2012 IEEE 27-th Convention of Electrical and Electronics Engineers in Israel

An Analysis of The Steam Community Network Evolution

Roi Becker, Yifat Chernihov, Yuval Shavitt and Noa Zilberman
School of Electrical Engineering
Tel Aviv University
Email: shavitt,noa@eng.tau.ac.il

***Abstract*—The Steam community network is a large social network of players on the Steam gaming platform, with over 30 million users to date. In this paper we introduce an analysis of the Steam community network in 2011, looking at the characteristics of the users network and the connectivity graph. We next present the evolution of the network over time and show how the network has changed over the years. Last, we analyze the role of games and groups in the Steam community. This work is the first to analyze the Steam network, and to provide a large scale analysis of the characteristics of gaming platforms communities.**

I. Introduction

Online gaming is a multi-billion industry with tens to hundreds of millions of users around the world [15], [18]. One of the compelling features of these games is sharing them with friends as well as making new friends through them. For this reason, several of the main online gaming platforms have a social network revolving around the games. While there is a considerable study of social networks over the last years, little attention was paid to social networks of online games from an engineering perspective.

A large portion of the works done on online gaming communities revolved around social networks's games and specifically Facebook, most notably Nazir *et al.* [13] seminal work on network characteristics in Facebook's online games community. Balint *et al.* [3] examined the structure of an online bridge gaming community and compared it to a face-to-face bridge community structure. Kegan *et al.* [7] studied gold farming in online gaming, focusing on the EverQuest II game. They studied aspects of the gold diggers network in terms of centrality, degree, clustering and more and reach the conclusion that it is much like the structure of drug traffic networks.

Steam [16] is a digital distribution multiplayer and communications platform developed by Valve Corporation. It distributes online games, both from small independent developers to larger software houses. It was revealed to the public in 2002, and has over 30 million registered users to date. It is currently ranked by Alexa [1] among the top 10 gaming websites. The Steam Community is comprised of all Steam users. When first installing Steam, one must open a user account, which results in joining the Steam community. The community also consists of groups which users can open and join based on common interests, such as shared games or friendship connections. Each group has a public web page which, except for some rare occasions, can be viewed by anyone surfing the web. Quite a lot of works have studied the Steam network, however their focus was either financial [8], gaming characteristics [12], [6] or platform performance [2], [5]. Blackburn *et al.*[4], who analyzed cheaters activity in the Steam network, provided some high level analysis of the network, mainly in the context of cheaters' friendship connections.

The contribution this paper is threefold. We first introduce an analysis of the Steam community network in 2011, looking at the characteristics of the users network and the connectivity graph. We next present the evolution of the network over time. Last, we analyze the role of games and groups in the Steam community.

II. Dataset

The Steam community is based on publicly accessible webpages: every user in the Steam community has a webpage dedicated to his profile, as well as webpages dedicated to groups and games. A user's webpage includes information about the user: profile name and real name, origin country, Steam ranking, membership date and playing time. For every user there is a dedicated webpage that lists all his friends and the date they became friends. Every user also has a privacy setting: public, private, not opened yet, and closed. A private profile does not allow access to the users personal info nor to his friends page. A user's profile can be defined as not opened yet if the user did not add any personal information to his profile. Every user has a games page, listing all the games that he has been playing, and a groups page, listing all the groups he is a member of.

Each registered user can open a Steam group that can be joined by other users. Groups can be created around shared interests, friendship relationships etc. and are pivotal in organizing events, most often game dates. Much like users profiles, so do groups have a privacy status: private, public, official, and closed. Official groups are public groups that were opened by the Steam network operators, usually around a game. These groups tend to have a very large number of members. A group's webpage includes information about the game (name, announcements, number of members etc.) as well as a list of its members.

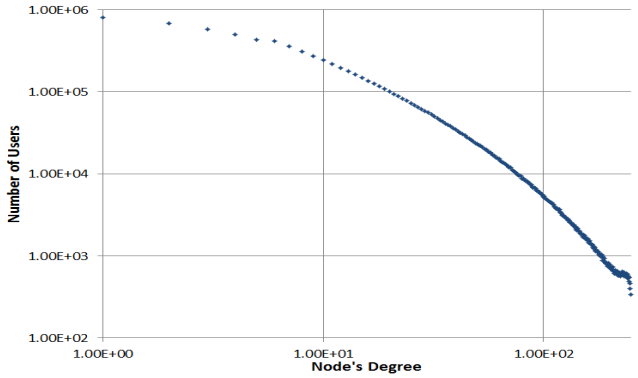

Fig. 1: Steam Community Node Degree Distribution

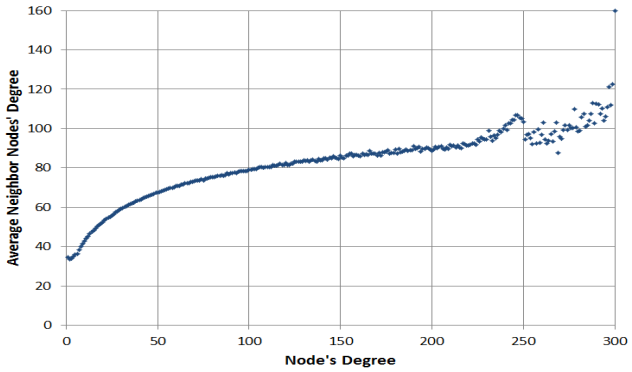

Fig. 2: Average Neighbor Node's Degree to Node's Degree

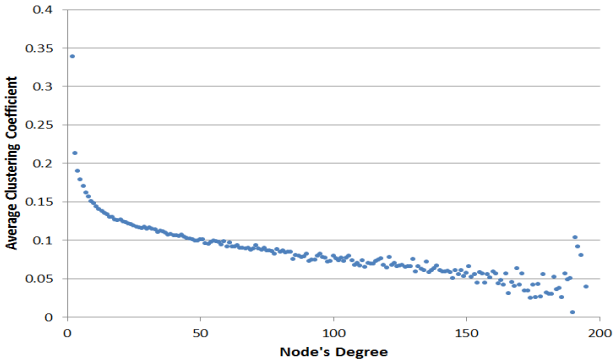

Fig. 3: Average Clustering Coefficient to Node's Degree

The last type of webpages that we crawl are dedicated for games. Each game available to purchase through Steam has a web page describing the game, its platform requirements, key features, and more.

The Steam community dataset was acquired by crawling the network during October, 2011. The crawl covered all of Steams groups and games that are publicly listed by Steam. Users webpages were accessed through one of three sources: group member's pages, friends pages of other users and through user ID seed number. The resulting dataset covers over nine million users and 82.2 million friendship edges. There were 1824 games at the time of the crawl and over 1.98 million groups. Friendship bonds' information is available starting September, 2008 and membership information starting 2003. This information allows us to track the network's evolution over time, since typically friendship connections are not removed so we miss a negligible connection from early dates.

III. Network Analysis

We study the Steam network characteristics, starting with the basic ones. Every user in the community is referred to as a node, and every friendship connection represents an edge. The graph is clearly very sparse: only 82 million edges are discovered, for the 9 million nodes, which results in an average degree of 18.2, meaning that on the average user has about eighteen friends in the Steam community. Figure 1 shows the node degree distribution. The graph largely behaves by Zipf law, except for high degrees: the Steam network limited the number of friends per user to 250 until 2011, and is currently limiting it to 300. Looking at the shape of the graph, high degrees are indeed distorted as instead of a long tail, there are clusters at 250 and 300 users. Since only a few months have passed since the friends limit was increased and until the data was sampled, a degree of 250 is still dominant.

In a community, people tend to connect to other people similar to them. In many social networks it is, thus, seen that high degree nodes tend to connect to other high degree nodes, a term called *assortativity*. Other networks, like the Internet AS graph, are disassortative, as high degree nodes there tend to connect to small degree nodes [14]. To check assortativity in the Steam network, Figure 2 depicts the average neighbor node's degree to node's degree distribution. The clear monotone increase in the graph for node degree 1 to 250 shows assortativity as expected. The discontinuity at 250 is due to the forced limit on the node degree which was lifted too close to our crawling

Another important property of a social graph is a high clustering coefficient, which represents the cliquishness of a typical neighborhood [20]. A clustering coefficient of a node in the network is defined by the ratio between the number of triangles among its immediate neighbors and the maximum possible number of such triangles. The higher the clustering coefficient is, the stronger is the cliquishness of the graph. The clustering coefficient (cc) of a network is calculated by averaging the cc of all its nodes. Since the Steam network is very large, calculating the local clustering coefficient for all the nodes in the network is not feasible. We use the "Forest Fire" method introduced by Leskovec *et al.* [10] and shown to preform well for scaled-down clustering coefficient calculation [9] to sample the network. The scaled-down set contained a million users (11% of the full graph) using a forward burning probability $p_f = 0.6$ and a backward burning probability $p_b = 0$. The average clustering coefficient of the sampled set is 0.092. A breakdown of the average clustering coefficient per node degree is shown in Figure 3.

A naive calculation of the Steam graph's diameter will require running $|V|$ times a BFS algorithm (since the graph

is unweighted we can avoid the costlier Dijkstra shortest path algorithm) with a time complexity of $O(|V^2| + |V||E|)$, which is in the order of quadrillion operations. Instead, we estimate the diameter of the graph using our algorithm on an approximation heuristic presented by Magnien *et al.* [11] and Takes and Kosters [17], which runs in $O(|V| + |E|)$. We ran 50 iterations of BFS search, each starting from the furthest node found in the previous iteration. In case there was more than one such node, we chose the one with the smallest degree. The combination of the two (furthest and lowest degree) suggests we run the BFS iterations from the most remote nodes in the graph, implying a tight bound on the diameter calculation. The result approximated diameter of the graph is 19. In small world networks that have power-law degree distributions (similar to the Steam networks without the maximum degree limit) the graph diameter behaves like $O(\ln|V|/\ln\ln|V|)$, so assuming the constant is the big-O notation is one the expected diameter of the Steam network should be 8. Further investigating our results, we found out that a very small number of nodes are responsible for the rather large diameter. Removing only 22 nodes from the graph (we are investigating the largest connected component which consists of over 8.2 million nodes, see next paragraph) will result in a diameter of 12.

The Steam community is not fully connected, it is comprised of a large connected component (LCC) which is comprised of the majority of the nodes, and additional smaller connected component. In the Steam network the LCC consists of 8244178 nodes, which holds 91.2% of the network nodes. Looking at the nodes that are not part of the LCC, we found that they are all part of very small connected components, none is larger than 14. It was validated that the lack of large clusters is not an artifact of the crawling algorithm.

additional statistical analysis of the network's structure is beyond the scope of this paper.

IV. Network Evolution

An intriguing aspect of Steam community is its evolution over time. Some aspects of the network, such as the number of nodes, can be observed since the introduction of the network. The friendship ties, however, are available only starting late 2008, thus can be tracked only over three years of network activity. We thus sample the connectivity graphs is periods of six months, starting six months after the introduction the this feature.

The number of users in Steam grows in a consistent manner since the introduction of the platform, as shown in Figure 4. The connectivity of the network, meaning the friendship connections introduced in 2008, rises exponentially and as expected quickly overtakes the number of users. We note that in 2011 there was a sharp growth in the number of users and connections, which is consistent with reports by Steam [19].The average node degree grows from 2.05 in 2009 to 18.18 at the end of 2011, in a rather linear manner - increasing the degree by an average of three every six months.

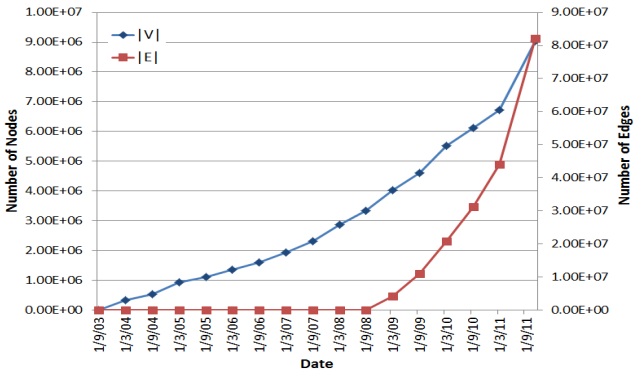

Fig. 4: Steam Community Size and Connectivity Evolution

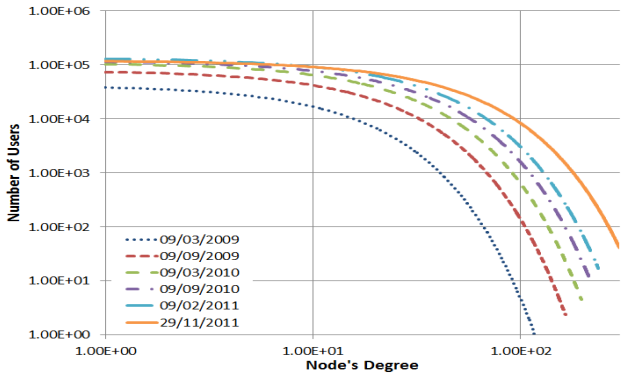

Fig. 5: Steam Community Degree Distribution Evolution

The degree distribution of the Steam community is studied during the last three years since friendship ties were introduced, as shown in Figure 5. As friends are added but not removed, it is expected that the network's degree will rise, however the rise indicates that new users joining in 2011 had more friends soon after they joined than new users in previous years. The maximal degree of a user rose from 176 in 2009 to 300 in 2011. Further more, while in 2009 only a single user had the maximal amount of friends, in 2011 there are over a dozen such users. Considering neighbor nodes degree distribution versus node degree distribution, a similar behavior is observed, as the average neighbor degree for a given degree gets higher as years go by. While in 2009, the average neighbor degree of a user with a single friend was 7.4, in 2011 it is already 34.4. For users with a hundred friendship ties, the average neighbor degree had risen from 27.6 to 79.1, in accordance.

The average clustering coefficient change over time was calculated as before using sampled nodes. For every 999 users, one random user was sampled, accommodating for the differences in graph size. While in 2011 the clustering coefficient was 0.2, back in 2009 it was only 0.084, growing to 0.15 in 2010. Though the clustering coefficient was lower in the past, it was still significantly higher than in a random graph, which shows that even when it emerged the Steam community presented a small world network.

Figure 6 presents the evolution of the LCC (marked by

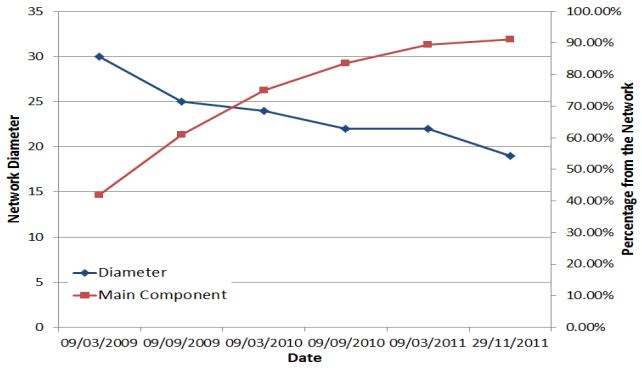

Fig. 6: Diameter and Main Component Evolution

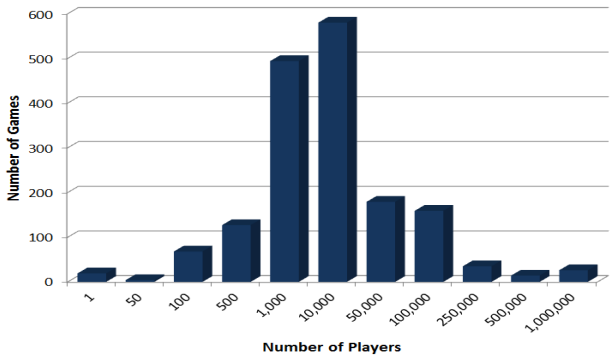

Fig. 7: Games Size Distribution

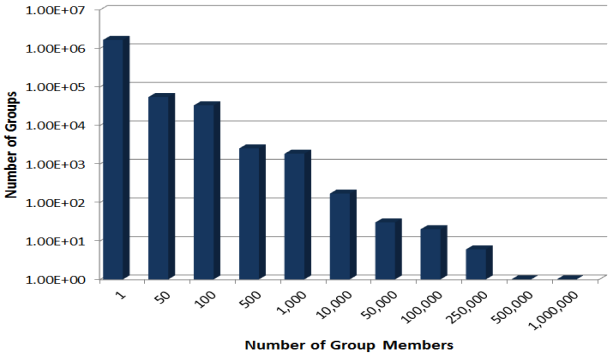

Fig. 8: Groups Size Distribution

rectangles) over time and consequently the diameter of the network's graph (marked by diamonds). The LCC is presented as a percentage of the overall graph, and the diameter is calculated over this component. In 2009, the LCC covered only 41.9% of the graph, with many small connected components. The nodes were connected within the graph in a loose manner (compared to the following years), thus the diameter of the graph was 30. As the main component expanded, the connectivity within it has tightened as well, leading to a diameter of 24 in 2010, when the LCC covered 75% of the network, and to a diameter of 19 in 2011, where 91.18% of the network were covered by the LCC. While the size of the LCC is already very large and cannot grow by more than a few percentages, the diameter of the graph is expected to keep decreasing as the connectivity within the exiting components continues to grow over time.

V. Games and Groups

The Steam platform had 1824 active games. We refer to a game as active if is available through the Steam store. Not all the games have players: we find that the users played only 1710 games. Some of the games played by users are no longer active, such as beta versions to games that were later released.

Figure 7 shows the number of games versus the number of games' players. Most of the games have between a thousand to a few tens of thousands of players. Twenty six games have been played by more than a million users, with nine of the games being played by over two million user, posing a quarter or more of the Steam community. It is interesting to note that two of these games are no longer active and have been beta versions of other active games. For one of these games, Team Fortress 2, the number of players of the released game is 1.1 million more than the beta. For the second game, Counter Strike: Source, there is a similar number of players for both beta and released version.

The number of groups in the Steam community is 1.98 million. Most of the groups are small: less than fifty members, as shown in Figure 8. Only a small number of groups, 27, have a hundred thousand members or more, with the largest group hosting 900K members. The large groups are either dedicated to popular games or link to game servers.

While users may participate in many groups and games, only a fraction of the users leverage this: most of the users take part in up to five groups or play up to five games, as shown in Figure 9. Less than 0.5% of the users participate in a hundred groups or more and less than 0.25% play a hundred or more games. The highest count of groups per unique user is 716 and the highest number of games attributed to one user is 3540 (followed by 2351).

The relation between a user's friendship ties and activity in the Steam community is of an interest. The average number of games per user is 6.5 and Figure 10 shows the number of games played by a user compared to his node's degree, as a scatter plot. The light line indicates the average number of

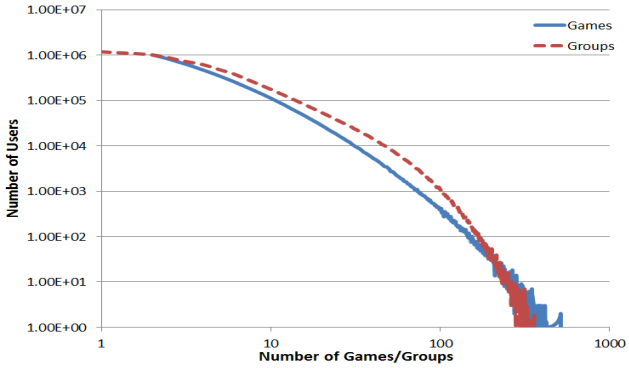

Fig. 9: Users Participation in Games and Groups

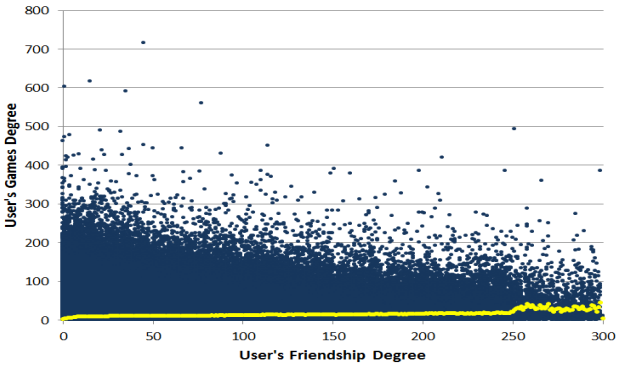

Fig. 10: Users' Friendship Degree vs. Played Games

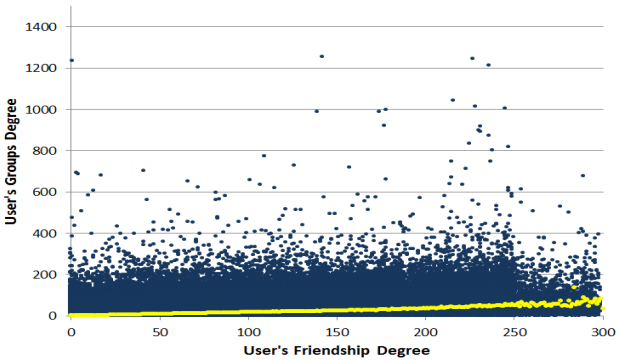

Fig. 11: Users' Friendship Degree vs. Groups Participation

games for each node degree. The average number of games per user grows with its degree: from an average of three games for a user with a single friend, to almost 30 games for users with a node degree of 250 or above. The range of number of games for each node degree is large, and the maximal number of games is played by a user with a rather low friendship degree of 46. The activity of users in groups and its correlation to the number of friends is shown in Figure 11 [1]. The light line indicated the average number of groups for each node degree. For users with a single friend, the average number of groups is only one, compare to sixty for users near the end of the degree scale. Surprisingly, the user participating in most groups, has only nine friends. We can therefore conclude that the activity of users in groups and games does not reflect their strength of position in the community, i.e. their ability to reach many users with a minimal number of hops.

VI. Conclusion

In this paper we presented an analysis of the Steam community network. The network shows small world graph characteristics, having a high clustering coefficient and a small diameter. The evolution of the network was surveyed since its introduction, focusing on the fast growth of connectivity within the network, Last, we observed characteristics of users activity in games and groups and referred to the relation between node's degree and participation in games and groups. In the future we plan to study commercial aspects of the network structure and its evolution in time.

[1]For clarity, we omit from the figure the few users participating in more than 1500 groups

Acknowledgments

We would like to thank Amir Shavitt for introducing us to the STEAM network and its intrinsics.

References

[1] Alexa. www.alexa.com [Accessed: November 17,2009].
[2] G. Armitage. Client-side adaptive search optimisation for online game server discovery. In *NETWORKING 2008 Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*, volume 4982 of *Lecture Notes in Computer Science*, pages 494–505. 2008.
[3] M. Balint, V. Posea, A. Dimitriu, and A. Iosup. An analysis of social gaming networks in online and face to face bridge communities. In *The third international workshop on Large-scale system and application performance*, LSAP '11, pages 35–42, 2011.
[4] J. Blackburn, R. Simha, N. Kourtellis, X. Zuo, M. Ripeanu, J. Skvoretz, and A. Iamnitchi. Branded with a scarlet "c": cheaters in a gaming social network. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 81–90, 2012.
[5] C. Chambers, W.-c. Feng, S. Sahu, and D. Saha. Measurement-based characterization of a collection of on-line games. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, IMC '05, pages 1–1, 2005.
[6] C. Chambers, W.-C. Feng, S. Sahu, D. Saha, and D. Brandt. Characterizing online games. *IEEE/ACM Trans. Netw.*, 18(3):899–910, June 2010.
[7] B. Keegan, M. Ahmed, D. Williams, J. Srivastava, and N. Contractor. Dark gold: Statistical properties of clandestine networks in massively multiplayer online games. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 201 –208, aug. 2010.
[8] C.-A. Larsen. Financial aspects of the online gaming service industry, 2011.
[9] J. Leskovec and C. Faloutsos. Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '06, pages 631–636, 2006.
[10] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, KDD '05, pages 177–187, 2005.
[11] C. Magnien, M. Latapy, and M. Habib. Fast computation of empirically tight bounds for the diameter of massive graphs. *J. Exp. Algorithmics*, 13, Feb. 2009.
[12] C. Moore. Hats of affect: A study of affect, achievements and hats in team fortress 2. In *Game Studies, The international journal of computer game research*, volume 11. 2011.
[13] A. Nazir, S. Raza, and C.-N. Chuah. Unveiling facebook: a measurement study of social network based applications. In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, IMC '08, pages 43–56, 2008.
[14] R. Pastor-Satorras and A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge Univ Press, 2007.
[15] S. E. Siwek. Video games in the 21st century, the 2010 report. Technical report, 2010.
[16] STEAM. http://store.steampowered.com/.
[17] F. W. Takes and W. A. Kosters. Determining the diameter of small world networks. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 1191–1196, 2011.
[18] The PC Gaming Alliance. Horizons research 2011. Technical report, 2011.
[19] Valve. Steam closes record 2011, Jan. 2012.
[20] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684), June 1998.