

Studying micro-communities in Facebook Communities

Andrea De Salve

Institute for Informatic and Telematic,
CNR
Pisa, Italy
andrea.desalve@iit.cnr.it

Barbara Guidi

Department of Computer Science,
University of Pisa
Pisa, Italy
guidi@di.unipi.it

Andrea Michienzi

Department of Computer Science,
University of Pisa
Pisa, Italy
andrea.michienzi@di.unipi.it

ABSTRACT

In the visionary view of the future Internet, named the Next Generation Internet, a current idea is to have a user-centric approach where human behavior models will be used to define the networks or to manage services. During the last years, a great trend in current Social Media platforms is to offer the opportunity to establish and join groups of people online. Despite human behaviour in current Online Social Media have been studied in depth, characteristics of these aggregations of people in content-based communities are still unknown. In this paper, we propose an evaluation of micro-communities of users inside the big network of Facebook groups to understand how and when users are active, and to evaluate the evolution of these micro-communities over time. Results show that almost all groups showed interactions-based communities. We found out that in all cases there is one massive core community which attracts small communities.

CCS CONCEPTS

• **Human-centered computing** → **Social networks; Social media; Social network analysis;**

KEYWORDS

Complex Networks, Social Networks Analysis, Community Detection, Facebook Groups Analysis

ACM Reference Format:

Andrea De Salve, Barbara Guidi, and Andrea Michienzi. 2018. Studying micro-communities in Facebook Communities. In *International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18)*, November 28–30, 2018, Bologna, Italy. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3284869.3284904>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Goodtechs '18, November 28–30, 2018, Bologna, Italy

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6581-9/18/11...\$15.00

<https://doi.org/10.1145/3284869.3284904>

1 INTRODUCTION

In the visionary view of the future Internet, named the Next Generation Internet, a current idea is to have a user-centric approach, where the decentralization of services will be the first goal¹. In this wide scenario, Social Media will be one of the main topics². Indeed, the Internet of People (IoP), a new paradigm introduced in [2, 3, 12], confirmed the trend of the future Internet to define a human-centric approach where human behavior models will be used to define the networks or to manage services. As just introduced in [5, 6, 14], as concerns as the Online Social Networks, the study of human behaviour is an important step to evolve to a new generation in which services are decentralized and where there is a high level of privacy.

During the last years, a trend in current Social Media platforms is to offer the opportunity to establish and join groups of people online. People join together in groups for several reasons, such as offline relationships (e.g. family groups), job opportunities (e.g. work groups), or for simple common interests and hobbies. Despite human behaviour in current Online Social Media have been studied in depth, characteristics of these groups of people are still unknown, due to the lack of data. Indeed, groups, such as in Facebook, can have different visibility status: public or private (only visible to members). Public groups do not represent real content-based communities due to the free join opportunity which can be exploited by people who are not interested enough. Instead, private groups, which usually require a check by admins, represent an interesting scenario that is unexplored.

In this paper, we propose an evaluation of social groups by studying a set of Facebook groups. The aim consists of studying the presence of micro-communities of users inside the big network of a Facebook group to understand how and when users are active. A micro-community can be defined as a dense portion of active users tightly connected, discovered by exploiting the graph of the interactions of each group. In particular, we analyse the characteristics of communities discovered by using the User Interaction Temporal Graph

¹<https://ec.europa.eu/digital-single-market/en/policies/next-generation-internet>

²<http://www.eismd.eu/next-generation-internet-summit/>

(UITG), and we study the evolution of these community over time. Results show that Facebook groups are characterised by a single massive community which tends to attract smaller communities.

The paper is organized as follow. In Section 2 presents related works on Online Social Network (OSN) analysis and Online Social groups. In section 3 we introduce the scenario and give a definition of our problem, while in section 4 we present the dataset and the analyses we performed on it. We conclude the paper in section 5 presenting some possible future works.

2 RELATED WORK

Current Online Social Media provide to users the possibility to form communities and sub-communities. As explained in [11], two major types of groups exist. The first one consists in self-created groups by sorting social contacts and place them (buddies, friends, or followers) into these sets. The second can be analogous to clubs in the offline world: open to anyone, closed, or secret. Facebook groups, for example, can be considered a kind of community based on similar interests where people can know or not the others members. Even if Online Social groups have been studied in several fields, such as in the medical field [13], they are still unknown in terms of their graph properties. An important concept in Social Network Analysis is community. A community does not have a concrete definition, but it is general defined as a group of nodes tightly connected with each other than with nodes belonging to other sets [8]. Community detection algorithms serve to identify groups of nodes in a network. Several algorithm have been proposed during the years [7]. The main difference consists of the definition of community. In [9, 10], authors demonstrated the need of community analysis in Online Social Networks in order to develop specific solutions that can be applied in the Next Generation Internet, by studying the ego networks of Facebook users. As concerns the combination of Facebook groups and the community concept, a recent work [15] analyses the community in a Facebook group by focusing on communities. Indeed, authors identify the dynamic communities detected in the Facebook group, based on the sentiment expressed in the posts during a certain period of time.

3 DEFINITIONS AND PROBLEM STATEMENT

By taking into account the activity of users inside a group we can define the User Interaction Temporal Graph (UITG). The UITG is a multigraph where nodes are the member of the group and edges are the interactions between them. Temporal information in the UITG is embedded directly on the edges using labels. Each edge, in fact, is labeled with the time at which the interaction took place. Edges in the UITG also have direction. The source of the edge is the user

that started the interaction and the destination of the edge is the user to whom the interaction is referring to. For the sake of our analyses, since we are interested in building groups of users that interact each other densely, we do not consider the direction of the edges. We built two versions of UITG by considering its temporal nature: the static UITG (s-UITG) and the incremental UITG (i-UITG). In the s-UITG we completely ignore all temporal information and we consider all the interactions at the same time. On the other hand, in the i-UITG at time t we find all the interactions with temporal label $\leq t$. Trivially, s-UITG = i-UITG at each $t \geq$ the time of the last interaction. While s-UITG provides a static general view of the interactions, i-UITG is able to provide the evolution of interactions through time, up to the final state.

Given this graph definition, we analyse and try to understand how social groups of people interact and if there are common characteristics. Community detection is exploited to search for dense zones in the UITG. Dense zones in the UITG represent groups of users which interact a lot with each other thus forming micro-communities within themed groups. As community detection tool, we use DEMON [4] because of its low time complexity.

4 DATA COLLECTION AND ANALYSIS

We start this section by briefly presenting the dataset we used for the analyses, then we move to the analyses performed. In particular, the information collected from the Facebook's groups are obtained by exploiting an HTTP-crawler that based on Selenium automates browsers library which, given a Facebook group G , periodically retrieve the following set of information:

- **Members** we retrieve the member list of the group;
- **Interactions** we collect interactions occurred between members of the groups, such as posts, comments, reply, likes to post, likes to comments, and reactions.

In particular, we were able to retrieve the activity occurring on the groups and the associated timestamps, which are useful to capture the evolution of the groups. The information collected about groups are periodically retrieved, stored in XHTML format, and parsed to extract relevant information. We selected 10 distinct groups which belong to three heterogeneous categories having different characteristics. Table 1 summarizes the main characteristics of each group by showing: the group's category (*Category*), the number of consecutive days on which our crawler collected the interactions (*Days*), the number of group's member (*Members*), the date of both the first (*min Date*) and the last post (*max Date*) retrieved by our crawler, and the total number of posts retrieved from the group in the monitored period (*Posts*)

Group	Cat.	Days	Users	min Date	max Date	Posts
Edu1	Edu	388	10,643	01/01/17	24/01/18	3555
Edu2	Edu	317	46,016	06/04/17	18/02/18	5271
Edu3	Edu	393	21,195	25/01/17	22/02/18	5060
Sport1	Sport	249	35,671	27/08/17	03/05/18	5588
Sport2	Sport	370	3,589	04/02/17	09/02/18	708
Sport3	Sport	28	107,459	13/02/18	14/03/18	6353
Work1	Work	406	26,901	02/01/17	12/02/18	1444
Work2	Work	418	4,925	04/01/17	26/02/18	945
Work3	Work	318	25,257	13/06/17	27/04/18	4809
Work4	Work	485	12,151	03/01/17	04/05/18	2651

Table 1: General description of the Facebook groups

Our crawler was able to collect the activities performed by members over a period of 365 days (i.e., 1 year) only for some groups (such as Edu1, Edu3, Sport2, Work1, Work2, and Work4). Indeed, groups having higher activity can overload the crawler due to the excessive number of interactions performed on some groups (such as, the Sport3 which has about 226 posts per day).

In general, the sizes of the groups are very heterogeneous and the maximum number of members of the selected groups does not exceed 107,459 users while the group Sport2 has the fewest number of members (3,589). The groups expose high activity level and the number of posts collected during the monitored period is higher than 4,000 for the majority of the groups (about 60%). As for instance, more than 6,000 posts have been published on group Sport3 during the monitored period of 28 days.

Communities in the s-UITG

We started our study by analysing the community structure, extracted using DEMON[4], of the s-UITG. Since each group has very peculiar characteristics in terms of number of users inside, number of posts/comments/replies per day, we decided to study each group separately. Table 2 contains the number and some statistical measures of static communities in each group. Values are aggregated by group, one group for each row of the table; the last row contains a recap where communities of all groups are aggregated together. From this table it is clear that the community structure in the groups is very heterogeneous, even among groups belonging to the same thematic area. In particular, we observe that the number of communities ranges between 8 and 139. Communities vary a lot also in size. The size of the biggest community for each group ranges between 14 and 5364. Low, with respect to maximum, mean values and high standard deviation values suggest us that the distribution of the size of the communities follows a power-law. This is also confirmed by the Cumulative Distribution Function (CDF) of the size of all

Group	Num	Min	Max	Mean	SD
Edu1	18	6	2005	575.0	613.5
Edu2	129	5	2626	241.2	394.2
Edu3	36	4	1494	339.3	417.6
Sport1	52	5	5364	617.9	1024.1
Sport2	24	4	76	19.4	20.5
Sport3	139	4	2766	165.6	401.0
Work1	69	4	684	92.9	141.5
Work2	10	4	848	313.0	328.7
Work3	8	4	14	6.75	3.5
Work4	8	5	15	8.5	4.0
Recap	493	4	5364	241.3	503.2

Table 2: Statistical measures on number and size of communities detected by DEMON in the s-UITG

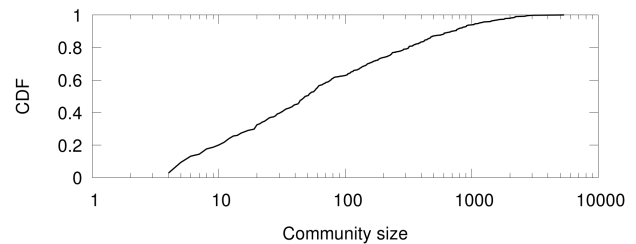


Figure 1: CDF of the size of all the communities extracted from the s-UITG

the communities in the s-UITG graphs of all the groups (Figure 1). Noticing that the x-axis of the figure has logarithmic scale, we see that half of the communities barely reach $\frac{1}{100}$ of the size of the largest community. Comparing tables 1 and 2 there is no clear relation between the number of users in the groups and the size of communities. Instead, there is a general trend in having more communities in larger groups. The groups Work3 and Work4 have instead a big number of users but a low number of communities which are also small in size. This is due to the fact that users in these groups had very few interactions over the observation time span.

Communities in the i-UITG

For what concerns the main contribution of this paper, we decided to investigate on how the community structure evolved over time in each group. To do so, we focused on the i-UITG. To be able to perform some analyses using the same tools we used for the s-UITG, we performed observations on the i-UITG at regular time intervals. These observations yielded static graphs containing all interactions happened up to a given point in time. A reasonable and significant time interval for our observations is 24 hours for two main reasons. The first, obvious, reason is that setting a too short time interval would produce observations of the evolving graph that

are extremely similar. A short time interval requires many interactions to happen in short time in order to see an evolution of the i-UITG. The second reason is connected to the fact that usually people's life follows a routine which lasts 24 hours. Therefore, setting a different time interval would not match typical usage patterns of the OSN. For the sake of clarity, the i -th observation contains all the interactions up to, and including, the i -th day. Before observing the community structure on the i-UITG, we made a preliminary study on the number of interactions of each group. Figure 2 shows the number of interactions for each observation for each group, divided by thematic area: 2(a) for the Education Groups, 2(b) for the Sport Groups, and 2(c) for the Work Groups. For the sake of clarity, this is not the number of new, unseen, interactions per day. These are the total amount of interactions happened per day, not considering that the same interaction (i.e. same source user and same destination user) may be happened a previous day, or even multiple times in the same day. In all the groups the number of interactions is, overall, quite stable. There seems to be no correlation with the number of users inside the group. Nevertheless, it seems that groups in the Sport thematic area contain users that are more active with respect to the other groups. Or let's consider, for instance, the group Sport1: it counts 35,000 users and have a little less than 1,000 interactions per day, versus the group Education2 which has 46,000 users but is far from 1,000 interactions per day.

To study the evolution of communities, we apply DEMON to each observation. Figure 3 shows the number of communities over time for each group, divided by thematic area. A first, surprising result is the fact that in most of the cases we have a rapid growth in the number of communities during the first days of activity of the groups. After this rapid growth, lasting 50 days on average, the number of communities remains stable (Edu1, Edu3, Sport1, Sport2, Work1 and Work2). This is suggesting us that people interacting in Facebook groups tend to join or to be included in pre-existing interaction groups rather than forming new ones. Interestingly enough, this trend is followed both by groups for which we were able to gather all interactions since the creation of the group and by groups for which we were able to gather interactions up to an arbitrary date. The only group not strictly following this trend seems to be Edu2. The group has a very steep growth in number of communities at the beginning of the observations (days from 100 to 150), as expected. But after the initial growth the number of communities still grows at a decreasing pace until the end of observations (right after day 400). What we suspect is that in this group we have a second long phase before stabilization during which the growth of the number of communities slows down. A similar trend can be observed in the group Work1. Mostly likely, the

group Sport3 is following a similar trend and the observation ended right during this second phase.

Having seen that the number of communities tend to remain stable during a group life, we are now interested in the nodes that compose such communities. To this aim, we computed the number of active nodes in each group. A node is said to be active if it belongs to at least one community. Figure 4 shows the number of active nodes for each group, divided by thematic area. From these plots we can see in all cases the number of active users grows over time. In most of the cases the growth is linear (Edu3, Sport2, Sport3, Work2, Work3, and Work4), but in some cases it is well-structured: Edu2 shows again a two phases growth, Work1 seems to follow a logarithmic-like growth, Sport1 has a unique shape. The ever-increasing number of active nodes is surprising if compared to the fact that the number of communities remain constant after the initial growth. Indeed, these plots show that, nevertheless the number of nodes in the graph increases, these nodes tend to join existing communities.

Since we are interested in how interaction groups evolve over time, we also evaluated how communities grow over time. In our study we extracted communities independently from each observation of the i-UITG, so there was no clear and easy way to determine the evolution of the communities. Therefore matching communities from adjacent observations a posteriori using a certain rule was not a good idea. In fact, we observed that in our case the match was extremely difficult and inaccurate due to the fact that community evolve in unexpected ways. Using some techniques, we observed communities shrinking in size or even split in adjacent observations. Recalling the fact that the i-UITG is built incrementally, this was clearly some counterintuitive behaviour because we expected communities to grow and merge as more interactions are added. So, instead of trying to match communities, we evaluated how many new active nodes a community has gained compared with the communities in the previous observation. For this evaluation, we define the *growth rate* as follows:

$$GR(C_i^j) = \frac{|C_i^j|}{|AN(C_i^j)^{j-1}|} - 1 \quad (1)$$

where C_i^j is the i -th community at the j -th observation and $AN(C_i^j)$ is the set of active nodes of the community C at the j -th observation. For the sake of clarity, we defined the growth rate of a community as the ratio between the size of that community and the number of nodes that were active in the previous observation, minus 1 so that if the grow rate of a community is 0 means that no new nodes were attracted by that community. Our definition of growth rate resembles the one given in [1], from which we took inspiration. The

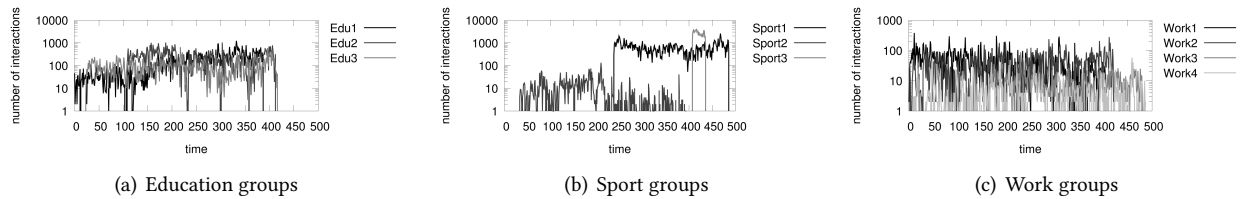


Figure 2: Number of interactions in the i-UITG

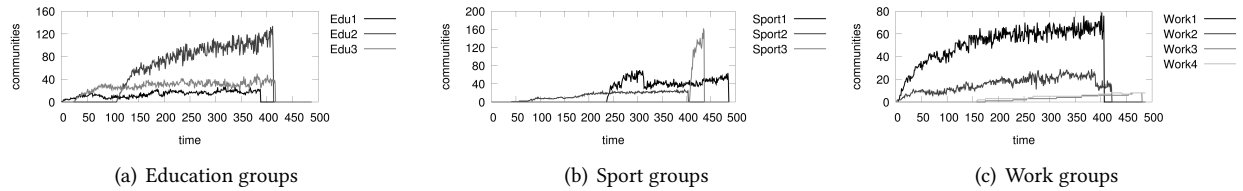


Figure 3: Number of communities in the i-UITG

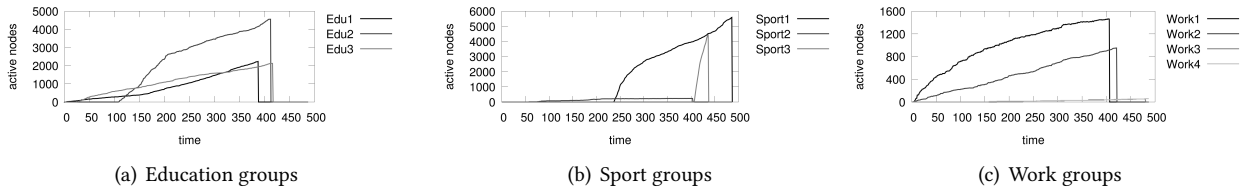


Figure 4: Number of active nodes (involved in at least one community) in the i-UITG

main differences is that they do not consider communities smaller than 100 nodes in their computation, due to high instability of small communities, and that they know exactly the evolution of each community. Figure 5 shows the average growth rate of the communities for each group, divided by thematic area. All groups in all categories show a very similar behaviour, with an extremely high growth rate when the first interactions in the group appear and then a constant, very low, growth. Edu3, Sport1, and Sport3 have the most emphasised peaks, while Edu1, Work2, Work3 and Work4 have the least emphasised ones. This behaviour suggests us that the important communities of each group forms right at the beginning of the history of the groups. Then, these big communities continuously grow during the group life, but at a very slow pace.

Finally, since we are interested in the main actors of the groups, we studied how the size of the biggest community evolve over time. Again, it was not possible to safely match communities detected in adjacent observation, so we assume that the biggest community is the same during all the history of the group. Figure 6 shows the size of the biggest community for each group, divided by thematic area. Again, we see a similar behaviour in all groups which resembles the one of the number of active nodes (Figure 4). The size of the biggest community almost matches the number of active nodes in the following groups: Edu1, Edu3, and Work2. Nevertheless,

in the other groups, the size of the biggest community is always much more than half of the active nodes. This fact tells us that each group has a single huge community made of a large portion of users in that group. This community is the core of the group and other smaller groups and new users are attracted to it as they make interactions in the group. A peculiar characteristic is that the size of the biggest community is quite unstable, dropping and rising at each observation. What happens is that, thanks to some users making their first interaction, the biggest community may be initially splitted into few shards, one of them sensibly bigger than the others. This splitting is caused by the nature of the community detection tool used, which is based on the label propagation algorithm. After few observations, the new users interact with more other users and the shards of the old biggest community join together to make up an even bigger community.

5 CONCLUSIONS AND FUTURE WORK

In this paper we studied the behaviour of social groups members by studying the interactions and the communities they build based on their interactions in heterogeneous groups. We carried out our analyses on 11 Facebook groups with the only requirement of them showing some activity. Except in one case, all groups showed interaction based communities. When studying how these communities form and evolve

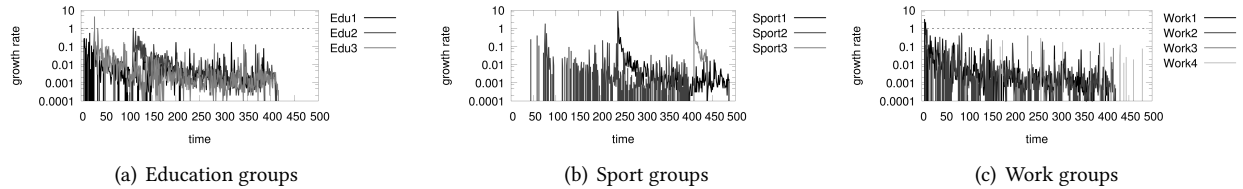


Figure 5: Growth rate of the communities in the i-UITG

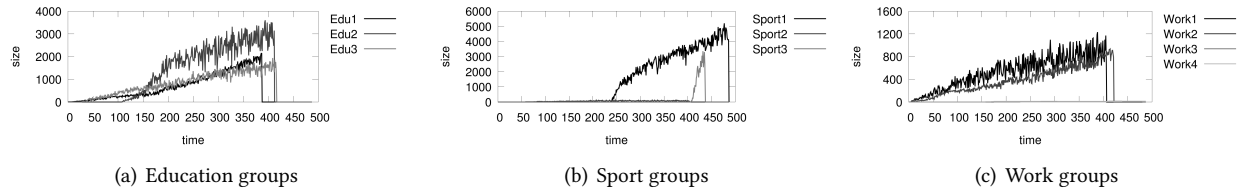


Figure 6: Size of the biggest community in the i-UITG

over time, we found out that in all cases there is one massive core community. This community forms right at the very beginning of the history of the group and explodes in size in a very short time. After the first explosion, the core community keeps growing but at a much slower pace, attracting both existing smaller communities and users new to the group.

As future work we plan to study more in detail the community structure using more sophisticated techniques. In particular, we plan to study communities in a dynamic fashion, giving interactions a lifespan. In this way we aim to study micro-communities at a finer temporal grain. What we expect to see is that the massive community has a relatively small core of users which are the ones that keep the community, and therefore the group, alive.

REFERENCES

- [1] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, 44–54.
- [2] Marco Conti and Andrea Passarella. 2018. The Internet of People: A human and data-centric paradigm for the Next Generation Internet. *Computer Communications* (2018).
- [3] Marco Conti, Andrea Passarella, and Sajal K. Das. 2017. The Internet of People (IoP): A new wave in pervasive mobile computing. *Pervasive and Mobile Computing* 41 (2017), 1 – 27.
- [4] Michele Coscia, Giulio Rossetti, Fosca Giannotti, and Dino Pedreschi. 2014. Uncovering Hierarchical and Overlapping Communities with a Local-First Approach. *ACM Trans. Knowl. Discov. Data* 9, 1, Article 6 (Aug. 2014), 27 pages.
- [5] Andrea De Salve, Marco Dondio, Barbara Guidi, and Laura Ricci. 2016. The impact of user's availability on on-line ego networks: a facebook analysis. *Computer Communications* 73 (2016), 211–218.
- [6] Andrea De Salve, Barbara Guidi, Laura Ricci, and Paolo Mori. 2018. Discovering Homophily in Online Social Networks. *Mobile Networks and Applications* (2018), 1–12.
- [7] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3-5 (2010), 75–174.
- [8] Santo Fortunato and Marc Barthélemy. 2007. Resolution limit in community detection. *Proceedings of the National Academy of Sciences* 104, 1 (2007), 36–41.
- [9] Barbara Guidi, Andrea Michienzi, and Giulio Rossetti. 2018. Dynamic Community Analysis in Decentralized Online Social Networks. In *Euro-Par 2017: Parallel Processing Workshops*. Springer International Publishing, 517–528.
- [10] Barbara Guidi, Andrea Michienzi, and Giulio Rossetti. 2018. Towards the Dynamic Community Discovery in Decentralized Online Social Networks. *Journal of Grid Computing* (2018), 1–22.
- [11] Jan H Kietzmann, Kristopher Hermkens, Ian P McCarthy, and Bruno S Silvestre. 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons* 54, 3 (2011), 241–251.
- [12] Javier Miranda, Niko Mäkitalo, Jose Garcia-Alonso, Javier Berrocal, Tommi Mikkonen, Carlos Canal, and Juan M Murillo. 2015. From the Internet of Things to the Internet of People. *IEEE Internet Computing* 19, 2 (2015), 40–47.
- [13] Stephanie R Partridge, Patrick Gallagher, Becky Freeman, and Robyn Gallagher. 2018. Facebook groups for the management of chronic diseases. *Journal of medical Internet research* 20, 1 (2018).
- [14] Andrea De Salve, Barbara Guidi, and Laura Ricci. 2018. Evaluation of Structural and Temporal Properties of Ego Networks for Data Availability in DOSNs. *Mobile Networks and Applications* 23, 1 (01 Feb 2018), 155–166.
- [15] Laura Sani, Gianfranco Lombardo, Riccardo Pecori, Paolo Fornaciari, Monica Mordonini, and Stefano Cagnoni. 2018. Social Relevance Index for Studying Communities in a Facebook Group of Patients. In *Applications of Evolutionary Computation*. Springer International Publishing, 125–140.