Characterizing Twitter Users: What do Samantha Cristoforetti, Barack Obama and Britney Spears Have in Common?

Characterizing Twitter Users:

What do Samantha Cristoforetti, Barack Obama and Britney Spears Have in Common?

Alessia Antelmi
Dipartimento di Informatica
Università degli Studi di Salerno
Fisciano, Italy
aless.antelmi@gmail.com

Delfina Malandrino
Dipartimento di Informatica
Università degli Studi di Salerno
Fisciano, Italy
dmalandrino@unisa.it

Vittorio Scarano
Dipartimento di Informatica
Università degli Studi di Salerno
Fisciano, Italy
vitsca@unisa.it

***Abstract*—The exponential growth in the use of digital devices and the ubiquitous online access produce a huge amount of structured and unstructured data that can be mined and analyzed to gather insights into several domains. In particular, since the advent of Web 2.0, Online Social Networks (OSNs) represent a rich opportunity for researchers to collect real user data and to explore OSNs users behavior. This study represents a first attempt to characterize and classify OSNs users according to their level of activity through the use of user profile attributes. We analyzed four case studies from the Twitter platform for a final total of around 721 thousand users, divided into four sub-datasets and examined over a period of at least six months in 2017. Following a data-driven methodology, we found that static, *profile-based information* - based on the entire lifetime of the users - can help to recognize users *influence* in Twitter online communities. On the other hand, these profile attributes are not enough to characterize user *activity* on the microblogging platform.**

***Index Terms*—User Behavior, Role Discovery, Online Social Networks, Data-driven analysis**

I. Introduction

The ever-increasing use of Internet - due to the exponential growth in the use of digital devices and the ubiquitous online access - produces a huge amount of structured and unstructured data that can be mined and analyzed to gather insights into many domains, such as health, education and marketing [1]. In this context, online social networks (OSNs) represent a rich opportunity for researchers to collect real user data and to explore and study OSNs user behavior. Jin et al. in their survey [2] define OSNs user behavior as identified by the various social activities that users can do online - friendship creation, content publishing, messaging and commenting - and describe how its characterization can be useful to different Internet entities in several aspects. Among these, they cite OSN service providers, interested in understanding their customers' attitudes to generate advertisement placement policies and OSN users whose behavior study is important to enhance their usability experience and to inform a better site design. Furthermore, accurate models of user behavior in OSNs are crucial in social studies and viral marketing [3]. Since the advent of Web 2.0, considered as the platform for the evolution of Social Media, researchers focused their attention on the visible activities done by the users of OSNs to characterize their behavior and investigate their online social roles. During the last decade, however, their attention is turning to a more challenging task: the analysis and characterization of *passive* users behaviors. In 2006, the expert of User Interface Jakob Nielsen introduced his 90-9-1 rule[1] of thumb, describing an online community as composed by 90% of lurkers, 9% by users who contribute from time to time, and only 1% by users who account for most contributions. This means that multi-user communities and OSNs share one property: most users do not participate and simply lurk in the background. Due to clear privacy reasons (e.g. the lack of log data), the identification of this user category, i.e., lurkers, is not a trivial task and the definition of the concept of *activity* itself is not straightforward. The problem lies in how we can determine several activity levels through the study of users' behavioral patterns in OSNs.

Our study represents a first attempt to characterize OSNs users according to their level of activity, (i) using a data-driven approach and (ii) focusing on both active and passive users. As a case study, we chose the microblogging platform Twitter because of the easiness of silent information consumption and the openness of its APIs. By analyzing around 721 thousand Twitter users divided into four sub-datasets, we investigate whether the data collected from the Twitter profile of the users are able to capture their activity level and the general concept of activity on the social platform. Our main result can be summarized as follows:

- Profile-based information are useful in characterizing users *influence* level and highlighting *famous* accounts in a Twitter online community.

II. Related Work

Rossi et al. [4] provide an overview of role discovery approaches, discussing graph-based, feature-based and node-attribute data. They give evidence that feature-based roles are more flexible and capable of capturing more complex roles. After introducing a framework for feature-based role discovery, they discuss classes of approaches for role feature construction and role assignment. In particular, they survey two main categories of methods for assigning roles using the graph-based feature representation: low-rank approximation techniques and role clustering methods. Revelle et al. [5] use

[1] https://www.nngroup.com/articles/participation-inequality

non-negative matrix factorization (NMF) for discovering node roles: they decompose independently several node-attribute matrices, each describing a network snapshot, using NMF. The resulting basis vectors correspond to node roles in the network. Cheng et al. [6] use role-labeled nodes to identify the roles of unlabeled ones. In this work, roles are defined in terms of functional occupations in an organization - that is, in terms of domain knowledge. A feature-based approach is presented by White et al. [7]: nodes are clustered using node features derived from the network structure and then a qualitative assignment of nodes to roles follows. Another feature-based approach for automatic detection of user roles in online forums is presented by Chan et al. [8]. They use principal component analysis and agglomerative clustering of feature profile data to find roles, where each cluster corresponds to a role. Maia et al. [9] investigate typical user behaviors and present a methodology to identify characteristics that define a user as part of a group that share the same overall behavior using K-means clustering. Based on this work, O'Donovan et al. [10] group users according to features of their behavior on Facebook, and then review the demographic and psychological data associated with each cluster to interpret the inferred formal roles. Our work follows the methodology described in [9], [10], but it differs fundamentally because we focus on the characterization of active and passive behaviors exhibited by the users other than analyzing a different social platform.

III. Dataset

To build a dataset as heterogeneous as possible, we collected four sub-datasets of users: a random sample of Twitter users and three groups representing as many Twitter communities. It is worth noting that we consider a *Twitter community as a set of Twitter users who share a common interest* (e.g., some followers of a TV series' Twitter account), motivated by the research of Java et al. [11]. After the data collection step, we cleaned our dataset from spammers and churners.

A. Dataset Construction

1) Random sub-dataset construction: due to the absence of Twitter APIs specifically designed to handle this task, we performed the following steps in order to construct our dataset.

1) We collected a tweets sample through the Twitter streaming APIs, from which we extracted 241,000 unique users (*starting set*). We could not stop our collecting process at this point because this would have meant working on a biased dataset where all users had posted at least one status update[2].
2) To reduce the bias introduced in the first step, we randomly selected 200 users from the *starting set*, and we collected all their followers (*followers set*).
3) Then, we randomly picked 50 from the *followers set*. This set of 50 users will represent the *seed* of our dataset.
4) For each user in the *seed set*, we randomly chose up to 60 his/her followers, adding them to the final dataset (*extended seed set*) and obtaining around 3,000 users.
5) For each user in the *extended seed set* we repeated step (4), ending up with around 180,000 users.
6) Finally, we removed all users with a private account, obtaining 131,301 unique users (*final set*).

In order to carry out our study, we collected the profile snapshots of the users in the *final set* at regular time intervals from August 3rd to December 3rd, 2017.

2) Communities sub-datasets construction: we considered three Twitter fans communities characterized by a very different nature - one related to the TV show Game of Thrones (GoT), the other to the British rock band Coldplay and the last one to the National Geographic (NatGeo) Magazine. We chose the GoT community due to the fact that the Game of Thrones TV show has established a reputation for being widely discussed on social media channels, specifically on the Twitter platform. We picked a Coldplay community for a similar reason: with a gross of 523 million dollars and 5.39 million fans attending their tour in 2017, the pop/rock band Coldplay is one of the most famous in the last decade. In both cases, Twitter plays the role of an import medium to engage fans. The NatGeo community was chosen for its potential broader topics discussed in contrast with the first two sub-datasets and, consequently, for its very diverse public. We retrieved the complete followers list of the GoT Twitter account and then randomly picked 350,000 users from among them. We collected at regular time intervals the profile data from 300,332 users among this random group whose profile was publicly shared over a 3-months timeframe from September 3rd to December 3rd, 2017. We followed the same process to analyze a random subset of the Coldplay official Twitter account's followers (obtaining 290,237 users) and a random subset of the NatGeo official Twitter account's followers (considering 314,610 users).

B. Removing spammers and churners

A limitation of the above dataset is that it may include accounts identified as spammers and *churners*, typically defined as users who do not post for a significantly long period of time [12]. Both these typologies of users add noise and bias in our analysis and we need to remove them from our dataset. To filter out users belonging to the first group, we identified all accounts marked as spammers by the Twitter Support Team. However, pointing out the users from the second category is not straightforward: due to the limited access to Twitter users' data, it is impossible to know who exactly is a churner. To deal with this problem, we used a similar strategy described in [13]: if a user did no action during the whole period of observation, then we considered him/her as a churner and we definitely removed him/her from our dataset. It is worthwhile highlighting that we consider the following activities as an action: posting a new status update (tweet, retweet, quote or comment), tapping a like, following or unfollowing a new account, modifying one's own profile data (screen name, description, location, image or banner image). Table I summarizes the number of unique users obtained per sub-dataset at the end of the cleaning process.

[2]Twitter does not reveal how the samples are generated.

IV. Data Analysis

In this section, we detail the features analyzed and the clustering methodology used to look for behavioral patterns within the dataset.

A. Clustering Features

We represent each user with a unidimensional features vector, containing the following information extracted from the Twitter profile of the referred user:

i. **P-Likes**, the number of tweets the user has liked in the account's lifetime;
ii. **P-Statuses**, the number of tweets (including retweets) issued by the user;
iii. **P-Lists**, the number of public lists that the user is a member of;
vi. **P-Friends**, the number of users the account is following;
v. **P-Followers**, the number of followers the account currently has.

These features represent a static snapshot of a user's profile based on his/her *whole* history of interaction.

B. Methodology

Our main goal is to characterize the users according to their level of activity and to group them according to their behavioral patterns through a data-driven approach, without attempting to match behaviors to a pre-defined set of roles. In this work, we follow a methodology similar to the one described in previous studies [9], [10], using the clustering algorithm K-means and the Euclidean distance as the distance measure. The K-means algorithm is a common data-mining approach to extract grouping patterns without any prior knowledge of their characteristics; furthermore, its scalability allows the analysis of datasets with thousands of elements. Before clustering our data, we first standardized the feature values by removing the mean and scaling to unit variance. We then applied the K-means clustering algorithm to these new feature vectors, using the K-means implementation from the Python library *scikit-learn*. We derived the optimal number of clusters using the Silhouette index, a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). To study the evolution of the user activity over time, we applied the K-means algorithm to each snapshot of the dataset, where a snapshot corresponds to the set of user profiles collected in a given time. We analyzed the differences among the clusters produced partitioning two consecutive snapshots and the associated *migration* patterns.

In this work, we focus only on the partition obtained clustering the datasets using feature values from the last snapshot of the users' profile as no interesting migration patterns have been found. Furthermore, the partitions obtained from the different snapshots were almost the same. The clustering methodology described has been run independently on all four sub-datasets.

V. Results

In the first part of this Section, we describe the cluster analysis performed to identify behavioral patterns, in terms of activity level among Twitter users. We then explore the composition of some of the clusters obtained through a qualitative analysis by examining the users within them.

A. Latent Roles

Based on the Silhouette values, we found that 4 was the best number of clusters for each sub-dataset. To confirm the significance of the cluster distributions within a sub-dataset, for each feature and for each pair of clusters we applied the ANOVA and the Tukey's HSD tests, to check whether groups in the same sample statistically differ (ANOVA) and which of them in specific have significance differences (Tukey's test).

The number of users within each cluster are shown in Table I. Although each sub-dataset has been clustered independently from the others, all partitions exhibit the same structure. There is: (i) one larger group containing the majority of the users (*Passive* clusters) and (ii) at least another smaller cluster containing less than 10 users. The similarity in terms of the sizes of these use cases suggests that it may be possible to find a similarity in terms of behavioral patterns regardless of the nature of each community. In Figure 1 - where each subfigure refers to a single use case - we plotted the scaled feature values for each cluster centroids. As clearly visible from these charts, all datasets manifest an analogous cluster composition, with the exception of the Coldplay community.

· **Passive**. In every sub-dataset, all cluster centroids referring to the largest cluster (i.e., $Cluster_0$) have value 0. This result means that the users within this group have done few or none actions on the social platform. In other words, this cluster should contain all **passive** or **low-engagement** users.

· **Prosumers**[3]. The second role shared across all use cases (see Cluster_1) is characterized by a moderate amount of statuses and likes and very small values for the remaining features. These users tend to be both content **producers** when posting a new status update, and content **consumers** when liking someone's else post.

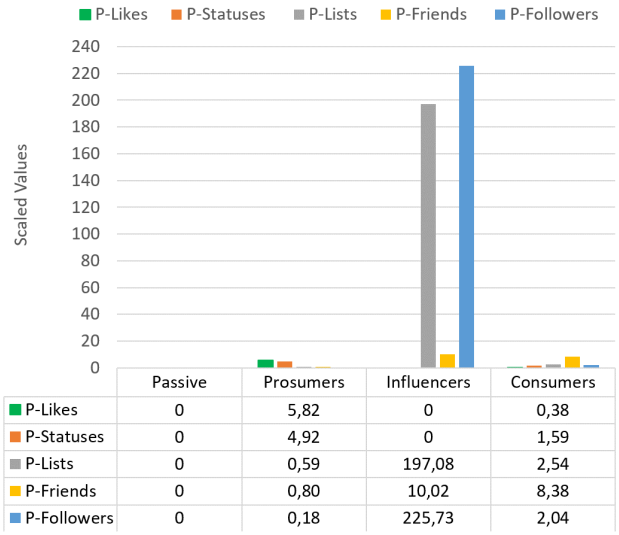
· **Influencers**. All sub-datasets, but the Coldplay community, reveal a third common role (see Cluster_2), defined by a prevalence of two features: *P-Lists*, i.e., the number of lists a user is member of, and *P-Followers*, i.e., the number of followers the user currently has. These features are two typical and basic popularity metrics within the Twitter world, suggesting that the users contained in this group might be the so-called **influencers**.

· **Consumers**. The last role identified and clearly noticeable in Figures 1b and 1c (see Cluster_4) is characterized by a high value of the feature *P-Friends*, indicating the number of accounts the user is following. This kind of users have a behavior **consumer-like** because they have the possibility

[3] *Prosumer* is a portmanteau word - i.e., a linguistic blend of words - coined in 1980 by the American futurist Alvin Toffler from the words *producer* and *consumer*. This term refers to a person who consumes and produces a product.

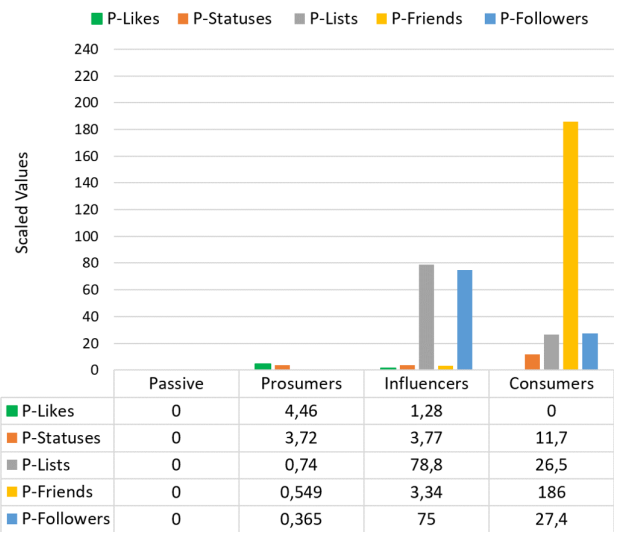
TABLE I: Number of users in each cluster per sub-dataset

Dataset	Number of unique users	$Cluster_0$ (Passive)	$Cluster_1$ (Prosumers)	$Cluster_2$ (Influencers)	$Cluster_3$ (Consumers)
Random	122,894	120,551	1,375	2	966
GoT	215,864	211,734	4,107	19	4
NatGeo	169,385	168,612	760	11	2
Coldplay	212,757	212,318	415	16	8 (*Prosumer-like*)



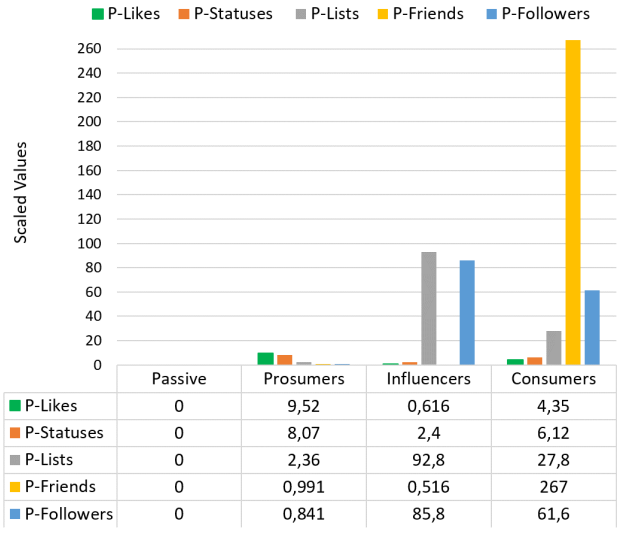
	Passive	Prosumers	Influencers	Consumers
P-Likes	0	5,82	0	0,38
P-Statuses	0	4,92	0	1,59
P-Lists	0	0,59	197,08	2,54
P-Friends	0	0,80	10,02	8,38
P-Followers	0	0,18	225,73	2,04

(a) Random dataset



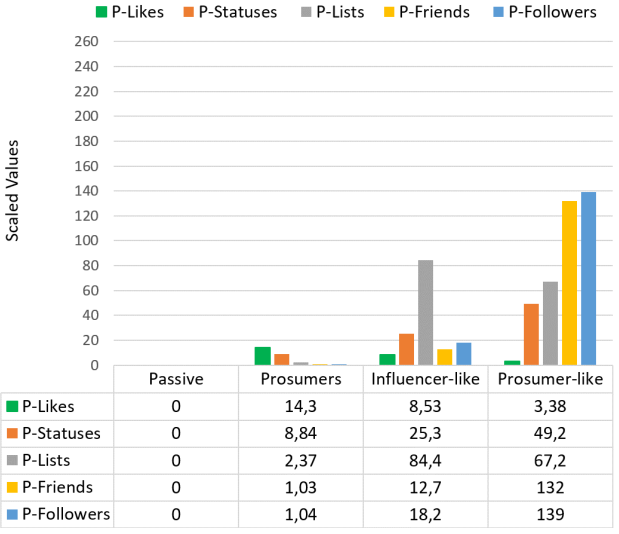
	Passive	Prosumers	Influencers	Consumers
P-Likes	0	4,46	1,28	0
P-Statuses	0	3,72	3,77	11,7
P-Lists	0	0,74	78,8	26,5
P-Friends	0	0,549	3,34	186
P-Followers	0	0,365	75	27,4

(b) Game of Thrones dataset



	Passive	Prosumers	Influencers	Consumers
P-Likes	0	9,52	0,616	4,35
P-Statuses	0	8,07	2,4	6,12
P-Lists	0	2,36	92,8	27,8
P-Friends	0	0,991	0,516	267
P-Followers	0	0,841	85,8	61,6

(c) National Geographic dataset



	Passive	Prosumers	Influencer-like	Prosumer-like
P-Likes	0	14,3	8,53	3,38
P-Statuses	0	8,84	25,3	49,2
P-Lists	0	2,37	84,4	67,2
P-Friends	0	1,03	12,7	132
P-Followers	0	1,04	18,2	139

(d) Coldplay dataset

Fig. 1: Cluster centroids

to gather information from several accounts. Although with a very different magnitude, also the Random dataset exhibit this role (Figure 1a).

The clustering analysis run on the Coldplay dataset highlights a slightly different picture from the ones just described (see Figure 1d).

· **Influencer-like**. *$Cluster_2$* describes users with a balanced set of features with a peak for the *P-Lists* variable. These values suggest a more uniform interaction pattern on Twitter, in terms of content posted and interaction with other users. However, the high value related to the *P-Lists* variable let us surmise that the content posted by these users is of interest on the social

platform.

· **Prosumer-like**. *Cluster*$_3$ characterizes users with high values for the social relationship features, namely *P-Friends* and *P-Followers*. This group seems to be both producers - given the significant high value of the features *P-Statuses* and *P-Lists* - and consumers - given the considerable number of friends.

To summarize, we have identified four main distinct behaviors - plus two extra found in the Coldplay dataset - in which Twitter users can be classified based on their profile attributes. However, it is worthwhile noting that some of these roles describe only a small fraction of the overall population analyzed. We will discuss these outcomes in Section VI.

B. Qualitative Analysis

The modest dimension of some of the clusters obtained allowed us to analyze under a qualitative lens these groups. In particular, we focused on the **Influencers** clusters: to validate our hypothesis whether it contains real influencers or not, we manually checked the account of the users labeled with this role. We analyzed this category in the Random, GoT and NatGeo datasets, without considering the Coldplay use case. Table II reports the results of the qualitative analysis for the Random dataset. In this cluster, we found well-known personalities, namely *Barack Obama* and *Britney Spears*. They are the 3^{rd} and the 14^{th} accounts most followed on Twitter[4] and both can be definitely defined as popular accounts. What distinguishes both users is not only the huge amount of lists they are member of and the number of their followers, as previously described but also the greater number of statuses with respect to the number of likes - especially for the account of Barack Obama. These characteristics suggest that this typology of users interact in a *unidirectional* way with the Twitter platform, meaning that they only share information related to their own person on the microblog and do not interact with other accounts as other activity. For space reason, we do not present the detailed description of all the *Influencers* analyzed in the GoT and NatGeo datasets. However, as in the Random case, all these users have a verified account, even though some of them act as fan-page accounts rather than referring to real people. In the GoT dataset, examples of *fan-page* accounts found are the official account of the American technology company *NVIDIA* and the official account of the TV show of the American comedian television host Jimmy Fallon (*Fallon's Tonight*). Examples of real people's accounts from the same dataset are the American comic-book writer and actor *Stan Lee*, the American DJ *Steve Aoki* and the American actor *Dylan Sprouse*. As for Barack Obama and Britney Spears, all these accounts - independently from their nature of fan-page or not - are characterized by a high number of lists they are member of, a significant amount of followers and a greater number of statuses posted with respect to the number of likes. The same consideration made for Barack Obama and Britney Spears hold for this group as well. In the NatGeo dataset, examples of *fan-page* accounts found are *Greenpeace Brasil*, the Brazilian section of the non-governmental environmental organization, and *500px*, a Twitter community for photographers. An example of real people's accounts from the same dataset is the Italian astronaut *Samantha Cristoforetti*. Even though in a low measure, this group exhibits the same characteristics of the previous groups described and the consideration made before are true also in this case. It is interesting noting how many verified accounts found in the GoT and NatGeo datasets are related to the Game of Thrones TV show and to the naturalistic theme, respectively. For instance, the TV host Jimmy Fallon interviewed several GoT cast members during his TV show Fallon's Tonight, while Steve Aoki [5] used to post several status updates about the TV series. In the NatGeo dataset, we have users like Samantha Cristoforetti, Greenpeace Brasil and 500px. All of them are strictly related to the naturalistic theme on which the magazine is based, even though under different points of view.

[4]http://friendorfollow.com/twitter/most-followers/

VI. Discussion

This work represents a first attempt to characterize OSNs users according to their level of activity using attributes extracted from the users' Twitter profile and following a data-driven methodology. Clustering independently all four sub-datasets, we found that:

· The majority of the users is grouped into a single group (**Passive**), whose centroid values are 0. This means that these users exhibit a *low-engagement* behavior. However, when we explored the feature values of the users within this cluster, we discovered that a large part of them had high values for both the number of statuses and the number of likes during their entire account lifetime. Even though these users were characterized by such values, they did not share enough other distinct characteristics for the K-means algorithm to require additional clusters. Analyzing the partition obtained clustering the data using a greater value for the parameter k did not result in a very different configuration. This suggests that the selected features are not particularly efficient at discriminating the activity level and, therefore, users cannot be characterized only on the basis of their lifetime-account activity.

· The profile-based features are able to spot outliers in terms of *famous* or *influencer* accounts (**Influencers**). However, can we define these users as truly active? Recalling that the attributes *P-Lists* and *P-Followers* are typical popularity metrics and that these users have a higher number of statuses with respect to the number of likes, it seems that this typology of users tend to interact with the social in a *unidirectional* way, i.e., not to interact with their contact network, but only to publish a status update. A recent Facebook study[6] states that simply broadcasting a status is not enough to contribute to subjective well-being: people have to maintain an active engagement with their own network by sending or receiving messages and/or

[5]http://www.steveaoki.com/lifestyle/the-new-game-of-thrones-trailer-is-here-and-were-pumped/

[6]https://newsroom.fb.com/news/2017/12/hard-questions-is-spending-time-on-social-media-bad-for-us

TABLE II: Feature values describing the Influencers contained in the Random dataset

Account	Account Description	P-Statuses	P-Likes	P-Lists	P-Friends	P-Followers
Barack Obama	44th President of the United States of America	15482	10	227234	626350	97418015
Britney Spears	American singer, dancer and actress	5333	2461	122239	389238	56144590

comments to have positive benefits. According to this work, *influencer* accounts might not be considered truly active users.
· Based on the previous consideration, the only active group detected by the K-means algorithm is the *producers/consumers* group (**Prosumers**). During their account lifetime, these users have done a comparable amount of tweets/retweets and likes, interacting not only with the platform (tweets) but also with their contacts network (retweets, likes). This role - together with the **Consumers** one - are consistent with the outcomes presented in previous studies [9].
· The Coldplay community represents an exception to the behavioral patterns found in the remaining sub-datasets. For what concerns the **Influencer-like** role, we hypothesize that it will evolve towards the *Influencers* role observed in the other use cases. A possible explanation about the absence of *real influencers* in this dataset could be the short account lifetime of the majority of the Coldplay users analyzed. By verifying the subscription date of these users, we found out that many of them signed up the first time to Twitter during Summer 2017 in correspondence of both the benefit concert One Love Manchester held on 4th June 2017 and the USA A Head Full of Dreams tour of the Coldplay band scheduled during the whole Summer 2017. In regards to the **Prosumer-like** role, by observing the feature value of this group, we surmise that these users will change their behavior towards a more *consumer-like* attitude. In future work, we aim to verify it by collecting and applying the same methodology to the new users' profile data.

From these considerations follows that the selected features are helpful to spot outliers in terms of *popular* and *influencer* accounts. Through our methodology, we are able to characterize this typology of users, defined by a high number of followers, lists they are member of and more statuses with respect to their likes. On the other hand, they fail to capture the activity patterns happening on the social platform and they are not a good behavior discriminator as they result in one single large group and in a very unbalanced partition. Improving our understanding about *participation inequality* in OSNs can be useful not only to traditional application domains, such as recommender systems for advertisements in OSNs or the study of new marketing strategies, but also to deepen our knowledge about how subjective well-being is related to the increasing use of social media [14], in order to face problems like bullying and online harassers.

VII. Conclusion

Improving our understanding in how people interact and express themselves on social media and how this is translated into different behavioral patterns is helpful not only to traditional application domains, but also to figure out in what extent subjective well-being is related to the increasing use of such platforms. In this work, we showed that the static *profile-based* features, based on the entire account lifetime of the users, are able to recognize users *influence* in Twitter online communities. As future work, we aim to further explore the characterization of the OSNs user activity by exploiting the dynamic features deriving from the direct observation of the actions (in terms of tweets, retweets, quotes, replies and likes) posted over time by the same set of users analyzed in this study. This may enable us to compare the effectiveness in describing active and passive behaviors of the OSNs users through the use of different sets of features. Another interesting line of inquiry is to keep monitoring the evolution of the Coldplay community to verify if the behavioral patterns found in the other datasets will also emerge in this community.

References

[1] M. Salampasis, G. Paltoglou, A. Gianchanou, Using social media for continuous monitoring and mining of consumer behaviour, Int. J. Electron. Bus. 11 (1) (2013) 85-96.

[2] J. Long, C. Yang, W. Tianyi, H. Pan and V. Athanasios. (2013). Understanding user behavior in online social networks: A survey. IEEE Communications Magazine, 51(9), 144-150. Communications Magazine, IEEE. 51. 144-150.

[3] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. 2009. Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement (IMC '09). ACM, New York, NY, USA, 49-62.

[4] R. Rossi, N.K. Ahmed, et al. *Role discovery in networks.* Knowledge and Data Engineering, IEEE Transactions on, 27(4):1112-1131, 2015.

[5] Revelle M., Domeniconi C., Johri A. (2016) *Persistent Roles in Online Social Networks.* In: Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016. Lecture Notes in Computer Science.

[6] Y. Cheng, A. Agrawal, A. Choudhary, H. Liu, and T. Zhang. *Social role identification via dual uncertainty minimization regularization.* In Data Mining (ICDM), 2014 IEEE International Conference, pp 767-772, IEEE, 2014.

[7] A. J. White, J. Chan, C. Hayes, and B. Murphy. *Mixed membership models for exploring user roles in online fora*. In Proceedings of the 6th International Conference on Web and Social Media, 2012.

[8] J. Chan, C. Hayes, and E. M. Daly. *Decomposing discussion forums and boards using user roles*. In Proceedings of the 4th International Conference on Web and Social Media, volume 10, pages 215-218, 2010.

[9] Maia, Marcelo & Almeida, Jussara & Almeida, Virgilio. (2008). Identifying User Behavior in Online Social Networks. Proceedings of the 1st Workshop on Social Network Systems.

[10] O'Donovan, F.T., Fournelle, C., Gaffigan, S., Brdiczka, O., Shen, J., Liu, J., & Moore, K.E. (2013). Characterizing user behavior and information propagation on a social multimedia network. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 1-6.

[11] Java, Akshay & Song, Xiaodan & Finin, Tim & Tseng, Belle. (2007). Why we Twitter: Understanding microblogging usage and communities. of the 9th WebKDD and 1st SNA. 43. 56-65.

[12] Oentaryo, Richard & Lim, Ee-Peng & Lo, David & Zhu, Feida & Prasetyo, Philips Kokoh. (2012). Collective Churn Prediction in Social Network. 210-214. 10.1109/ASONAM.2012.44.

[13] W. Gong, E. Lim, F. Zhu. Characterizing Silent Users in Social Media Communities. In Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM), 2015.

[14] Verduyn P, Ybarra O, Resibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Iss Policy Rev 2017 Jan 13;11(1):274-302.