Multi-scaling Analysis of Social Network Users' Profiles

Paulo Salvador, António Nogueira

Instituto de Telecomunicações, Aveiro, Portugal
DETI, University of Aveiro, Portugal
E-mails: salvador@ua.pt, nogueira@ua.pt

***Abstract*—Online social networks are extremely popular services on the Internet, having a relevant impact on the social, political and economical aspects of our daily lives. Being able to characterize social network users based on their temporal behavior and activity profiles can be useful to several engineering, management and design tasks. This paper characterizes the activity profiles of Facebook users, as well as their time evolution, based on data extracted using the Facebook Graph API and corresponding to social networking activity of 200 users during the years of 2011 and 2012. Using multiscale analysis based on a wavelet decomposition of the dataset, the proposed methodology is able to identify three main periodicity trends on the activity profiles: one of the identified profiles exhibits a visible periodicity around the 24 hours time scale (corresponding to users that have a daily interaction with Facebook), a second profile exhibits relevant components in the complete frequency range (intensive interaction) and a third profile includes important low periodicity components (users that tipically use Facebook in time periods higher than 24 hours). By looking at the temporal evolution of the different activity profiles, it is also possible to understand the dynamics of the profile changes.**

***Keywords*–Facebook users activity; user profile; scalogram; clustering.**

I. Introduction

Some Online Social Networks (OSN), such as Facebook, Twitter or Google+, are among the most popular services on the Internet, enabling individual users to connect to other participating users, share and find contents, and disseminate information through the network, while collecting the reactions from other users without any constraints, either geographic or temporal. Several OSN sites provide social links, like for example networks of professionals and contacts (e.g., LinkedIn, Facebook, MySpace) and networks for sharing contents (e.g., Flickr, YouTube). People access OSNs using both traditional personal computers and new emerging mobile devices. So, these services are really having a deep impact on the social, political and economical aspects of our daily lives.

The success of a social network depends on the behaviour of its users. Characterizing users based on their temporal behavior and activity profiles can be useful in many ways: studies of user behavior will allow to evaluate the performance of existing systems, leading to better site designs [1][2] and advertisement placement policies [3]; social studies can exploit models of user behavior and interaction, while viral marketers can use those models to spread their promotions quickly and widely [4][5]; understanding the workload of OSNs can be very important to design the next-generation Internet infrastructure and content distribution systems [6]; activity profiles can also be used for user classification or detection of anomalous behaviors, such as bots or compromised accounts [7].

Characterizing user activity in major OSNs is challenging because the distribution of most user characteristics is very skewed and shows wide variations, while the majority of participating users have a low (to moderate) level of activity. In this paper, we will characterize the activity profiles of Facebook users, as well as their time evolution, in order to understand how their behaviors change over time. The periodicity trends of the activity profiles are identified using multiscale analysis based on a wavelet decomposition of the time series. In fact, it is known that one of the main purposes of multiscaling analysis is to identify the most important time-scales of (pseudo-periodicity) activity and quantify the constancy of that pseudo-periodicity. The conducted study relies on a medium size dataset, which was extracted using the Facebook Graph API [8], containing social networking activity of 200 users during the years of 2011 and 2012.

From the conducted analysis, we were able to identify three types of activity profiles: an activity profile that exhibits a periodicity around the 24 hours time scale, a second profile that exhibits relevant components in the complete frequency range and a profile that includes important low periodicity components. In other words, the first group of users has an activity periodicity roughly equal to one day; the second group presents significant energy components at the different time scales, which means that these users interact intensively with Facebook, being almost permanently active; the third group uses Facebook in time periods that are generally higher than one day.

By looking at the temporal evolution of the different users behaviors, we could identify a global decrease on the number of users corresponding to the first profile and a steady increase on the number of users corresponding to the other two profiles. This means that the number of users that typically have daily access to Facebook is decreasing and those users tend to behave according to second and third profiles. We also tracked how users changed their profiles in the two years period of our study by calculating the number of users that changed between two specific profile types: we could verify that the first profile was actually the one who lost more users: the majority of those users (aroud 33%) changed their profiles to the second type, although around 14% of them changed their profiles to the third type.

The remaining of the paper is organized as follows: Section II presents some relevant related worksin this area; Section III briefly presents the multiscale analysis methodology that will be used to identify the pseudo-periodicity characteristics of the Facebook users' activity, as well as a summarized description of the k-means clustering algorithm; Section IV presents the social network activity analysis that was conducted, together

 ISBN: 978-1-61208-413-8

with its most relevant results; Section V presents the main conclusions and discusses some possible directions for future work.

II. Related Work

The huge amount of data generated by social media users has been used to understand their behavior. By analyzing workloads from three information networks, Guo et al. [9] showed that users' posting behavior exhibited strong daily and weekly patterns. Benevenuto et al. [10] used click-stream data from a social network aggregator to compare user behavior across different OSNs. Papagelis et al. [11] investigated the causality between individual behavior and social influence by observing the information diffusion among users. In order to understand how users' news interests change over time, Liu et al. [12] conducted a large scale analysis of Google News users click logs that were conveniently anonymized. These authors developed a Bayesian framework for predicting users' current news interests from the activities of that particular user and the news trends demonstrated in the activity of all users.

Several works have been studying the properties of user interactions in OSNs. Cha et al. [13] analyzed the spread of favorite-marking of Flickr photos, showing that social links are a primary way users find and share information in social media. Also in 2009, Valafar et al. [14] conducted a measurement study on the Flickr OSN that was able to show that only a small fraction of users in the main component of the friendship graph is responsible for the vast majority of user interactions. Wilson et al. [1] were able to show that interaction activity is significantly skewed towards a small portion of each user's social links by analyzing interaction graphs derived from Facebook user traces. Also using Facebook data, Gilbert and Karahalios [15] were able to demostrate that the "strength of ties" varies a lot, ranging from pairs of users who are best friends to pairs of users who even wished they were not friends. Viswanath et al. [16] studied the evolution of activity between Facebook users, investigating how pairs of users in a social network interact and examining how the varying patterns of interaction affect the overall structure of the activity network.

In a very interesting work, Burke et al. [17] analyzed the activities of a huge number of Facebook users in order to study the roles of user interactions. Visible actions, such as wall posts and comments, and silent actions, such as consumption of friend's content, was conveniently quantified and the authors were able to show that a typical Facebook user communicates with a small subset of their entire friendship network, although it usually maintains relationships with a larger group.

Several efforts have been made to characterize and detect malicious forms of interactions in online social networks. Gao et al. [18] the authors were able to quantify and characterize spam campaigns launched using accounts on online social networks, particularly Facebook. Benevenuto et al. [19] built a test collection of real YouTube users and classified them as spammers, promoters and legitimates, based on social and content attributes. Markines et al. [20] proposed a methodology to detect spam on social bookmarking sites, while Webb et al. [21] placed honeypot accounts on MySpace and study the captured social spammers. Grier at al. [22] presented a characterization of spam on Twitter, finding that this OSN is a highly successful platform for coercing users to visit spam pages. Vasconcelos et al. [23], authors analyzed how Foursquare users exploit the tips, dones and to-dos features to uncover different behavior profiles. This study revealed the existence of very active and influential users that seem engaged in posting tips at a large variety of venues while also receiving a great amount of user feedback on them. Besides, the paper also provided evidence of spamming, showing the existence of users that post tips whose contents are unrelated to the nature or domain of the venue where the tips were left. Wang et al. [7] proposed a machine learning approach to distinguish spam bots from normal ones in Twitter. To facilitate the spam bots detection, three graph-based features, such as the number of friends and the number of followers, are extracted to explore the unique follower and friend relationships among users.

User navigation and usage on OSN websites have also been intensively studied in the last years. Schneider et al. [24] analyzed OSN clickstream data extracted from network traffic, identifying typical user navigation patterns; Jiang et al. [25] used traces from a chinese social network to obtain statistics of profile visits, showing that silent interactions are much more prevalent and frequent than visible events, and that profile popularity are uncorrelated with the frequency of content updates. Joinson et al. [26] identified seven unique reasons for users to use Facebook: social connection, shared identities, content, social investigation, social network surfing, and status updating. Using data from Facebook, Burke et al. [2] studied user motivations for contributing to OSN sites. The authors concluded that newcomers who see their friends contributing to the OSN share more content themselves. Furthermore, those who were initially inclined to contribute, receiving feedback and having a wide audience, were also predictors of increased sharing. Chapman and Lahav [27] analyzed the ethnographical differences in the usage of OSNs. Finally, Caverlee and Webb [28] analyzed over 1.9 million MySpace profiles in order to understand who is using such large OSNs networks and how they are being used.

III. Pseudo-periodicity Analysis

The proposed pseudo-periodicity analysis of the social network users' activity is based on a wavelet decomposition through the Continuous Wavelet Transform (CWT) [29]. Let us consider that $x(t), t \in \{t_0, t_0+\Delta, \ldots, t_0+\Delta i\}, i = 1, 2, \ldots$ quantifies the number of activities performed by a user, within a social network, between $t-\Delta$ and t. Using wavelet decomposition, it is possible to analyze the social interaction process $x(t)$ in both time and frequency domains. The CWT of a process $x(t)$ can be defined as [30]:

$$\Psi_x^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{+\infty}^{-\infty} x(t)\psi^*(\frac{t-\tau}{s})dt \qquad (1)$$

where $*$ denotes the complex conjugation, $\frac{1}{\sqrt{|s|}}$ is used as an energy preservation factor, $\psi(t)$ is the mother wavelet, while τ and s are the translation and scale parameters, respectively. The first parameter is used for shifting the mother wavelet in time, while the second parameter controls the width of the window analysis and, consequently, the frequency that is being analyzed. By varying these parameters, a multiscale analysis of the entire captured process can be performed, providing a description of the different frequency components present in the decomposed process together with the time-intervals where each one of those components is located. A Wavelet Scalogram can be defined as the normalized energy $\hat{E}_x(\tau, s)$

over all possible translations (set $\mathbf{T}$) in all analyzed scales (set $\mathbf{S}$), and is computed as:

$$\hat{E}_x(\tau, s) = 100 \frac{\left|\Psi_x^{\psi}(\tau, s)\right|^2}{\sum_{\tau' \in \mathbf{T}} \sum_{s' \in \mathbf{S}} \left|\Psi_x^{\psi}(\tau', s')\right|^2} \quad (2)$$

To facilitate the analysis of the scalograms and enable the discovery of the different frequency (periodicity) components, we choose to average the normalized energy over time (set $\mathbf{T}$) obtaining the average energy at timescale s:

$$\bar{e}_x(s) = \frac{1}{|\mathbf{T}|} \sum_{\tau \in \mathbf{T}} \hat{E}_x(\tau, s), \forall s \in \mathbf{S} \quad (3)$$

where $|.|$ represents cardinality of a set. The existence of a peak in the average energy at a low timescale indicates the existence of a high-frequency (high-periodicity) component in the analyzed time-series, while a peak in the average energy at a high timescale corresponds to the existence of a low-frequency (low-periodicity) component.

In order to classify the distinct users behaviors, the unsupervised learning k-means algorithm was applied to the average normalized energies. The k-means algorithm is one of the simplest unsupervised learning algorithms that is used to solve the clustering problem [31]. The number of clusters, k, is fixed *apriori* and their centers are defined, one for each cluster. These centers should be placed in a cunning way, because different locations causes different results: the better choice is to place them as much as possible far away from each other. In our problem, we verified that three clusters were enough to classify the average normalized energies because an hypothetical forth cluster was simply a replication of one of the other clusters.

The next step of the k-means algorithm is to take each point belonging to a given dataset and associate it to the nearest center. When no point is pending, the first step is completed. At this point, we have to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. A new binding has to be done between the same dataset points and the nearest new center. An iterative approach has been generated and, as a result of this loop, we notice that the k centers change their location step by step until no more changes are done. This algorithm aims at minimizing an objective function, known as squared error function, which is given by:

$$J(\mathbf{C}_1, \ldots, \mathbf{C}_k) = \sum_{j=1}^{k} \sum_{i \in \mathbf{C}_i} \left(\|\bar{e}_i - \mu_j\|\right)^2 \quad (4)$$

where $\bar{e}_i$ is the s-dimensions data point (average normalized energies) for user i, k is the number of cluster centers, sets $\{\mathbf{C}_1, \ldots, \mathbf{C}_k\}$ contain the data points binded to each one of the k clusters, $\{\mu_1, \ldots, \mu_k\}$ is the set of (s-dimensions) cluster centers, and $\|\bar{e}_i - \mu_j\|$ is the Euclidean distance between $\bar{e}_i$ and μ_j.

IV. Social Network Activity Analysis

As as proof of concept, we used a dataset containing social networking activity (specifically, wall posting) in Facebook of a group of 200 users during 2011 and 2012. The dataset was extracted from a group of Facebook friends of the authors of this paper using the Facebook Graph Application Program Interface (API) [8]. Facebook Graph API is a low level HTTP-based API used to retrieve data from Facebook's Social Graph. Data queried using the Facebook Graph API is returned in JavaScript Object Notation (JSON) format, which can be easily post-processed to extract relevant statistics. The data was divided in four datasets, one per semester: Jan-Jun 2011, Jul-Dec 2011, Jan-Jun 2012 and, Jul-Dec 2012. We considered an interval of analysis of 15 minutes ($\Delta = 15$ minutes).

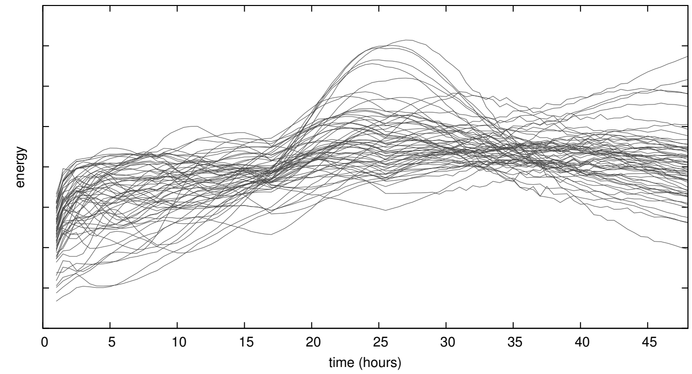

Figure 1. Pseudo-periodicity sample profiles.

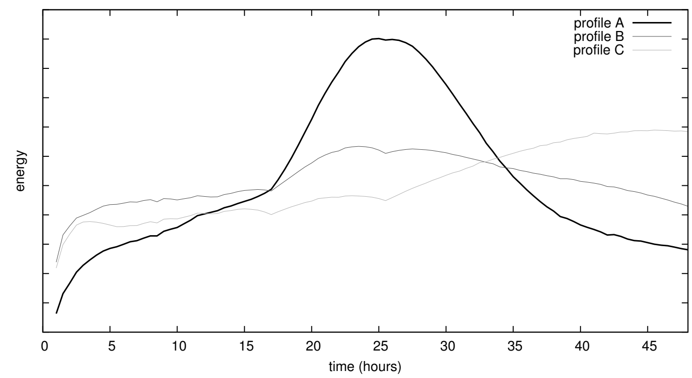

Figure 2. Cluster centroids corresponding to the first semester of 2011.

Note that this type of API usually provides well structured data, but are generally limited in terms of which data, how much data, and how often data can be retrieved. Conditions vary significantly between services: in contrast to Twitter, for example, Facebook is quite restrictive in terms of what data can be accessed, but imposes few limits on the request frequency. Besides, we know that companies retain the right to modify or close their data interfaces, which can lead to substantial problems for researchers.

A. Activity Patterns

Figure 1 shows the average normalized energy coefficients over time for some of the users included in the dataset. Although this plot contains a high number of curves, different profiles can be distinguished: some users have activity profiles that exhibit a noticeable periodicity around the 24 hours time scale (corresponding to the typical bell-shaped curve); for other users, the activity profiles exhibit relevant components in the complete frequency range, while a third group of users include important low periodicity components. The first group of users is the one that presents high energy coefficients around the 24 hours time scales, that is, the periodicity of its activity profile is

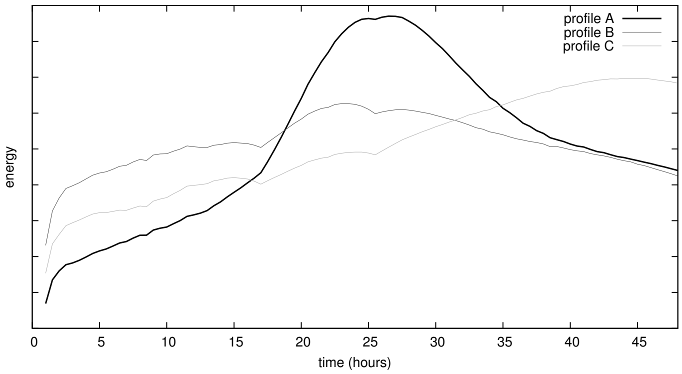

Figure 3. Cluster centroids corresponding to the second semester of 2011.

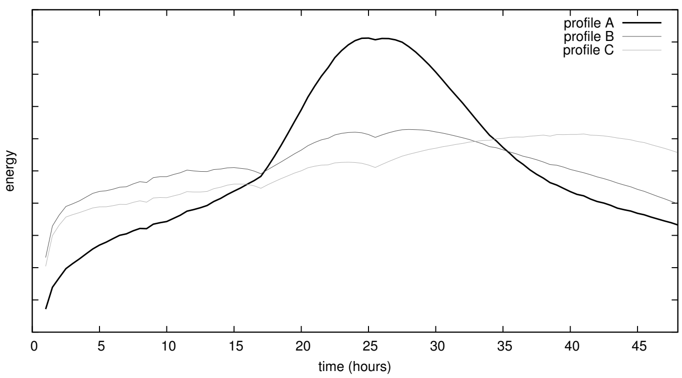

Figure 4. Cluster centroids corresponding to the first semester of 2012.

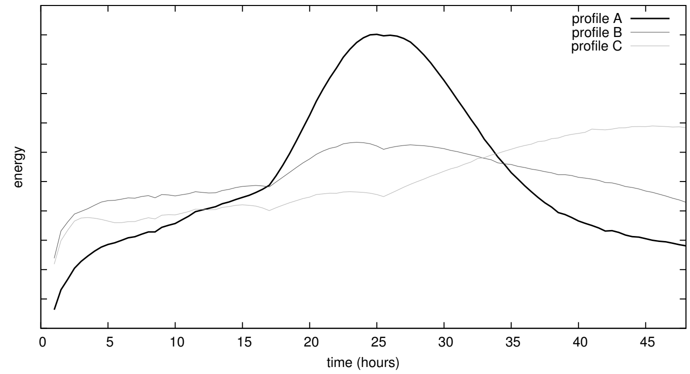

Figure 5. Cluster centroids corresponding to the second semester of 2012.

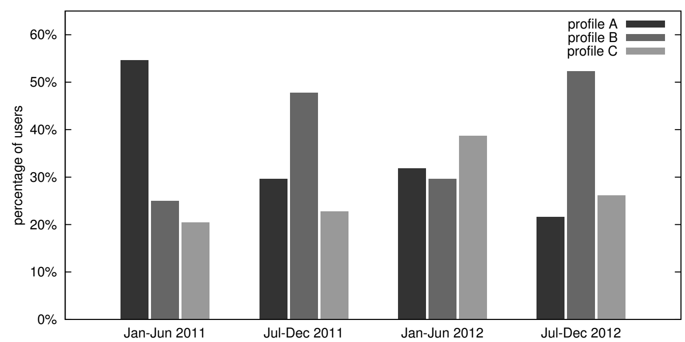

Figure 6. Users' pseudo-periodicity classification over time.

roughly equal to one day. So, these users have typically a daily interaction with Facebook. The second group of users is the one that presents significant energy components at the different time scales, which means that these users interact intensively with facebook, being permanently active. We know that there are several types of Facebook interactions besides wall posting, like messaging, use of different applications, photo upload, chat, among others, but we cannot establish a straighforward relationship between their interaction frequencies. Nevertheless, high activity periods are very likely to include different types of user interactions. Finally, users of the third group use Facebook in a less infrequent way, interacting with the OSN (through wall posting, in this case) in periods that are generally higher than one day. So, these profiles include significant energies for periodicities of several hours.

B. Users Classification

In order to better identify the different activity profiles, the k-means unsupervised clustering algorithm was applied to each one of the four datasets. As previously explained, three clusters were enough to describe the different profiles. In fact, when we used more clusters, some of them tend to overlap.

The different cluster centroids corresponding to the four datasets are presented in Figures 2 to 5. Note that cluster centroids are simply the means of each clustering variable for each cluster. For all datasets, three similar profiles can be clearly identified: profile A exhibits significant energy components around the 24 hours time scale, with lower energy components at the low and high frequency regions of the plot; profile B corresponds to users that include relevant energy components in the complete frequency range; profile C represents users that have significant energy components in the medium to low frequeny range and lower energy components in the high frequency range.

These profiles, clearly revealed by the clustering analysis, are in accordance to our previous analysis. In terms of their wall posting activity, Facebook users can actually be classified into three groups, corresponding to their typical usage behavior.

C. Users Behavior Evolution

Finally, it is important to study the temporal evolution of the different users behavior. Figure 6 shows the size (in terms of the percentage of users) of the different profiles in the four semesters that are included in this study. The main trend that can be identified is a global decrease on the number of users corresponding to profile A and a steady increase (although with some fluctuations) on the number of users corresponding to profiles B and C. This means that the number of users that typically have daily access to Facebook is decreasing; these users tend to behave according to profile B (users that are almost permanently connected) or profile C (users that access to Facebook at longer intervals).

In order to prove this observation, we tried to track how users changed their profiles in the two years period of our study by calculating the number of users that changed between two specific profile types. Figure 7 shows the percentage of users that made a particular change on their profile, for all possible changing combinations. As previously observed, profile A was the one that lost more users: now, we can conclude that the majority of profile A users (aroud 33%) changed their profiles

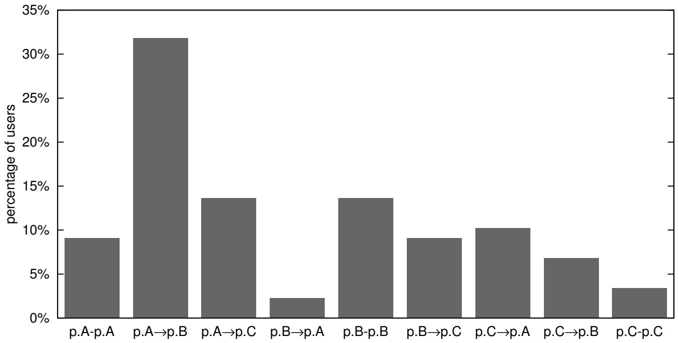

Figure 7. Comparison of users' profiles between Jan-Jun 2011 to Jul-Dec 2012.

to type B, although around 14% of them changed to profile type C. A significantly lower number of users whose profiles were classified as type B changed their profiles: among the ones that changed, the vast majority changed their profiles to type C. Finally, most of the users whose profiles were classified as type C changed their profiles: in this case, a transition to profile type A has a higher probability than a transition to profile type B.

V. Conclusions and Future Work

Some online social networks are among the most popular services on the Internet, having a deep impact on the social, political and economical aspects of our daily lives. Characterizing social network users based on their temporal behavior and activity profiles can be useful to evaluate the performance of existing systems, conduct social studies, develop viral marketing strategies, design new content distribution systems, detect anomalous behaviors, such as bots or compromised accounts, among many other tasks. Based on a medium size dataset, extracted using the Facebook Graph API and containing social networking activity of 200 users during the years of 2011 and 2012, this paper was able to characterize the activity profiles of Facebook users, as well as their time evolution. The periodicity trends of the activity profiles were identified using multiscale analysis based on a wavelet decomposition of the dataset. Three types of activity profiles were identified: an activity profile that exhibits a periodicity around the 24 hours time scale, a second profile that exhibits relevant components in the complete frequency range and a profile that includes important low periodicity components: the first group of users has an activity periodicity roughly equal to one day; users belonging to the second group interact intensively with facebook, while users belonging to the third group use Facebook in time periods that are generally higher than one day. By looking at the temporal evolution of the different users behavior, we were able to identify a global decrease on the number of users corresponding to the first profile and a steady increase on the number of users corresponding to the other two profiles. In fact, the majority of users from the first group (aroud 33%) changed their profiles to the second type, although around 14% of them changed their profiles to the third type.

The identification methodology that was used in this study, which relies on multiscale analysis, will now be applied to the detection of social bots. Social bots, which are automatic or semi-automatic computer programs that mimic humans and/or human behavior in online social networks, can attack users (targets) to pursue a variety of latent goals, such as to harvest private users' data such as email addresses, phone numbers, and other personal data that have monetary value, to spread information or to influence targets. Although many techniques have been proposed to automatically identify spambots in OSNs based on their abnormal behavior, security defenses are not prepared for detecting or stopping a large-scale infiltration caused by social bots.

Acknowledgments

This work was supported by the Fundação para Ciência e Tecnologia (FCT) through PEst-OE/EEI/LA0008/2013, UID/EEA/50008/2013 and PTDC/EEI-TEL/2016/2012.

References

[1] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in Proceedings of the 4th ACM European Conference on Computer Systems, ser. EuroSys '09. New York, NY, USA: ACM, 2009, pp. 205–218. [Online]. Available: http://doi.acm.org/10.1145/1519065.1519089

[2] M. Burke, C. Marlow, and T. Lento, "Feed me: Motivating newcomer contribution in social network sites," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 945–954. [Online]. Available: http://doi.acm.org/10.1145/1518701.1518847

[3] Y.-M. Li, C.-Y. Lai, and C.-W. Chen, "Discovering influencers for marketing in the blogosphere," Inf. Sci., vol. 181, no. 23, Dec. 2011, pp. 5143–5157. [Online]. Available: http://dx.doi.org/10.1016/j.ins.2011.07.023

[4] T. Rodrigues, F. Benevenuto, M. Cha, K. Gummadi, and V. Almeida, "On word-of-mouth based discovery of the web," in Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, ser. IMC '11. New York, NY, USA: ACM, 2011, pp. 381–396. [Online]. Available: http://doi.acm.org/10.1145/2068816.2068852

[5] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," ACM Trans. Web, vol. 1, no. 1, May 2007. [Online]. Available: http://doi.acm.org/10.1145/1232722.1232727

[6] J. M. Pujol et al., "The little engine(s) that could: Scaling online social networks," SIGCOMM Comput. Commun. Rev., vol. 40, no. 4, Aug. 2010, pp. 375–386. [Online]. Available: http://doi.acm.org/10.1145/1851275.1851227

[7] A. H. Wang, "Detecting spam bots in online social networking sites: A machine learning approach," in Proceedings of the 24th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy, ser. DBSec'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 335–342. [Online]. Available: http://dl.acm.org/citation.cfm?id=1875947.1875979

[8] "Graph API - Facebook Developers," https://developers.facebook.com/docs/graph-api, 2015, [Online; accessed May-2015].

[9] L. Guo, E. Tan, S. Chen, X. Zhang, and Y. E. Zhao, "Analyzing patterns of user content generation in online social networks," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 369–378. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557064

[10] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, ser. IMC '09. New York, NY, USA: ACM, 2009, pp. 49–62. [Online]. Available: http://doi.acm.org/10.1145/1644893.1644900

[11] M. Papagelis, V. Murdock, and R. van Zwol, "Individual behavior and social influence in online social systems," in Proceedings of the 22Nd ACM Conference on Hypertext and Hypermedia, ser. HT '11. New York, NY, USA: ACM, 2011, pp. 241–250. [Online]. Available: http://doi.acm.org/10.1145/1995966.1995998

[12] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in Proceedings of the 15th International Conference on Intelligent User Interfaces, ser.

IUI '10. New York, NY, USA: ACM, 2010, pp. 31–40. [Online]. Available: http://doi.acm.org/10.1145/1719970.1719976

[13] M. Cha, A. Mislove, and K. P. Gummadi, "A measurement-driven analysis of information propagation in the flickr social network," in Proceedings of the 18th International Conference on World Wide Web, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 721–730. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526806

[14] M. Valafar, R. Rejaie, and W. Willinger, "Beyond friendship graphs: A study of user interactions in flickr," in Proceedings of the 2Nd ACM Workshop on Online Social Networks, ser. WOSN '09. New York, NY, USA: ACM, 2009, pp. 25–30. [Online]. Available: http://doi.acm.org/10.1145/1592665.1592672

[15] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 211–220. [Online]. Available: http://doi.acm.org/10.1145/1518701.1518736

[16] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in Proceedings of the 2Nd ACM Workshop on Online Social Networks, ser. WOSN '09. New York, NY, USA: ACM, 2009, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/1592665.1592675

[17] M. Burke, C. Marlow, and T. Lento, "Social network activity and social well-being," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1909–1912. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753613

[18] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 35–47. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879147

[19] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves, "Detecting spammers and content promoters in online video social networks," in Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR '09. New York, NY, USA: ACM, 2009, pp. 620–627. [Online]. Available: http://doi.acm.org/10.1145/1571941.1572047

[20] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web, ser. AIRWeb '09. New York, NY, USA: ACM, 2009, pp. 41–48. [Online]. Available: http://doi.acm.org/10.1145/1531914.1531924

[21] S. Webb, J. Caverlee, and C. Pu, "Social honeypots: Making friends with a spammer near you," in Proceedings of Fifth Conference on Email and Anti-Spam, CEAS, Silicon Valley, California, USA, 2008.

[22] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The underground on 140 characters or less," in Proceedings of the 17th ACM Conference on Computer and Communications Security, ser. CCS '10. New York, NY, USA: ACM, 2010, pp. 27–37. [Online]. Available: http://doi.acm.org/10.1145/1866307.1866311

[23] M. A. Vasconcelos, S. Ricci, J. Almeida, F. Benevenuto, and V. Almeida, "Tips, dones and todos: Uncovering user profiles in foursquare," in Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, ser. WSDM '12. New York, NY, USA: ACM, 2012, pp. 653–662. [Online]. Available: http://doi.acm.org/10.1145/2124295.2124372

[24] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding online social network usage from a network perspective," in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, ser. IMC '09. New York, NY, USA: ACM, 2009, pp. 35–48. [Online]. Available: http://doi.acm.org/10.1145/1644893.1644899

[25] J. Jiang et al., "Understanding latent interactions in online social networks," in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ser. IMC '10. New York, NY, USA: ACM, 2010, pp. 369–382. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879190

[26] A. N. Joinson, "Looking at, looking up or keeping up with people?: Motives and use of facebook," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1027–1036. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357213

[27] C. N. Chapman and M. Lahav, "International ethnographic observation of social networking sites," in CHI '08 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA '08. New York, NY, USA: ACM, 2008, pp. 3123–3128. [Online]. Available: http://doi.acm.org/10.1145/1358628.1358818

[28] J. Caverlee and S. Webb, "A large-scale study of myspace: Observations and implications for online social networks," in Proceedings of International Conference on Weblogs and Social Media, Seattle, Washington, USA, 2008.

[29] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," SIAM Journal on Mathematical Analysis, vol. 15, no. 4, 1984, pp. 723–736.

[30] J. Slavic, I. Simonovski, and M. Boltezar, "Damping identification using a continuous wavelet transform: application to real data," Journal of Sound and Vibration, vol. 262, no. 2, 2003, pp. 291 – 307.

[31] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, "An efficient k-means clustering algorithm: analysis and implementation," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, Jul 2002, pp. 881–892.