWoods	Athlete	124	cocorocha	Community
5	EmWatson	Actor_director	65	lancearmstrong	Athlete	125	SeaviewSurvey	Community
6	VanessaHudgens	Actor_director	66	carmeloanthony	Athlete	126	Disney	Company
7	ActorLeeMinHo	Actor_director	67	ReggieBush	Athlete	127	adidasoriginals	Company
8	ZacEfron	Actor_director	68	FloydMayweather	Athlete	128	Dior	Company
9	channingtatum	Actor_director	69	rioferdy5	Athlete	129	LACOSTE	Company
10	charliesheen	Actor_director	70	CP3	Athlete	130	CHANEL	Company
11	priyankachopra	Actor_director	71	andy_murray	Athlete	131	espn	Company
12	akshaykumar	Actor_director	72	Joey7Barton	Athlete	132	Ford	Company
13	iansomerhalder	Actor_director	73	Tbayne21	Athlete	133	DCComics	Company
14	RealHughJackman	Actor_director	74	paulocoelho	Author	134	ELLEmagazine	Company
15	ThatKevinSmith	Actor_director	75	DeepakChopra	Author	135	TechCrunch	Company
16	mark_wahlberg	Actor_director	76	ScottKelby	Author	136	victoriabeckham	Company
17	msleamichele	Actor_director	77	jr_raphael	Author	137	GM	Company
18	shaymitch	Actor_director	78	MoosePeterson	Author	138	Virgin	Company
19	okanbayulgen	Actor_director	79	carlzimmer	Author	139	dpreview	Company
20	peterfacinelli	Actor_director	80	jenny8lee	Author	140	Zagat	Company
21	WilliamShatner	Actor_director	81	edmunds	Automobiles_and_parts	141	thehipmunk	Computers_internet_website
22	JohnCleese	Actor_director	82	BMWGroup	Cars	142	GoogleDiscovery	Computers_internet_website
23	rainnwilson	Actor_director	83	BBC_TopGear	Cars	143	CNET	Computers_technology
24	adriangrenier	Actor_director	84	InsideFerrari	Cars	144	coachella	Concert_tour
25	rosemcgowan	Actor_director	85	Audi	Cars	145	DaniellePeazer	Dancer
26	katewalsh	Actor_director	86	NissanUSA	Cars	146	OMGFacts	Education_website
27	edward_burns	Actor_director	87	Porsche	Cars	147	WhatTheFFacts	Education_website
28	jimmyfallon	Actor_director	88	lamborghini	Cars	148	BlackBerry	Electronics
29	caseymckinnon	Actor_director	89	Jeep	Cars	149	nokia	Electronics
30	NASA	Aerospace_defense	90	Kia_Motors	Cars	150	LucianoHuck	Entertainer
31	SpaceX	Aerospace_defense	91	chevrolet	Cars	151	BeingSalmanKhan	Entertainer
32	zyngapoker	App_page	92	VW	Cars	152	fluffyguy	Entertainer
33	AngryBirds	App_page	93	Toyota	Cars	153	bellathorne	Entertainer
34	CandyCrushSaga	App_page	94	Cadillac	Cars	154	ParisHilton	Entertainer
35	Zoosk	App_page	95	MercedesAMG	Cars	155	SteveMartinToGo	Entertainer
36	slotomania	App_page	96	ChryslerAutos	Cars	156	NickCannon	Entertainer
37	BubbleWitchSaga	App_page	97	FiskerAuto	Cars	157	SabrinaSatoReal	Entertainer
38	pizap	App_page	98	FoodRev	Cause	158	shwood	Entertainer
39	tetrisbattle	App_page	99	VictoriasSecret	Clothing	159	AshleyEsqueda	Entertainer
40	9GAG	App_page	100	ZARA	Clothing	160	TokyoOtakuMode	Entertainment_website
41	instagram	App_page	101	LEVIS	Clothing	161	sitevagalume	Entertainment_website
42	GangsofBoomtown	App_page	102	Burberry	Clothing	162	eonline	Entertainment_website
43	GeErgen	Artist	103	hm	Clothing	163	IGN	Entertainment_website
44	adde_adesokan	Artist	104	gucci	Clothing	164	Oatmeal	Entertainment_website
45	garybaseman	Artist	105	HollisterCo	Clothing	165	younghollywood	Entertainment_website
46	RodneyPike	Artist	106	PUMA	Clothing	166	someecards	Entertainment_website
47	jpattersonphoto	Artist	107	Abercrombie	Clothing	167	CELEBUZZ	Entertainment_website
48	lomokev	Artist	108	armani	Clothing	168	Blogger	Entertainment_website
49	thomashawk	Artist	109	ASOS	Clothing	169	redbull	Food_beverages
50	Cristiano	Athlete	110	shopbop	Clothing	170	Starbucks	Food_beverages
51	nikefootball	Athlete	111	saks	Clothing	171	McDonalds	Food_beverages
52	JohnCena	Athlete	112	MaterialGirl	Clothing	172	Skittles	Food_beverages
53	KingJames	Athlete	113	KevinHart4real	Comedian	173	Pringles	Food_beverages
54	andresiniesta8	Athlete	114	werevertumorro	Comedian	174	MonsterEnergy	Food_beverages
55	RafaelNadal	Athlete	115	rustyrockets	Comedian	175	nutellaUSA	Food_beverages
56	Njr92	Athlete	116	marcoluque	Comedian	176	drpepper	Food_beverages
57	DwyaneWade	Athlete	117	DaniloGentili	Comedian	177	guarana	Food_beverages
58	3gerardpique	Athlete	118	DavidSpade	Comedian	178	frappuccino	Food_beverages
59	tonyhawk	Athlete	119	kassemg	Comedian	179	BWWings	Food_beverages
60	robdyrdek	Athlete	120	PapaCJ	Comedian	180	TacoBell	Food_beverages

Table 7: User List 2

#	TW_id	fb_category	#	TW_id	fb_category	#	TW_id	fb_category
181	pepsi	Food_beverages	241	rafinhabastos	Monarch	301	onedirection	Musician_band
182	DunkinDonuts	Food_beverages	242	avatarmovienews	Movie	302	tiesto	Musician_band
183	dominos	Food_beverages	243	Twilight	Movie	303	slipknot	Musician_band
184	WholeFoods	Food_beverages	244	TitanicMovie	Movie	304	u2com	Musician_band
185	kraftfoods	Food_beverages	245	transformers	Movie	305	RollingStones	Musician_band
186	CadburyUK	Food_beverages	246	RealDirtyDance	Movie	306	jason_mraz	Musician_band
187	TheSims3	Games_toys	247	JACKoffSPARROW	Movie	307	JessieJ	Musician_band
188	EASPORTSFIFA	Games_toys	248	Iron_Man	Movie	308	KingsOfLeon	Musician_band
189	NeedforSpeed	Games_toys	249	starwars	Movie	309	MariahCarey	Musician_band
190	gameloft	Games_toys	250	TheHungerGames	Movie	310	Slash	Musician_band
191	SkylandersGame	Games_toys	251	GrownUpsMovie	Movie	311	TaioCruz	Musician_band
192	SenSanders	Government_official	252	DespicableMe	Movie	312	radiohead	Musician_band
193	whitehouse	Government_organization	253	MuppetsStudio	Movie	313	Ricardo_Arjona	Musician_band
194	Sephora	Health_beauty	254	saatchi_gallery	Museum_art_gallery	314	jasonderulo	Musician_band
195	sarahpotempa	Health_beauty	255	vangoghmuseum	Museum_art_gallery	315	30SECONDSTOMARS	Musician_band
196	epicurious	Home_garden_website	256	TheGRAMMYs	Music_award	316	TPAIN	Musician_band
197	casacombr	Home_garden_website	257	rihanna	Musician_band	317	inna_ro	Musician_band
198	danieltosh	Interest	258	Eminem	Musician_band	318	ladyantebellum	Musician_band
199	hootsuite	Internet_software	259	shakira	Musician_band	319	MirandaCosgrove	Musician_band
200	OneLouderApps	Internet_software	260	michaeljackson	Musician_band	320	backstreetboys	Musician_band
201	UberSoc	Internet_software	261	ladygaga	Musician_band	321	manaoficial	Musician_band
202	Piaget	Jewelry_watches	262	katyperry	Musician_band	322	SeanKingston	Musician_band
203	MarceloTas	Journalist	263	linkinpark	Musician_band	323	deadmau5	Musician_band
204	CarlosLoret	Journalist	264	LilTunechi	Musician_band	324	4PlanB	Musician_band
205	stshank	Journalist	265	bobmarley	Musician_band	325	PearlJam	Musician_band
206	stran9ee	Local_business	266	taylorswift13	Musician_band	326	souljaboy	Musician_band
207	GlamourMagUK	Magazine	267	AvrilLavigne	Musician_band	327	WakaFlockaBSM	Musician_band
208	RollingStone	Magazine	268	OfficialAdele	Musician_band	328	Jason_Aldean	Musician_band
209	NewYorker	Magazine	269	davidguetta	Musician_band	329	nickjonas	Musician_band
210	InStyle	Magazine	270	usherraymondiv	Musician_band	330	bigtimerush	Musician_band
211	GQMagazine	Magazine	271	BrunoMars	Musician_band	331	Ludacris	Musician_band
212	NatGeo	Media_news_publishing	272	thebeatles	Musician_band	332	iamwill	Musician_band
213	WWE	Media_news_publishing	273	enrique305	Musician_band	333	tamerhosny	Musician_band
214	Playboy	Media_news_publishing	274	Pitbull	Musician_band	334	juanes	Musician_band
215	CNN	Media_news_publishing	275	Metallica	Musician_band	335	arminvanbuuren	Musician_band
216	FoxNews	Media_news_publishing	276	50cent	Musician_band	336	amrdiab	Musician_band
217	nytimes	Media_news_publishing	277	GreenDay	Musician_band	337	mirandalambert	Musician_band
218	TheOnion	Media_news_publishing	278	coldplay	Musician_band	338	CodySimpson	Musician_band
219	VEJA	Media_news_publishing	279	aliciakeys	Musician_band	339	NicoleScherzy	Musician_band
220	TheEconomist	Media_news_publishing	280	LMFAO	Musician_band	340	Sugarlandmusic	Musician_band
221	WSJ	Media_news_publishing	281	MileyCyrus	Musician_band	341	blakeshelton	Musician_band
222	TIME	Media_news_publishing	282	britneyspears	Musician_band	342	ricky_martin	Musician_band
223	NBCNews	Media_news_publishing	283	keshasuxx	Musician_band	343	agnezmo	Musician_band
224	MotorTrend	Media_news_publishing	284	SnoopDogg	Musician_band	344	ivetesangalo	Musician_band
225	WIRED	Media_news_publishing	285	ChiliPeppers	Musician_band	345	train	Musician_band
226	EW	Media_news_publishing	286	pinkfloyd	Musician_band	346	Avicii	Musician_band
227	CBSNews	Media_news_publishing	287	JLo	Musician_band	347	KeithUrban	Musician_band
228	latimes	Media_news_publishing	288	BonJovi	Musician_band	348	joejonas	Musician_band
229	Reuters	Media_news_publishing	289	wizkhalifa	Musician_band	349	AlejandroSanz	Musician_band
230	NoticiasCaracol	Media_news_publishing	290	jtimberlake	Musician_band	350	greysonchance	Musician_band
231	msnbc	Media_news_publishing	291	Nirvana	Musician_band	351	PAULVANDYK	Musician_band
232	CNNMoney	Media_news_publishing	292	maroon5	Musician_band	352	edsheeran	Musician_band
233	nerdist	Media_news_publishing	293	Pink	Musician_band	353	christinaperri	Musician_band
234	bbcgoodfood	Media_news_publishing	294	evanescence	Musician_band	354	luansantana	Musician_band
235	verge	Media_news_publishing	295	ddlovato	Musician_band	355	ClaudiaLeitte	Musician_band
236	arstechnica	Media_news_publishing	296	Nickelback	Musician_band	356	morissette	Musician_band
237	BostonGlobe	Media_news_publishing	297	systemofadown	Musician_band	357	CalvinHarris	Musician_band
238	20m	Media_news_publishing	298	daddy_yankee	Musician_band	358	johnlegend	Musician_band
239	intelligence2	Media_news_publishing	299	TreySongz	Musician_band	359	chickenfootjoe	Musician_band
240	allthingsd	Media_news_publishing	300	DONOMAR	Musician_band	360	Anahi	Musician_band

Table 8: User List 3

#	TW_id	fb_category	#	TW_id	fb_category	#	TW_id	fb_category
361	RamazzottiEros	Musician_band	421	David_Cameron	Politician	481	jamieoliver	Public_figure
362	steveaoki	Musician_band	422	JerryBrownGov	Politician	482	rodrigovesgo	Public_figure
363	JoshGrobanNews	Musician_band	423	GavinNewsom	Politician	483	LaurenConrad	Public_figure
364	paurubio	Musician_band	424	marianorajoy	Politician	484	MarthaStewart	Public_figure
365	LilJon	Musician_band	425	_Rubalcaba_	Politician	485	LittlecBeadles	Public_figure
366	TravieMcCoy	Musician_band	426	tylerperry	Producer	486	eddieizzard	Public_figure
367	TizianoFerro	Musician_band	427	RayWJ	Producer	487	PhillyD	Public_figure
368	LittleMixOffic	Musician_band	428	YouTube	Product_service	488	sirdickbranson	Public_figure
369	karminmusic	Musician_band	429	PlayStation	Product_service	489	newtgingrich	Public_figure
370	dolly_parton	Musician_band	430	Skype	Product_service	490	pattiemallette	Public_figure
371	llcoolj	Musician_band	431	SamsungMobileUS	Product_service	491	DjASHBA	Public_figure
372	kaskade	Musician_band	432	intel	Product_service	492	Jess_Stam	Public_figure
373	imogenheap	Musician_band	433	firefox	Product_service	493	Calle13Oficial	Public_figure
374	chamillionaire	Musician_band	434	dcshoes	Product_service	494	edans	Public_figure
375	vidialdiano	Musician_band	435	Badoo	Product_service	495	Ze_Frank	Public_figure
376	pauloakenfold	Musician_band	436	AppStore	Product_service	496	gbiffle	Public_figure
377	sherinamunaf	Musician_band	437	MediaFire	Product_service	497	TheSharkDaymond	Public_figure
378	PretaGil	Musician_band	438	Marvel	Product_service	498	Yunus_Centre	Public_figure
379	Fiuk	Musician_band	439	Sony	Product_service	499	april_summerz	Public_figure
380	kinagrannis	Musician_band	440	havaianas	Product_service	500	Plaid_Page	Public_figure
381	paulwallbaby	Musician_band	441	TOMS	Product_service	501	gailsimmons	Public_figure
382	yokoono	Musician_band	442	Dropbox	Product_service	502	violetblue	Public_figure
383	sammyhagar	Musician_band	443	travelocity	Product_service	503	MariaBartiromo	Public_figure
384	petewentz	Musician_band	444	engadget	Product_service	504	annecurtissmith	Public_figure
385	perryfarrell	Musician_band	445	BuzzFeed	Product_service	505	TreyRatcliff	Public_figure
386	terranaomi	Musician_band	446	earthoutreach	Product_service	506	nishjamvwal	Public_figure
387	RealMadAnthony	Musician_band	447	GooglePlay	Product_service	507	randomhouseau	Publisher
388	afgansyah_reza	Musician_band	448	googledrive	Product_service	508	iHeartRadio	Radio_station
389	gilbertogil	Musician_band	449	echofon	Product_service	509	RyanSeacrest	Radio_station
390	YellaBoy	Musician_band	450	realmadrid	Professional_sports_team	510	Armada	Record_label
391	dariamusk	Musician_band	451	chelseafc	Professional_sports_team	511	SUBWAY	Restaurant_cafe
392	kompascom	News_media_website	452	Lakers	Professional_sports_team	512	Walmart	Retail_and_consumer_m
393	portalR7	News_media_website	453	acmilan	Professional_sports_team	513	Target	Retail_and_consumer_m
394	Estadao	News_media_website	454	Arsenal	Professional_sports_team	514	amazon	Retail_and_consumer_m
395	Slate	News_media_website	455	LFC	Professional_sports_team	515	Macys	Retail_and_consumer_m
396	la_patilla	News_media_website	456	chicagobulls	Professional_sports_team	516	Forever21	Retail_and_consumer_m
397	rtve	News_media_website	457	GalatasaraySK	Professional_sports_team	517	BestBuy	Retail_and_consumer_m
398	CBSSports	News_media_website	458	MiamiHEAT	Professional_sports_team	518	neimanmarcus	Retail_and_consumer_m
399	firstpostin	News_media_website	459	Indiancrickteam	Professional_sports_team	519	muyinteresante	Science_website
400	big_picture	News_media_website	460	celtics	Professional_sports_team	520	LaughingSquid	Society_culture_website
401	NewsHour	News_media_website	461	Yankees	Professional_sports_team	521	googlechrome	Software
402	sengineland	News_media_website	462	MLB	Professional_sports_team	522	nimbuzz	Software
403	VOA_News	News_media_website	463	RedSox	Professional_sports_team	523	NBA	Sports_league
404	detikcom	News_media_website	464	nyknicks	Professional_sports_team	524	ufc	Sports_league
405	TheAcademy	Non	465	Giants	Professional_sports_team	525	nfl	Sports_league
406	marchmadness	Non	466	dallasmavs	Professional_sports_team	526	ATPWorldTour	Sports_league
407	carrolltrust	Non	467	okcthunder	Professional_sports_team	527	MLS	Sports_league
408	MotherJones	Non	468	SFGiants	Professional_sports_team	528	tatadocomo	Telecommunication
409	TEDx	Non	469	santosfc	Professional_sports_team	529	hotel_urbano	Travel_leisure
410	davos	Non	470	BrooklynNets	Professional_sports_team	530	KLM	Travel_leisure
411	oxfamgb	Non	471	pdxtrailblazers	Professional_sports_team	531	CarnivalCruise	Travel_leisure
412	EFF	Non	472	roushfenway	Professional_sports_team	532	EmiratesAirLDN	Travel_leisure
413	TheGPP	Non	473	KimKardashian	Public_figure	533	bookingcom	Travel_leisure
414	AbuDhabiFF	Non	474	annoyingorange	Public_figure	534	hotelsdotcom	Travel_leisure
415	twitter_es	Non	475	DalaiLama	Public_figure	535	British_Airways	Travel_leisure
416	Olympics	Organization	476	iamsrk	Public_figure	536	AmericanAir	Travel_leisure
417	MOCAlosangeles	Organization	477	BillGates	Public_figure	537	Delta	Travel_leisure
418	FragileOasis	Organization	478	tyrabanks	Public_figure	538	alpharooms	Travel_leisure
419	JornalOGlobo	Organization	479	mscristinereyes	Public_figure	539	TravelRepublic	Travel_leisure
420	BarackObama	Politician	480	CHRISDJMOYLES	Public_figure	540	AnimalPlanet	Tv_channel

Table 9: User List 4

#	TW_id	fb_category	#	TW_id	fb_category
541	multishow	Tv_channel	601	Yahoo	Website
542	redetelecine	Tv_channel	602	eBay	Website
543	ntvspor	Tv_channel	603	UOLEsporte	Website
544	rede_globo	Tv_channel	604	UOL	Website
545	FRANCE24	Tv_channel	605	deviantART	Website
546	NBCSports	Tv_channel	606	NETAPORTER	Website
547	MTV	Tv_network	607	HuffingtonPost	Website
548	cartoonnetwork	Tv_network	608	UEFAcom	Website
549	Discovery	Tv_network	609	Forbes	Website
550	HISTORY	Tv_network	610	googlemaps	Website
551	NickelodeonTV	Tv_network	611	TripAdvisor	Website
552	HBO	Tv_network	612	UberFacts	Website
553	FoodNetwork	Tv_network	613	googlestudents	Website
554	AJEnglish	Tv_network	614	damnitstrue	Website
555	PBS	Tv_network	615	ModernMom	Website
556	ABC	Tv_network	616	rosana	Writer
557	travelchannel	Tv_network			
558	Bravotv	Tv_network			
559	TheSimpsons	Tv_show			
560	FamilyGuyonFOX	Tv_show			
561	SouthPark	Tv_show			
562	HOUSEonFOX	Tv_show			
563	SpongeBob	Tv_show			
564	TwoHalfMen_CBS	Tv_show			
565	BigBang_CBS	Tv_show			
566	MeetatMacLarens	Tv_show			
567	GLEEonFOX	Tv_show			
568	iCarly	Tv_show			
569	AmericanDadFOX	Tv_show			
570	NCIS_CBS	Tv_show			
571	CSIMiami_CBS	Tv_show			
572	vampirediaries	Tv_show			
573	CrimMinds_CBS	Tv_show			
574	SHO_Dexter	Tv_show			
575	gossipgirl	Tv_show			
576	BONESonFOX	Tv_show			
577	cw_supernatural	Tv_show			
578	theofficenbc	Tv_show			
579	dahSyatMusik	Tv_show			
580	ModernFam	Tv_show			
581	SportsCenter	Tv_show			
582	106andpark	Tv_show			
583	ConanOBrien	Tv_show			
584	TheXFactorUSA	Tv_show			
585	GreysABC	Tv_show			
586	nbcagt	Tv_show			
587	GMA	Tv_show			
588	RachaelRayShow	Tv_show			
589	Legendarios	Tv_show			
590	TheMandyMoore	Tv_show			
591	AC360	Tv_show			
592	Eurovision	Tv_show			
593	JimmyKimmelLive	Tv_show			
594	CBSThisMorning	Tv_show			
595	kingsthings	Tv_show			
596	katiecouric	Tv_show			
597	RickiLake	Tv_show			
598	BBCClick	Tv_show			
599	FaceTheNation	Tv_show			
600	google	Website			