

A flexible workload model based on roles of interactive users in social networks

Pablo Nicolás Terevinto, Ana Pont, José A. Gil and Josep Domenech
Universitat Politècnica de València (Spain)
Email: {pabtecha,jdomenech}@upvnet.upv.es, {apont,jagil}@disca.upv.es

Abstract—Uses, applications and functionalities of Online Social Network (OSN) are continuously changing and adapting to the new habits of users. The massive adoption of smart mobile devices and the appearance of new roles such as community managers have had a strong influence in the wide use of these networks among all sectors of population regardless of age, gender or socio-economic status. Consequently, the characterization and modeling of OSN user's behavior is a difficult task because their habits and activities change rapidly. This fact motivates us to propose a new OSN user's workload model based on the active roles that they play at any instant when navigating the application. Role-based workload is a flexible way to have a fresh characterization of users because they can adopt new roles, stop using others, or simply modify the way they change between them. Roles and their interrelations can be easily defined in our workload model and generator thus providing a useful tool for fine grain performance evaluation and testability studies.¹

Index Terms—Online Social Network, User characterization, workload, user role.

1. Introduction

Social networking activity is a current global phenomenon that has surpassed all expectations. Statistics [1] show that more than 70% of all internet users are now active on social media, and experts project the number of social network users grow 12.5% each year. But maybe the most interesting point in this growth is the increasing presence of social networking in business and other activities different to leisure and entertainment, like for instance education, research and more timidly for e-governance.

But despite its penetration and massive use, the technical paradox in this kind of applications is the lack of tools, workload models and testbeds for performance evaluation and testability studies. This fact makes difficult, among others, provisioning hardware and software resources in an efficient and appropriate way according to the functional use and workload previsions. As with any other web-based application, the complexity of characterizing a wide spectrum of user's behaviors and the continuous emergence of

applications that change user's habits makes it extremely difficult to get representative workloads for these important studies. This fact is even more relevant due to the intrinsic difficulty for many research communities to get real traces which are usually owned and kept by big companies.

In a previous work [2], we proposed the Dynamic Web Workload Model (DWEB). This model makes it possible to characterize and reproduce the behavior of web users which is usually a difficult task due to the continuous interaction between them and the offered content. To do this, DWEB introduced two concepts that permit to consider different levels of user interactions. First, the user navigation concept allows us to represent dynamic reactions of users when they interact with web content and services. These reactions modify the user's response according to the content provided by the OSN server or other parameters, as for instance, response time or quality of service. This feature permits to create interactive users. Second, the user roles concept defines the different behaviors of users according to the characteristics of the visited site, their personal goals and their active involvement. By implementing these two concepts in our workload generator we can mimic the behavior of the actual web users' community.

Departing from the versatility offered by DWEB, this paper proposes a workload model based on roles representing a variety of OSN user's profiles. In this paper we propose a workload model based on roles. We present three different user's roles which have been currently identified in different OSN (Facebook, Twitter, LinkedIn, MySpace, Pinterest, Tuenti). These roles are representative of the most common user's profiles in these kind of networks and can be easily identified in one degree or another in all of them. Nevertheless, the activities that can be performed in each role can be different according to the services offered by each application. User's roles define a level of abstraction that allows us to create a flexible model which can be adapted to any kind of OSN by modifying the activities performed by each role.

In summary, the main contributions of this work are: i) a flexible and adaptable OSN workload model based on the main current user's roles, ii) a generation of a dynamic workload that accurately reproduces interactive OSN user.

The remainder of this paper is organized as follows: in Section 2 we present the roles of our model and describe the activities of an OSN which are later used to create the

1. This work has been partially supported by the Spanish Ministry Economy and Competitiveness under grant TIN-2013-43913-R

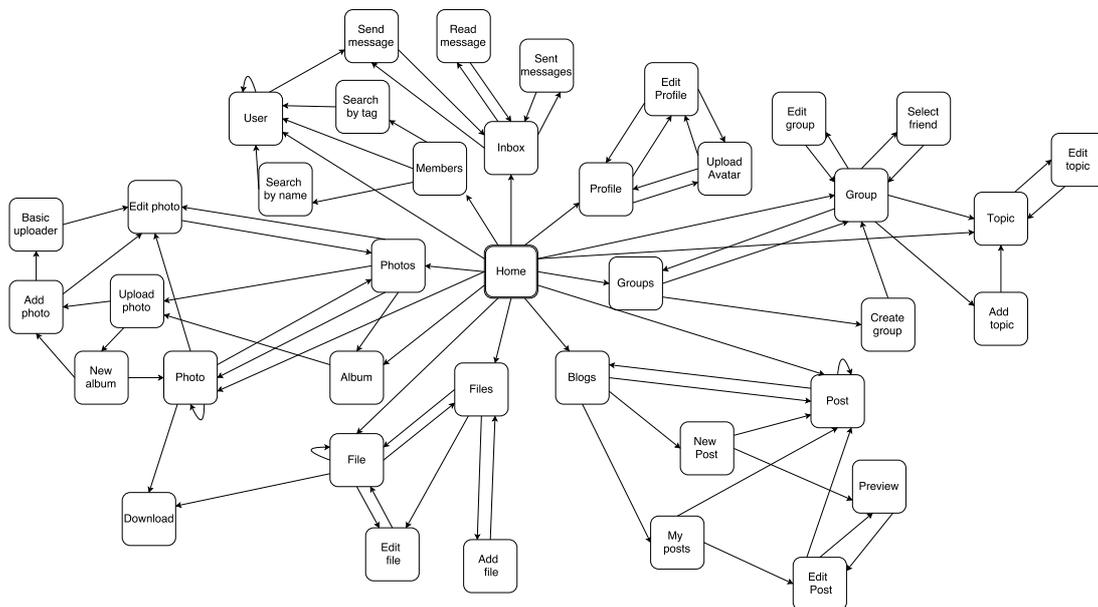


Figure 1. Navigational graph based on the OSN interface.

workload model. Then, the model is used to generate the workload as Section 3 describes. Later in Section 4 we provide a comparison of our work with the most related works that can be found in the open literature. Finally, we provide the main conclusions of our work and give some ideas about our future work.

2. Workload modeling based on user roles

In this section we present a workload model which is based on the concept of role implemented in DWEB. For this, we present the typical roles identified and an example of this model for an specific OSN. First of all, for sake of clarity, we define the nomenclature used in this work.

- **Role:** defines the user's behavior while interacts with the application. Roles define the active involvement of users and produce a set of sequences of a activities that can be done in parallel and/or sequentially by each user.
- **Activity:** is the sequence of actions performed by the user aimed at achieve a simple goal. Examples of activities are: uploading a file, posting a blog or commenting a photo. Many current OSN offer similar activities, but the way they are made can differ in each one. This is because the sequence of actions involved is different.
- **Action:** each of the interactions between the user agent and the application (OSN). They are originated from the interrelation between user and the application's interface. Actions can trigger the request of a new page or the change of state of the current

one. However, not all user's interactions produce an action. For instance, filling up a form may not produce an action, while sending it does. Actions are intrinsic to the application. Between the successive actions of an activity usually it elapses time, which is the latency between an action and the user's reaction, as a consequence of the previous ones (think time).

2.1. User's roles definition

An accurate definition of roles is important to later develop a flexible workload model that: i) can represent the activities done by users when they use this type of applications, ii) permits to conduct fine grain evaluation studies. The user roles that we propose in this paper are representative of the current behaviors of any individual when they participate in a OSN. Those roles are: user, generator user, and reader user. These profiles have been also identified in [3] after analyzing real traces from a blogosphere. This observation is also consistent with the user characterization presented in [4] and [5]. And finally, some internet monitoring websites [6] and [7] provide similar inputs about OSN user's active participation.

- **Social user:** this type of users browse the OSN, but also interact with other users and generate new content in the OSN.
- **Generator user:** content generator users are those whose main purpose is to update and upload new content to the social network and respond to other user's comments and messages. This is the typical

role for a social media manager or community manager for example.

- **Reader user:** users which generally log in to check the updates in the OSN. Reader user do not interact with other users or comment any content. They only browse the OSN.

2.2. Activities for an OSN

Although roles are quite independent from the type of OSN, the activities of each role are not. The activities and the actions that implement the activities are dependent of the services offered by the OSN. For this reason and in order to present an example model, an online social network powered by the open source social networking engine Elgg was used. The Elgg 1.8 version was selected with the plugins for the most common activities which users usually undertake while navigating OSN.

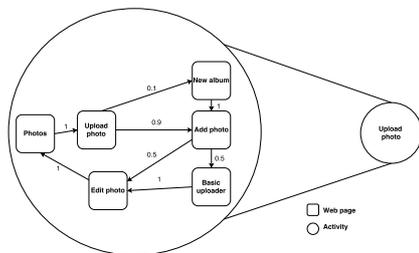


Figure 2. Grouping several actions into the activity *upload photo*

Our first step was an analysis of the OSN interface and the possible transitions between the different pages. Here we identified the starting point of the web site, the Home page. It provides a timeline with the latest visible changes occurred on the site and therefore, offers direct links to specific content such as photos, files, posts, albums, groups or members. Also, the OSN interface has fixed links to each of the types of content mentioned before. This analysis resulted in the navigational graph shown in Figure 1 where each square node represents a single page of the OSN. This graph shows the transitions present in each of the OSN web pages without considering the transitions between different types of content which are always present in every page. This omission aims to reduce the complexity of the graph. Although those transitions have been omitted in the figure, they will be taken into consideration in the following steps of the modeling process.

Using this graph we identify the activities that a user can perform in our OSN.

As an example, Figure 2 shows how we have grouped the different necessary actions to upload a photo to the social network into the activity *upload photo*. Each square node represents a web page while activities are represented with circles. Besides the transitions between pages, an activity also includes actions which provide interactivity. For instance, after each transition we can modify the user's response time or the length of the answer according to

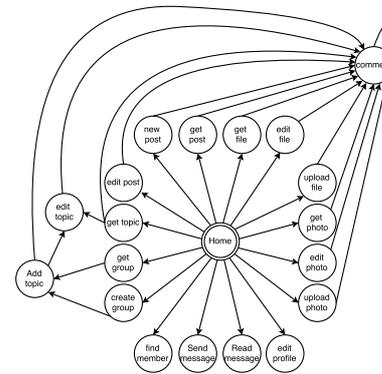


Figure 3. Navigation graph for a social user.

the amount of data of the server response. Responses with more information usually require more time for the user to process them. Quality of service could also affect an action because the user may decide to stop browsing after several unsuccessful connections to the server or due to high latency. These factors can influence the probabilities of transition, changing the activity of closing the session.

2.3. Linking activities to user's roles

At this point, we took into consideration the definitions of roles aforementioned and selected those activities which better suit each role. Additionally, the user's roles are finally defined by giving probabilities to the transition between activities

Social role model. A social user spends most of the time browsing the social network, but eventually interacts with other users or generates new content to share with the rest of users.

In our model, we considered the results presented in [4] stating that: i) 92% of the total amount of workload consist in browsing content, ii) users tend to repeat the same activity around 67% of the time, iii) users tend to do related activities instead of doing unrelated ones.

The navigational graph for the social role is the result of grouping all the activities described in the previous section. Figure 3 represents the initial Home page and the transitions to the different activities. For sake of clarity, the activities not accessible from the starting page where added to the graph but the transitions between them have been omitted. The complete representation of transitions is presented in Table 1. It represents the transitions between activities and the probabilities for each transition with values represented in percentage. These values and the ones presented in other tables are merely an example, although they are enough consistent with users navigations. Most of the time spent is browsing the site and mainly repeating the same activity. Also, changes between activities tend to occur to related activities.

Generator role model. Content generator users only perform activities which produce new content or modify

TABLE 1. PROBABILITIES (%) OF TRANSITIONS BETWEEN ACTIVITIES FOR A SOCIAL ROLE.

	Home	Get post	Get file	Get photo	Get group	Get topic	New post	Upload file	Upload photo	Create group	Add topic	Edit post	Edit file	Edit photo	Edit topic	Comment	Read message	Send message	Find member	Edit profile
Home	-	18	18	18	18	18	1	1	1	1	-	1	1	1	1	-	0.5	0.5	0.5	0.5
Get post	3	66	4	4	4	4	2	-	-	-	-	1	-	-	-	-	10	0.5	0.5	0.5
Get file	3	4	66	4	4	4	-	2	-	-	-	-	1	-	-	-	10	0.5	0.5	0.5
Get photo	3	4	4	66	4	4	-	-	2	-	-	-	-	1	-	-	10	0.5	0.5	0.5
Get group	13	19	19	19	4	19	1	1	1	1	1	-	-	-	-	-	0.5	0.5	0.5	0.5
Get topic	3	4	4	4	4	66	-	-	-	-	2	-	-	-	1	10	0.5	0.5	0.5	0.5
New post	12	50	5	5	5	5	3	-	-	-	-	3	-	-	-	10	0.5	0.5	0.5	0.5
Upload file	12	5	50	5	5	5	-	3	-	-	-	-	3	-	-	10	0.5	0.5	0.5	0.5
Upload photo	12	5	5	51	5	5	-	-	-	-	-	-	-	5	-	10	0.5	0.5	0.5	0.5
Create group	-	-	-	-	-	-	25	25	25	-	25	-	-	-	-	-	-	-	-	-
Add topic	12	5	5	5	5	50	-	-	-	-	3	-	-	-	3	10	0.5	0.5	0.5	0.5
Edit post	59	10	2	2	2	2	1	1	1	1	1	1	-	-	-	15	0.5	0.5	0.5	0.5
Edit file	59	2	10	2	2	2	1	1	1	1	1	-	1	-	-	15	0.5	0.5	0.5	0.5
Edit photo	59	2	2	10	2	2	1	1	1	1	1	-	-	1	-	15	0.5	0.5	0.5	0.5
Edit topic	59	2	2	2	2	10	1	1	1	1	1	-	-	-	1	15	0.5	0.5	0.5	0.5
Comment	8	16	16	16	16	16	1	1	1	1	1	1	1	1	1	1	0.5	0.5	0.5	0.5
Read message	20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	65	-	-
Send message	80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-
Find member	-	19	18	19	18	18	-	-	-	-	-	-	-	-	-	-	-	-	8	-
Edit profile	3.5	18	18	19	18	18	1	1	1	1	1	-	-	-	-	-	0.5	0.5	0.5	-

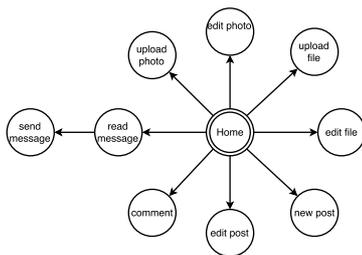


Figure 4. Navigation graph for a generator user.

existing one. We also consider that these users read and answer private messages from other users as well as they answer comments from other users on the generated content.

Figure 4 shows the navigation graph including only the transitions from the initial node (Home) to each activity. To reduce the complexity of the graph given the high density of connections between nodes, the remaining transitions are shown at Table 2 where each probability is expressed in percentage. As this type of users do not tend to browse content, the navigations had to be readjusted by giving more weight to the actions that create new content rather than those that modify existing ones. Also, we have considered that this kind of user uses the private messages mainly to answer other users. Therefore the activity *Send message* only follows a *read message* first.

Reader role model. The only interaction of this type of users with the OSN is requesting content. They do not interact with other users nor generate new content to contribute with the OSN growth. With this in mind, we have selected the activities which allowed the user to browse the site without creating new content. Figure 5 shows the graph

TABLE 2. PROBABILITIES (%) OF TRANSITIONS BETWEEN ACTIVITIES FOR A GENERATOR ROLE.

	Home	New post	Upload file	Upload photo	Edit post	Edit file	Edit photo	Comment	Read message	Send message
Home	-	18	18	18	2	2	2	20	20	-
New post	5	17	17	17	4	-	-	30	10	-
Upload file	5	17	17	17	-	4	-	30	10	-
Upload photo	5	17	17	17	-	-	4	30	10	-
Edit post	23	17	17	17	1	-	-	15	10	-
Edit file	23	17	17	17	-	1	-	15	10	-
Edit photo	23	17	17	17	-	-	1	15	10	-
Comment	20	16	16	16	-	-	-	22	10	-
Read message	10	13	13	14	-	-	-	-	15	35
Send message	45	15	15	15	-	-	-	-	10	-

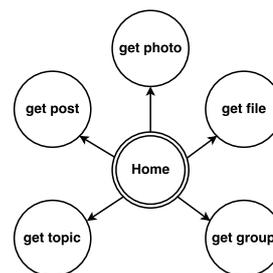


Figure 5. Navigation graph for a reader user.

with the Home page as the starting point, and the activities from the initial node.

Table 3 represents the transitions between activities and the probabilities for each transition with percentage values. Every activity has a similar behavior, except for *Get group* which tends to get topics, posts, photos or files of that

specific group. Also here it can be appreciated that the user tends to repeat the same activity.

TABLE 3. PROBABILITIES (%) OF TRANSITIONS BETWEEN ACTIVITIES FOR A READER ROLE.

	Home	Get post	Get file	Get photo	Get group	Get topic
Home	-	20	20	20	20	20
Get post	2	70	7	7	7	7
Get file	2	7	70	7	7	7
Get photo	2	7	7	70	7	7
Get group	-	25	25	25	-	25
Get topic	2	7	7	7	7	70

3. Workload generation

This characterization of user's behaviors in roles, and the proposed model are addressed to finally generate a representative workload for testing applications and for any type of fine grain performance evaluation studies. To this end, we implement a workload with interactive users using GUERNICA and LoadG. GUERNICA [8] is a web workload generator which allows the user to define interactive web workloads based on the concept of DWEB previously mentioned and considering also current browsers facilities. LoadG is a graphical configuration tool which provides a user-friendly interface to create the navigation and workload XML files to configure GUERNICA. LoadG helps to easily design the models by graphically adding new nodes and interconnect nodes assigning probabilities of transitions between them. Also, each node can be edited and the user can add one or more actions associated with the node. In this case, we define each activity as a node as shown in the previous sections and each node contains the different actions necessary to achieve the goal of the activity. Once the model is implemented, LoadG allows the user to export the model in a XML format which GUERNICA can interpret. At this point, the generator permits to define conditional transitions among activities, considering, for instance, the results of previous actions, QoS parameters or the time of day. Figure 6 is a screenshot of the reader user's model implemented with LoadG. This interface offers different colors and line thickness to distinguish between a single transition and multiple transitions.

4. Related Work

Unlike other computational workloads, web-based application ones involve people who interact and are sensitive to the offered content. That makes it difficult to have available representative workloads for performance evaluation or testability studies. In order to provide detailed and realistic workloads for web-based applications, the faithful characterization of users behaviors is a fundamental pillar. One of the earliest attempts to characterize in detail web users behavior can be found in [9] where the intrinsic characteristics of web

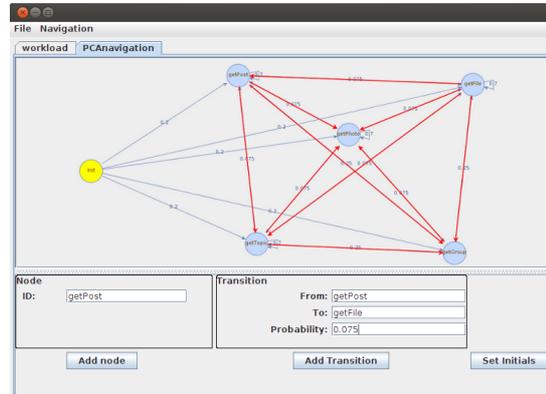


Figure 6. Reader user's navigation graph implemented with LoadG.

workloads were settled and the importance of considering the user interactivity was pointed out. For the case of e-commerce sites, this work identified the main user transactions which supposed a workload characterization based on user's activities according to our nomenclature.

This user model was later extended in [10] to capture application inter-request and data dependencies in order to consider a certain degree of interactivity in the workload. More recently, [11] characterized and modeled generic web user's navigations that include current browsers facilities such as the use of back button or opening new tabs, and they also represented the user dynamic interaction with the provided contents or according to QoS parameters.

With the growing and wide penetration of social network applications, it becomes necessary to define specific web workload models for them. A recent work [11] presents an interesting survey where several approaches to characterize OSN users are explored. They classified these attempts according to the point of view of connection, interaction, traffic activity, mobile social behavior, and malicious behavior. This work also focuses on the importance of understanding user's behavior both for Internet and Applications service providers. Also a better knowledge of user's behavior can help for enhancing user experience.

The most related work to ours [3] defines the workload for a blogspace. In this work, a similar idea to ours for identifying user's roles was made for the first time. They describe profiles like "blurker" and "commenter" and also distinguish between write session and read session associated to those profiles. But, this work only considers the scope of the blogosphere which is only a specific case of social network.

In [4], they characterize users behavior in online social networks by collecting and analyzing data obtained from a representative social network aggregator. While their model relies on user activities and the transitions between them, our work, provides a higher level of abstraction, by identifying the main user's roles in OSN according to the major functionalities and characteristics of these applications [12].

Other authors have focused on characterizing specific

user's actions or roles. For instance, the work presented in [13] models the user posting behavior on social media according to the influence of content factors but it does not consider other user roles which are also of huge interest.

Another approach can be found in [14], where a normal OSN user's behavior is characterized and modeled in order to identify significant deviations aimed at detecting anomalous or malicious activities. Authors use statistical techniques applied to user pattern accesses, such as the number of likes at day. This level of characterization does not permit to easily identify user's roles or specific profiles.

An interesting work to assist the workload generation process for OSN is [15]. They characterize the workload at the level of user navigation, identifying sessions, subsessions within a session, session durations, active and inactive times, inter arrival time, bytes per session, and they pay also attention to popularity. Nevertheless, some data should be reviewed for updating the results to the current context.

Aimed at detecting user's intentions and preferences for efficient recommendations system design, the work presented in [16] also analyzes user behaviors in social media systems considering temporal context. This work focuses only on the creation of more accurate actions and recommendations, but does not consider user's profiles

In general, pro-active users of social networks not only focus on one type of them but they are also active members of others. Therefore, to understand how users distribute their activities across different sites is also of interest. In these vein [17] investigates the relationship between user's patterns of two well-known social sites to better understand this phenomenon.

5. Conclusions and future work

In conclusion, in this paper we have presented a new method for designing a flexible model to represent the interactive behavior of OSN users.

Our model is based on the typical roles that active users play currently in these networks. Roles present a level of abstraction when modeling the workload that permit to offer a flexible and fresh characterization. Consequently, it can be easily adapted to the ever-changing environment of these applications, where new trends, functionalities and technologies are appearing constantly.

By implementing the resulting model in our workload generator (GUERNICA) we can reproduce in an accurate way the interactive actions performed by OSN users when navigating. Our generator allows conditional user navigations depending on the content offered or other type of parameters related, for instance, to QoS metrics. The combination of GUERNICA and LoadG, a powerful and friendly tool to graphically define and edit web based workloads, permits to generate detailed workloads for performance evaluation and testability studies.

The model presented in this work has been applied to the specific case of Elgg, an open source OSN that includes the typical functionalities of these type of networks. At the present, Elgg has been conveniently filled out with contents

to serve as back end for a future benchmark aimed, among others, at evaluating OSN.

References

- [1] I. W. Stats, "Usage and statistics," <http://www.internetworldstats.com/>.
- [2] R. Pea-Ortiz, J. Sahuquillo, A. Pont, and J. A. Gil, "Dweb model: Representing web 2.0 dynamism," *Computer Communications*, vol. 32, no. 6, pp. 1118 – 1128, 2009.
- [3] F. Duarte, B. Mattos, J. Almeida, V. Almeida, M. Curiel, and A. Bestavros, "Hierarchical characterization and generation of blogosphere workloads," 2006.
- [4] 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.
- [5] M. Maia, J. Almeida, and V. Almeida, "Identifying user behavior in online social networks," in *Proceedings of the 1st Workshop on Social Network Systems*, ser. SocialNets '08. New York, NY, USA: ACM, 2008, pp. 1–6.
- [6] Statista, "Social networks usage in spain," <http://www.statista.com/study/31830/social-networks-usage-statista-dossier/>.
- [7] S. M. Examiner, "Facebook and twitter user behavior changes: New research," <http://www.internetworldstats.com/>.
- [8] R. Pea-Ortiz, J. Sahuquillo, A. Pont, and J. A. Gil, "Generating dynamic workload for web performance evaluation," *XXI Jornadas de Paralelismo (JPPP'10)*, pp. 711–718, 2010.
- [9] D. A. Menasce and V. Almeida, *Capacity Planning for Web Services: Metrics, Models, and Methods*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.
- [10] M. Shams, D. Krishnamurthy, and B. Far, "A model-based approach for testing the performance of web applications," in *Proceedings of the 3rd International Workshop on Software Quality Assurance*, ser. SOQUA '06. New York, NY, USA: ACM, 2006, pp. 54–61.
- [11] L. Jin, Y. Chen, T. Wang, P. Hui, and A. Vasilakos, "Understanding user behavior in online social networks: a survey," *Communications Magazine, IEEE*, vol. 51, no. 9, pp. 144–150, September 2013.
- [12] J. Heidemann, M. Klier, and F. Probst, "Online social networks: A survey of a global phenomenon," *Computer Networks*, vol. 56, no. 18, pp. 3866 – 3878, 2012, the {WEB} we live in.
- [13] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang, "Modeling user posting behavior on social media," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '12. New York, NY, USA: ACM, 2012, pp. 545–554.
- [14] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug. 2014, pp. 223–238.
- [15] 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.
- [16] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang, "A temporal context-aware model for user behavior modeling in social media systems," in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '14. New York, NY, USA: ACM, 2014, pp. 1543–1554.
- [17] R. Ottoni, D. B. Las Casas, J. P. Pesce, W. Meira Jr, C. Wilson, A. Mislove, and V. Almeida, "Of pins and tweets: Investigating how users behave across image-and text-based social networks." in *ICWSM*, 2014.