Evolution of User Activity and Community Formation in an Online Social Network

Andreas Kalaitzakis, Harris Papadakis and Paraskevi Fragopoulou
Department of Applied Informatics and Multimedia
Technological Educational Institute of Crete
71500 Heraklion, Crete, Greece
Email: adanar@epp.teicrete.gr, fragopou@ics.forth.gr

***Abstract*—The paper performs an empirical study of the MySpace Online Social Network (OSN). Its purpose is threefold aiming to capture the evolution of user population, to examine user activity, and finally to characterize community formation using two well established community finding algorithms, namely the Fortunato et al. and the Clique Percolation algorithms. Both algorithms are known to be effective in identifying communities in large graphs, starting at seed nodes and utilizing only local interactions between nodes.**

One million user profiles were randomly collected in a month's period and stored in a local database for further processing. For each profile certain attributes were fetched: profile status (public, private, invalid), member since and last login dates, number of friends, number of views, etc. The profiles and their attributes were analyzed in order to reveal the evolution in user population and the activity of the participating members. Significant conclusions were drawn for the synthesis of the population based on profile status, the number of friends, and the duration MySpace members stay active.

Subsequently, a large number of communities were identified aiming to reveal the structure of the underlying social network graph. The collected data were further analyzed in order to characterize community size and density but also to retrieve correlations in the activity among members of the same community. A total of 171 communities were detected with Fortunato's algorithm, while using Clique Percolation this number was 201. Results demonstrate that MySpace members tend to form dense communities. For the first time, strong correlation in the last login date (the main attribute that shows user activity) for members of the same community was documented. It was also shown that members participating in the same community have similar values for other attributes like for example number of friends. Lastly, there is strong evidence that participation of users in communities inhibits them from abandoning MySpace. As a last observation, members that abandoned MySpace shortly after their account creation (the so-called Tourists), have very low connectivity and thus they do not participate in communities.

I. Introduction

In recent years, social networking [3] has been driving a dramatic evolution due to the increasing use of Web 2.0. The major impact of Online Social Networks (OSNs), such as MySpace [17] and Facebook [6], on Information Technology paved the way for popular Internet applications and attracted hundreds of million of users. This evolution also attracted the interest of the academic community and prompted several researchers to examine population synthesis, user activity, evolution, and structure, of popular OSNs [1], [3], [11], [14]–[16], [20].

First efforts on this direction included the analysis of the friendship relations of popular social networking sites from data collected online. Some studies focused on the evolution of user population based on data collected over a period of time [11], [14]. They also tried to capture the evolution of the friendship relations over time, considering this as the main element that characterize activity in social networks. In [16] authors aim to capture the decline in use activity of the MySpace OSN. Their research is based on the collection and analysis of a large number of user profiles and provides significant evidence on the activity of MySpace users from several different points of view. The authors claim that capturing decline in the activity of an OSN is a nontrivial task since network administrators tend to hide such events from public view, as much as possible, in order not to influence other members.

In this paper, we collect a large number of profiles from the MySpace OSN and try to characterize the way the synthesis of the population changed over time, the friendship graph, and the way user activity evolves in the network. To this purpose one million user profiles were randomly collected over a period of time, and they were analyzed accordingly. One of the important features of MySpace, is the fact that it enables researches to collect user profiles along with all accompanied attributes and to analyze them, deriving interesting results for the network.

Apart from characterizing user activity based on collective analysis of individual profiles, we go one step further, trying to derive interesting results by extracting a large number of communities from the underlying social graph, using two well known community detection algorithms found in the literature, namely the Fortunato et al. [12], [13] and the Clique Percolation [4], [19] algorithms.

A community is usually defined as a densely connected subgraph of an initial graph. Several attempts have been made to provide a formal definition for this generally described "community detection" concept in networks. A *strong community* was defined as a group of nodes for which each node of the community has more edges to other nodes of the same community than to nodes outside the community

This project is implemented through the Operational Program "ARCHIMEDE III: Education and Lifelong Learning" (project P2PCOORD) and is co-financed by the European Union (European Social Fund) and Greek national funds (National Strategic Reference Framework 2007 - 2013).

[7]. This is a relatively strict definition, in the sense that it does not allow for overlapping communities and creates a hierarchical community structure since the entire graph can be a community itself. A *weak community*, was later defined as a subgraph in which the sum of all node degrees within the community is larger than the sum of all node degrees toward the rest of the graph [10].

Variations also appear in the method used to identify communities: Most of the algorithms that appear in the literature follow an iterative approach starting by characterizing either the entire network, or each individual node as community, and splitting [8], [18] or merging [10], respectively. These methods produce a hierarchy of partitions. There is an entire hierarchy of communities, because communities are nested: small communities compose larger ones, which in turn are put together to form even larger ones. By merging or splitting communities one can build a hierarchical tree of community partitions called *dendrogram*. The modularity criterion defined in [8] is a measure of the quality of a partition, and can be used to identify a single optimal partition, i.e. the one corresponding to the largest modularity value.

Trying to extract communities from real application graphs put inevitably under reconsideration the capabilities of the existing community detection methods. They proved to be inefficient in adapting to the modern distributed environment and were rendered obsolete leaving a significant void in scientific literature. This led to the invention of new methods, able to deal with the new restrictions imposed by modern OSN's, such as the ones we utilize in this paper [4], [12]. In these approaches a community surrounding a seed node is identified by progressively adding nodes and expanding a small community. Other interesting community finding methods can be found in [5], [9], [24]. For the interested reader, two comprehensive and relatively recent surveys covering the latest developments in the field can be found in [12], [21].

We performed an empirical study of the MySpace OSN, collecting not only a large number of user profiles in order to characterize population synthesis, evolution in user population, and user activity, but also extracting a large number of community structures in order to analyze the underlying social network graph and to correlate characteristics and activity of users participating in communities. Two local algorithms were used to extract communities, namely, the Fortunato and the Clique Percolation methods. The main conclusions can be summarizes as follows: Population synthesis changed over time, as the difference between public and private profiles shrunk, and invalid profiled overpassed first private ones and subsequently the public ones. Use activity was shown to be diminishing, while the number of new users joining the network remains high. Drawing safe conclusions regarding MySpace was based on the extraction of a combination of data for every single profile, including among others, the status, the Member Since and Last Login dates, the number of friends. Indeed, these information alone had been very enlightening helping us to shed light to all crucial questions mentioned before.

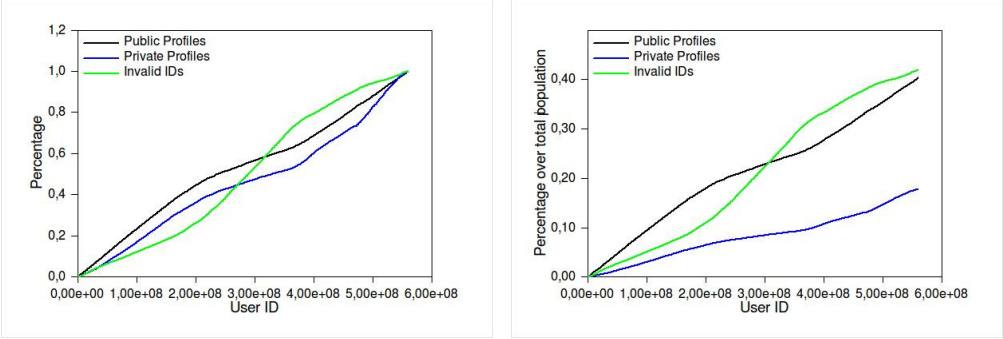
Based on the analysis of the communities, important conclusions were drawn on the size of the communities that are formed in the network, but also in the community density as shown by the number of edges among participating member. As an important result, strong correlation was demonstrated in the Last Login date among members of the same community, and other attributes like the number of friends. We draw the conclusion that participation in strong communities inhibits users from abandoning the OSN in order to migrate to other more popular networks that emerge at times. As a last observation, we notice that members that abandoned MySpace shortly after their account creation (the so-called Tourists), have very low connectivity and thus they do not participate in communities.

The remaining of the paper is organized as follows: In section II, we present the methodology we followed to collect the profiles used in our experiments. In section III, we analyze population evolution and user activity based on the collected profiles. The community detection algorithms we use are briefly presented in Section IV, and the experimental results related to the community analysis of the MySpace OSN are provided in Section V. Finally, we conclude in Section VI.

II. Profile Collection and Methodology

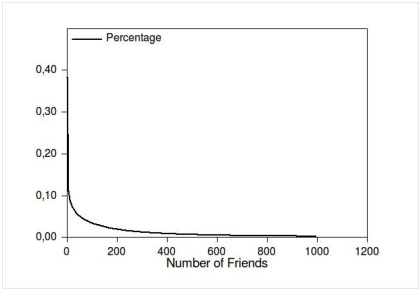
MySpace is a popular OSN with several hundreds of millions members [17]. MySpace allocates numeric user IDs in a sequential fashion. This claim was documented in [16] which collected and analyzed a large number of user profiles in February 2009. Another important characteristic of MySpace that facilitates its analysis, is that user profiles are accessible via HTTP requests, and unlike other social networks, like for example Facebook, one does not need to rely on the availability of enormous static datasets, released periodically by the company. A user, writing his own scripts is able to dynamically collect random user profiles according to its needs. For a certain numeric user ID, the corresponding profile is easily accessed through an http request with the following URL http://www.myspace.com/ID. The file downloaded from this URL provides all the available information for the user with the specified ID. Using html parsing to preprocess the downloaded profiles we extracted all necessary attributes for our research.

Each user profile records a number of information regarding the user. Apart from the numeric user ID, there is a Status attribute with values in {Public, Private, Invalid}. Other attributes are the Alias, the number of friends, and the number of views. Furthermore, for each profile the following dates are recorded, Member Since date, which is the date a user joined the OSN, and the Last Login date, which is the most recent date a user logged in the network. We verified that a user's Last Login is indeed updated each time a user logs into its MySpace account. Last Login and Member Since information is available only for public profiles. The relative age of user accounts can be also inferred by the user ID since, as we already mentioned, user IDs are monotonically assigned, thus a smaller ID has been created earlier in time compared to a larger one. Finally, for each profile we recorder the date it was fetched (Fetch Date) since the profiles were collected over a period of several days.



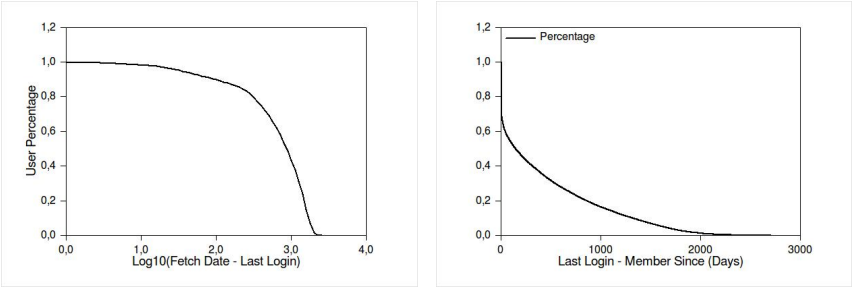
(a) (b)

Fig. 1. Synthesis of user population. (a) CDF of user ID based on profile status. (b) CDF of user ID based on profile status over total population.



(a)

Fig. 2. (a) CDF of number of friends.



(a) (b)

Fig. 3. (a) CCDF of (Fetch Date - Last Login) for public profiles. (b) CDF of (Last Login - Member Since), i.e. number of days profiles are active.

We collected a total of 999.937 user profiles over a period of one month, from the 2^{nd} to the 28^{th} of September 2011. To summarize, for each user profile we retrieved the following attributes:

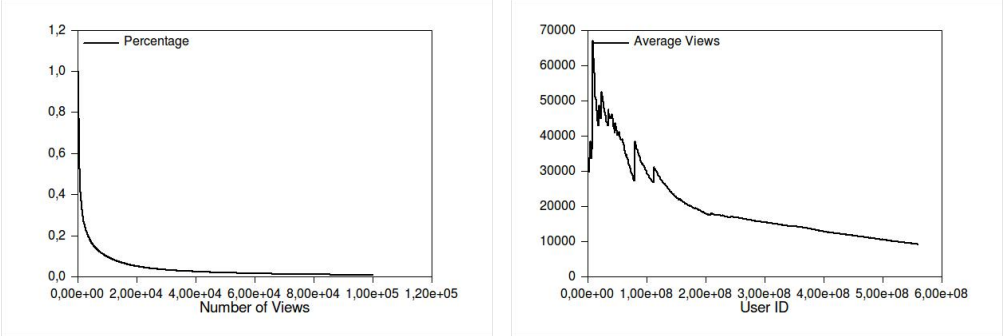
- ID
- Status: Public, Private or Invalid
- Alias
- Number of friends
- Number of views
- Last login date
- Member since date
- Fetch date

Based on Status (Public, Private, Invalid) the user population has the following composition:

- Total profiles: 999.937
- Public profiles: 403.183
- Private profiles: 177.664
- Invalid profiles: 419.090

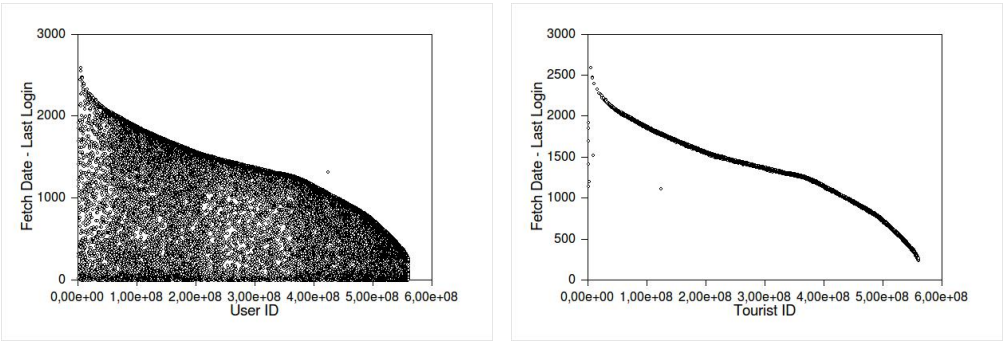
The first indications regarding population's composition are confirming our assumptions based on previous work [16]. Out of 999.937 randomly selected profiles, 419.090 were found to be invalid, corresponding to 41.91% of our sample population and allowing to safely conclude that a large number of MySpace users abandoned the network. The remaining 580.847 profiles (or 58.09% of the total) is shared between public and private, with public profiles being 403.183, corresponding to 40.32% for the sampled population, while private profiles are only 177.664, corresponding to 17.76% of the sampled population.

Compared to [16] which collected and analyzed a large number of MySpace user profiles in February 2009, at the time, the largest allocated user ID was 455 millions while



(a) (b)

Fig. 4. (a) CDF of profile number of views. (b) Mean absolute value of profiles number of views for profiles with ID less than X.



(a) (b)

Fig. 5. (a) (Fetch Date - Last Login) for all profiles. (b) (Fetch Date - Last Login) for Tourists only.

at the time of our experiments the largest allocated user ID was 650 millions. Thus in a period of one year and seven months 200 million new MySpace user accounts were created, indicating that while MySpace users tend to abandon the OSN, the arrival of new users continues at a rate of approximately 348.000 new user accounts per day.

III. Analyzing Population Evolution and User Activity

For all recorded IDs we followed the change in the synthesis of the population based on the percentage of profile status over time. Figures 1(a) and 1(b) present this information. Figure 1(a) shows the CDF (Cumulative Distribution Function) for each different type of profile (Public vs Private vs Invalid), while Figure 1(b) shows the CDF for each profile type as a percentage of the total population. These figures show the distribution of each group of profiles across the entire range of user IDs. The results indicate that for a large initial portion of the ID space, the percentage of public profiles was higher than that of the private profiles, while towards the end of the ID space this difference is vanishing. The results are even more noteworthy if we take into consideration Invalid profiles. While in the beginning of the ID space the percentage of Invalid profiles was lower compared to that of Private and Public ones, it increased over time and at some point, clearly shown in the figures, Invalid profiles overpassed Private ones, and a little later they also exceeded Private profiles. This is an evidence that users either abandoned MySpace OSN and migrated to other OSNs that gained popularity at the given time period, or OSN administrators aggressively removed profiles for violation of use.

Following the population synthesis based on ID Status, we approached the term activity in three different ways. The easiest way to determine the level of activity of a specific user is to examine its number of friends. In this context, a user that counts, for example, one hundred friends is considered to be more active than a user with a smaller number of friends. Figure 2(a) presents the CDF of IDs based on number of friends. From this figure we notice that a large number of MySpace users have a small number of friends. We could mention that about 5% of users have more that 100 friends.

The second way to characterize activity is based on the difference, expressed in days, between the date that the user had logged in for the last time (Last Login) and the date its profile was fetched for the experiments (Fetch Date). Figure 3(a) depicts the CCDF of the difference in days between a users Last Login and the date his profile was fetched (Fetch Date). According to our statistics only 1.8% of profiles did actually login, up to ten days before the fetch process. On the other hand, 8,3% were the users that logged in up to one hundred days before profile fetching took place. This leaves us with almost 90% of the profiles not having logged in for more than one hundred days. Another way to monitor the activity of users is by analyzing the number of profile views. The results

are shown in Figures 4(b) and 4(a).

Subsequently, we analyze a special category of users, the so-called Tourists [16]. The basic characteristic of tourists is the very small time-span between the date they created their account and the date they logged in to the system for the last time. As shown in Figure 3(b), about 33% percent of the users abandoned the network between one to ten days after they created their account. Moreover, another 13% percent has been found to be active for a period of ten to one hundred days. Practically this means that these users had a short stay in the network.

As a last indication for user activity, we study the relationship between a user ID and the users Last Login date, trying to collect evidence on the activity of users compared to account creation time. Figure 5(a) is a scattered plot presenting the Fetch date - Last login for Public profiles. The monotonic decrease of this plot is understandable and the appearance of an clear edge at the top is fully explained by the existence of Tourists, whose Last Login occurred shortly after their account creation time. Thus, in addition to Figure 3(b), the sharp edge isolated in Figure 5(a) also shows the distribution of Tourists over time.

IV. Community detection algorithms

In the second part of this paper an effort is made to reveal the structure of the underlying social graph of the MySpace OSN, but also to characterize community size and density and to retrieve correlation in the activity among members of the same community. To this purpose a large number of communities were identified using two well established community finding algorithms, namely the Fortunato et al. [12], [13] and the Clique Percolation [4], [19]. These algorithms share the same main principal, which is a local approach as they are based solely on local topological properties. Both algorithms are known to be effective in identifying communities in large graphs, starting at seed nodes and utilizing only local interactions between nodes. These algorithms have a crucial advantage over the prevailing clustering methods that follow global approach [8], [18], demanding the manipulation of the whole network. Although global methods can be applied in synthetic networks and datasets for academic purposes, many challenges exist for real world datasets, let alone modern social networks that consist of hundreds of millions users. Before proceeding to the presentation of the results, we describe the two distributed algorithms used to find the hiding communities.

A. The Fortunato et al. algorithm

An interesting method for community detection appears in [13]. A local algorithm is devised developing a community from a seed node and expanding around it. A community is identified as a subgraph that has a certain *fitness*. The authors provide an appropriate fitness function, whose calculation is based on the number of inter- and intra-community edges and a tunable parameter a. Starting at a node, at each iteration, the community is either expanded by a neighboring node that increases the community fitness, or shrinks by omitting a node if this action results in higher fitness for the community. The algorithm stops when the insertion of any neighboring node would lower the fitness of the community. This algorithm is local, and able to identify individual communities.

For a community S of the graph, the fitness f_S is calculated as follows:

$$f_S = \frac{K_{in}^S}{(K_{in}^S + K_{out}^S)^\alpha} \tag{1}$$

where K_{in}^S and K_{out}^S refer to the total number of intra-community and inter-community edges, respectively, for community S, and α is a positive real-valued parameter which controls the size of the community. A larger value of α results in larger communities. A commonly used value is $\alpha = 1$.

The process of revealing the natural community of node A is described below:

1) a loop is performed over all neighboring nodes of S not included in S;
2) the neighbor with the largest fitness is added to S, yielding a larger subgraph S';
3) the fitness of each node of S' is recalculated;
4) if a node turns out to have negative fitness, it is removed from S', yielding a new subgraph S'';
5) if (4) occurs, repeat from (3), otherwise repeat from (1) for subgraph S''.

The process terminates when there is no other candidate node with positive fitness for inclusion in the community. The entire overlapping and hierarchical structure of complex networks can also be found, if the algorithms is initiated from various graph nodes.

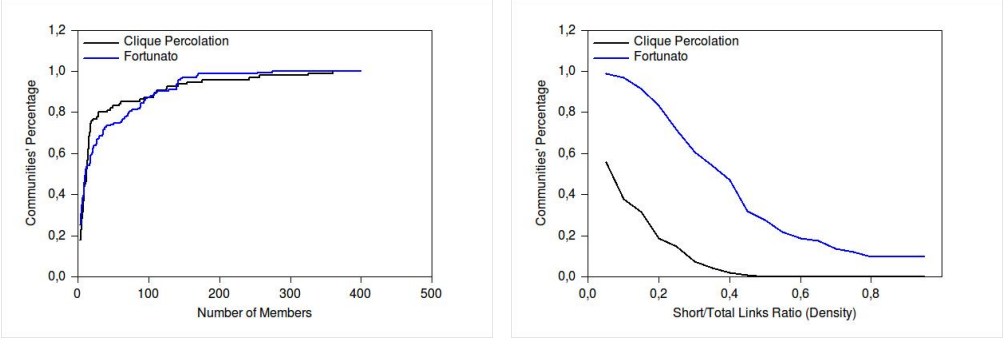
B. The Clique Percolation algorithm

Recognizing the importance of identification of overlapping community structures, Palla, Derenyi and Vicsek introduced Clique Percolation in 2005 [4], [19]. The method was applied in a variety of well established scientific disciplines extracting valuable information over network structures.

As the algorithm's name betokens, the method's cornerstone is the K-clique which correspond to a fully connected subgraph of K nodes. Two K-cliques are considered to be adjacent if they share K-1 nodes. A community is defined as the maximal union of K-cliques that can be reached from each other through a series of adjacent K-cliques.

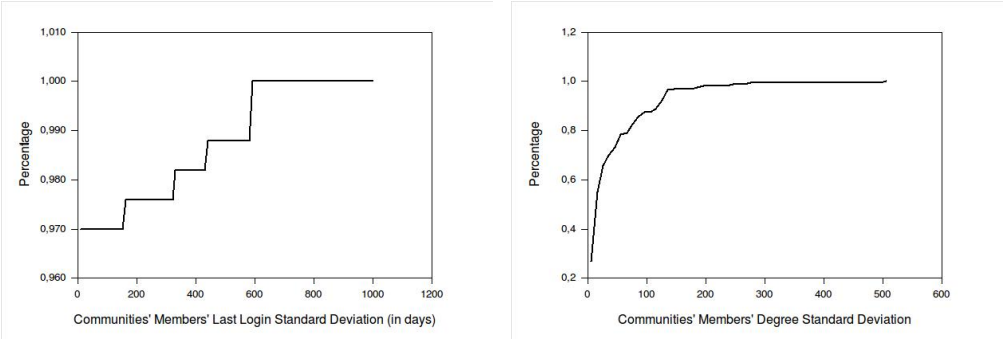
The whole procedure can be implemented with the help of a K-clique template which is an object isomorphic to a complete graph of K nodes. Such a template can be placed onto any K-clique in the graph, and rolled to an adjacent K-clique by relocating one of its nodes and keeping its other K-1 nodes fixed. Thus, a K-clique community of a network is a subgraph that can be fully explored by *rolling* a k-clique template in it.

The key in order to achieve acceptable results is the selection of the K parameter. Usually choosing K = 3 or K = 4 helps us extract valuable information, and currently these values of K have yielded, to our knowledge, the most relevant communities in practical applications. In our implementation K = 3 was chosen in order to maximize output communities' size and the possibility to successfully start the detection of the surrounding community from a random node.



(a) (b)

Fig. 6. (a) CDF of community size. (b) CDF of community density (x-axis is the ration of intra- to inter-community edges).



(a) (b)

Fig. 7. (a) Standard deviation of Last Login (b) Standard deviation of communities member number of friends (degree).

V. Experimental Results

Using Fortunato's algorithm to extract communities, we managed to successfully identify 171 different communities. A total of 6137 different profiles belonged to these communities, thus the average number of members per community is about 37. In order to extract the 171 successfully detected communities, we had to fetch a large number of nodes surrounding the communities. These were 18259 distinct nodes. Thus to summarize:

- Communities successfully detected - 171
- Total number of community members - 6132
- Average members in each community - 36.718
- Total number of profiles fetched - 18259

Using the Clique Percolation algorithm we identified a total of 201 communities. The total number of profiles that belonged to the communities is 3576, thus an average of about 18 members per community. In order to extract the 201 successfully detected communities, we had to fetch 176577 distinct nodes. To summarize: Thus to summarize:

- Communities successfully detected - 201
- Total number of community members - 3576
- Average members in each community - 17.791
- Total number of profiles fetched - 176577

A first question we are trying to answer is "What type of communities do nodes form?" in the MySpace OSN. We proceeded with analyzing the surrounding community of a series of randomly selected profiles identified with the two algorithms, Fortunato's (with α=1) and Clique Percolation (with K=3). Figure 6(a) shows the CDF of the community sizes identified with the two methods. It is interesting to mention that Fortunato's algorithm identifies larger communities compared to the Clique Percolation method. Figure 6(b) depicts the CDF of the community density for the two methods. Density is defined as the ration of the number of intra- to inter-community edges (mentioned in the figure as short/ total links ratio). The more this ratio approaches one, the more the specific community is dense and thus tends to be an independent community structure.

As shown in the plot, in case of Clique Percolation, 60% of the detected communities have a greater than 0.05 short/ total ratio value. For density values greater than 0.10, percentage drops to about 40%. At the same time, members of communities detected by Fortunato's method, seem to be more closely connected achieving density values as high as 0.80. Moreover, almost half of the detected communities are characterized by a very satisfying 0.40 density ratio. Clique Percolation's Achilles' heel comes as a result of the conditions that needs to be met in order to be possible for a node to participate into a community. A node's participation in an adjacent triangle alone is sufficient, regardless of the risk it carries to have disproportionately more edges leading to nodes outside the community. Apart from the qualitative difference that density represents, output also differs in terms of communities' size. Although difference is not as striking as in the case of

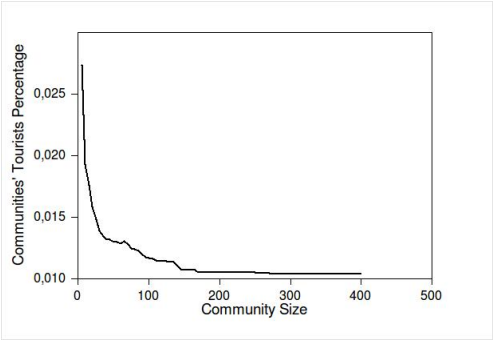


Fig. 8. Percentage of Tourists in communities.

density, we can still observe a difference in small to medium communities, counting a few dozens nodes as shown in Figure 6(b).

The next, less trivial questions we are trying to answer are "Do inactive users tend to abandon MySpace individually or the same tendency exists for their surrounding community?" and "Do tourists participate in community structures?". Taking into account the previously mentioned differences and considering the fact that Clique Percolation demands more information to be fetched in order to reveal the surrounding community of a node, we continued our analysis based on Fortunato's results.

In order to answer the first question we focus our efforts into locating members that have not given any signs of life for more than one hundred days. According to the collected data, the majority of community members is found to be satisfyingly active. It is characteristic that more than 85% of the communities consist of almost 90% of users that showed their presence in the network at least once in the last one hundred days. The vast majority of the inactive users are either tourists whose short stay period did not allow them to get connected with other profiles, or users with a very high degree. Both types are not favored by the different community definitions nor by the majority of the existing community detection methods which are operating by trying to maximize community's density and are thus eliminated. This can be also confirmed by the low to moderate average members' degree which characterizes the plurality of the communities. More interestingly, our revealed community structures show a significant homogeneity in members' level of activity. Figure 6(d) depicts the cumulative distribution for each communities' member's difference between the Last Login date and the date experiment took place (Fetch Date), expressed in days. As shown in the relevant Figure, more than 97% have a standard deviation value lower than 10. Equally impressively, the members of the detected communities show a significantly low dispersion over the total degree. The maximum standard deviation value is lower than 150. Moreover 60% of the communities have a standard deviation lower than 20. When percentage increases by another 20%, standard deviation increases by 30. In conjunction with the previous observation this demonstrates an important correlation between the nodes. Although we have only indications, all information provided above permits us to make the assumption that the participation of a user in a community, inhibits him to abandon MySpace, regardless of the existence of more attractive social networks.

Concerning the question "Do tourists participate in community structures?", Figure 8 clearly shows the very low participation of tourists in communities. Users that have been active for such a small period of time usually have a very low number of friends and are not favored by the different community definitions nor by the majority of the existing community detection methods which are operating by trying to maximize community's density.

VI. Conclusions

We performed an empirical study of the MySpace OSN, collecting not only a large number of user profiles in order to characterize population synthesis, evolution in user population, and user activity, but also extracting a large number of community structures in order to analyze the underlying social network graph and to correlate characteristics and activity of users participating in communities. Two local algorithms were used to extract communities, namely, the Fortunato and the Clique Percolation methods.

The main conclusions can be summarizes as follows: Population synthesis changed over time, as the difference between public and private profiles shrunk, and invalid profiled overpassed first private ones and subsequently the public ones. Use activity was shown to be diminishing, while the number of new users joining the network remains high.

Based on the analysis of the communities, important conclusions were drawn on the size of the communities that are formed in the network, but also in the community density as shown by the number of edges among participating member. As a last and most important result, strong correlation was demonstrated in the Last Login date among members of the same community, and other attributes like the number of friends. We draw the conclusion that participation in *strong* communities inhibits users from abandoning the OSN in order to migrate to other more popular networks that emerge at times. As a last observation, we notice that members that abandoned MySpace shortly after their account creation (the so-called Tourists), have very low connectivity and thus they do not participate in communities.

REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, M. S., and H. Jeong. Analysis of Topological Characteristics of Huge Online Social Networking Services. May 2007.

[2] J. P. Bagrow and E. M. Bollt. Local method for detecting communities. *Physical Review E*, 72(4):46–108, Oct 2005.

[3] B. Buter, N. Dijkshoorn, D. Modolo, Q. Nguyen, S. van Noort, B. van de Poel, A. Ali, and A. Salah1. Explorative visualization and analysis of a social network for arts: The case of deviantart. *Journal of Convergence Volume*, 2(1), 2011.

[4] I. Derényi, G. Palla, and T. Vicsek. Clique Percolation in Random Networks. *Physical Review Letters*, 94(16):160–202, Apr 2005.

[5] A. Dutta, I. Ghosh, and D. Mukhopadhyay. An advanced partitioning approach of web page clustering utilizing content & link structure. *Journal of Convergence Information Technology*, 4(3):65–71, 2009.

[6] Facebook. *http://www.facebook.com*.

[7] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee. Self-organization and identification of web communities. *IEEE Computer*, 35:66–71, March 2002.

[8] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, June 2002.

[9] P. F. Harris Papadakis and C. Panagiotakis. Distributed community detection: Finding neighborhoods in a complex world using synthetic coordinates. In *IEEE symposium on Computers and Communications*, pages 1145–1150, 2011.

[10] D. Katsaros, G. Pallis, K. Stamos, A. Vakali, A. Sidiropoulos, and Y. Manolopoulos. Cdns content outsourcing via generalized communities. *IEEE Transactions on Knowledge and Data Engineering*, 21:137–151, 2009.

[11] R. Kumar, J. Novak, and A. Tomkins. Structure and the Evolution of Online Social Networks. August 2006.

[12] A. Lancichinetti and S. Fortunato. Community detection algorithms: a comparative analysis. *Physical Review E*, 80(5 Pt 2):056117, Sep 2009.

[13] A. Lancichinetti, S. Fortunato, and J. Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015+, March 2009.

[14] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. August 2005.

[15] M. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. October 2007.

[16] W. W. Mojtaba Torkjazi, Reza Rejaie. Hot Today, Gone Tomorrow: On the Migration of MySpace Users. *WOSN'09, Barcelona, Spain*, pages 43–48, August 17 2009.

[17] MySpace. *http://www.myspace.com*.

[18] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, Feb 2004.

[19] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society, June 2005.

[20] G. Pallis, D. Zeinalipour-Yazti, and M. D. Dikaiakos. Online Social Networks: Status and Trends. *New directions in Web Data Management*, 1, 2011.

[21] S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27 – 64, 2007.

[22] S. Van Dongen. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.*, 30:121–141, February 2008.

[23] F. Wu and B. A. Huberman. Finding communities in linear time: a physics approach. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):331–338, March 2004.

[24] Y. Ye, X. Li, B. Wu, and Y. Li. A comparative study of feature weighting methods for document co-clustering. *International Journal of Information Technology, Communications and Convergence*, 1(2):206–220, 2011.