Network	# of nodes	# of edges	$\langle D_{uv} \rangle$ (km)	$\langle l_{uv} \rangle$ (km)	$\langle NL \rangle$
Steam Community	4,342,670	26,475,896	5,896	1,853	0.79
Steam Community Cheater-to-Cheater	190,041	353,331	4,607	1,761	0.79
BrightKite	54,190	213,668	5,683	2,041	0.82
FourSquare	58,424	351,216	4,312	1,296	0.85

Table 2: Location network properties: the number of nodes, edges, mean distance between users $\langle D_{uv} \rangle$, average link length $\langle l_{uv} \rangle$, average node locality $\langle NL \rangle$. The FourSquare and BrightKite properties are from [24].

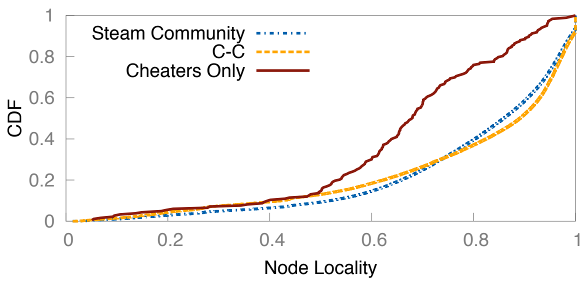


Figure 9: CDF of node locality.

graph (C-C), as well as just the cheaters within the location network (Cheaters Only). We first note that about 40% of users in the location network have a node locality of above 0.9, a phenomena exhibited by other geographic online social networks such as BrightKite and FourSquare [24]. This is strong evidence that Steam Community relationships exhibit geo-social properties, a characteristic to be expected in the context of multiplayer gaming where high network latencies cannot be well masked by current game infrastructure. Next, we observe that the cheater-to-cheater network and the Steam Community at large have similar node locality distributions. Finally, when considering only the cheaters embedded within the location network, we see drastically lower node locality, with only about 10% of cheaters having a node locality greater than 0.9.

These results lead to three observations: 1) friendships tend to form between geographically close users, 2) cheaters tend to form relationships with other nearby cheaters and these links are geographically closer than those formed by non-cheaters, and 3) as evidenced by their lower node locality when considering the entire location network and not only the cheater-to-cheater subgraph, cheaters appear to befriend geographically remote fair players. This might indicate that cheaters form relationships with other cheaters via a different mechanism than they form relationships with non-cheaters. Cheater-to-cheater relationships appear geographically constrained, while their relationships with non-cheaters are over larger distances.

Social proximity: We use *social proximity* as the second metric to characterize the strength of the relationships between Steam Community users and understand whether they materially differ for the cheater population. The social proximity metric is based on a previous study [22] that suggests that the overlap between the social neighborhood of two individuals is a good indicator of the strength of their relationship. We study the overlap of friends of users in the Steam Community networks to understand whether cheaters exhibit a stronger relationship with other cheaters than fair players do with fair players. We assess the strength of the relationship between two connected users by the overlap between their sets of friends, computed as follows:

$$Overlap_{uv} = m_{uv}/((k_u - 1) + (k_v - 1) - m_{uv})$$

where m_{uv} is the number of common neighbors between users u and v, k_u is the number of neighbors of user u and k_v is the number of neighbors of user v. This overlap is calculated for two groups of user pairs: the 1.5 million pairs of cheaters (i.e., all cheater pairs in the full social network) and 1.5 million randomly selected pairs of non-cheaters (i.e., about 2% of the existing non-cheater pairs). Additionally, we also calculate the same metric on the cheater-only as well as on the non-cheater-only graphs.

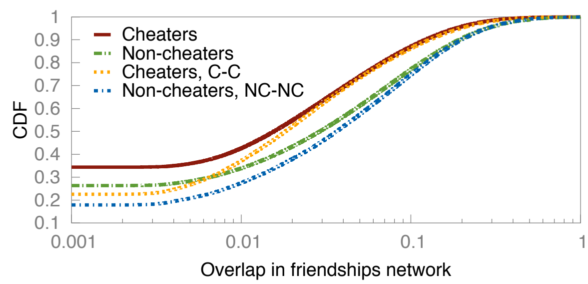


Figure 10: CDF of social proximity for cheater and non-cheater pairs when we consider all relationships, only cheater to cheater relationships (labeled C-C) and only non-cheater to non-cheater relationships (labeled NC-NC).

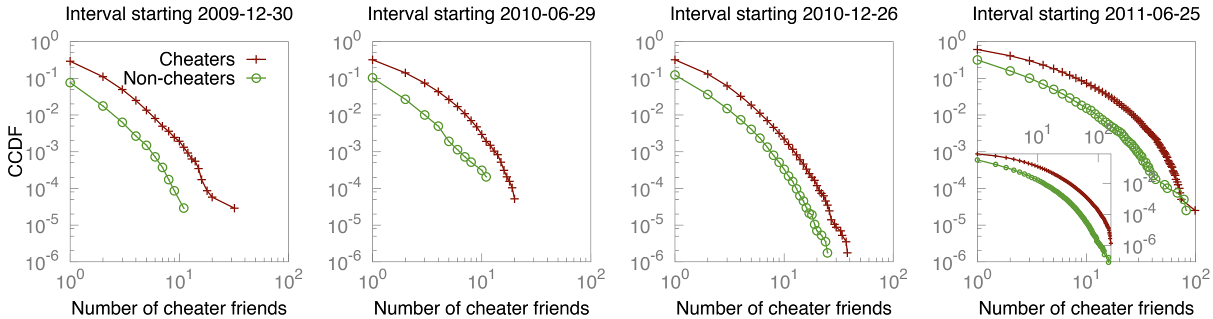
Figure 10 shows a higher overlap for cheater pairs in the cheater-only graph and non-cheater pairs in the non-cheater-only graph compared to the respective overlaps in the overall social network. This suggests that social relationships are weaker between different types of players (cheaters to non-cheaters) than within a uniform group.

5. PROPAGATION OF CHEATING

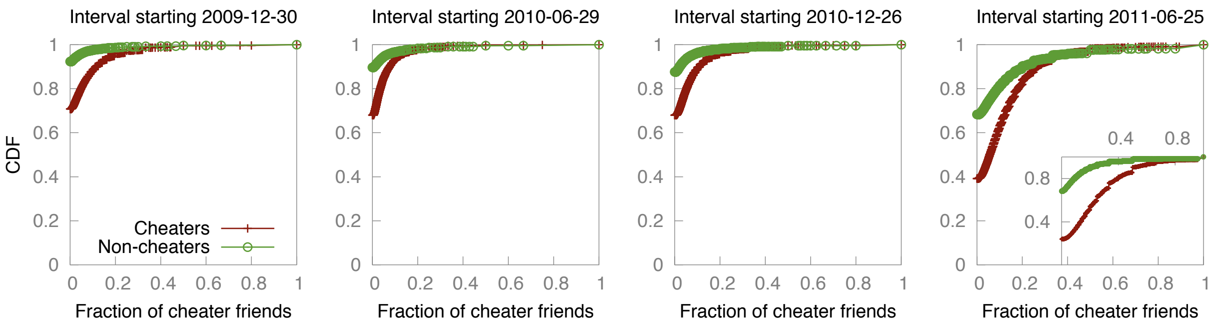
How does cheating behavior spread in the Steam Community? A first insight can be obtained by investigating how cheating bans propagate in the network over time (Section 5.1). Further insight can be obtained by understanding whether cheaters hold positions of influence in the social network (Section 5.2).

5.1 How Does Cheating Propagate over Time?

Based on the observations from Figure 3, we hypothesize that the friends of known cheaters are at risk of becoming cheaters themselves. To test this hypothesis, we explored whether cheaters discovered during a given time interval were more likely to be friends of previously discovered cheaters than to be friends with non-cheaters. Again, we stress that banned dates must be treated as "on or before"



(a) CCDF of the number of cheater friends of newly discovered cheaters and a random sample of non-cheaters.



(b) CDF of the fraction of cheater friends of newly discovered cheaters and a random sample of non-cheaters.

Figure 11: The spreading of cheating behavior in the Steam Community over four 180-day time intervals. The inset plot shows the final state of the network. A randomly selected control sample of non-cheaters is used for comparison in each time interval.

as opposed to exact timestamps. To mitigate the effects of this uncertainty, we chose to examine cheaters discovered over 180-day long intervals.

We begin by assuming that all users in the Steam Community friendship network are non-cheaters. We then initialize the network by marking the 94,522 users found to have a VAC ban on or before December 29, 2009 (i.e., the earliest date retrieved from vacbanned.com). For the first 180-day interval (between December 30, 2009 and June 28, 2010), 34,681 players were found to have a VAC ban. For these users, we calculated and plotted the number and fraction of their cheater friends (i.e., from the 94,522 cheaters found previously). We repeat these steps for another 3 time intervals, with 19,294, 571,975, and 43,465 cheaters found in each. The third interval (starting at December 26, 2010) contains the bulk of cheaters, since their VAC ban was first observed by our initial crawl (and not from `vacbanned.com`). However, as shown next, the differentiation between cheaters and non-cheaters holds true for all intervals. In addition to a best effort approximation of the timestamp of the VAC bans, the data constructed this way has another caveat: the social network is from our March/April 2011 crawl, but we show in Section 4.3 that the network is quite dynamic. We verified that despite the change in number of friends over time, the trend is preserved: we recalculated the fraction of cheaters in the 1 hop neighborhoods of users based on the state of their relationships as determined by our October 2011 re-crawl and we found the non-cheaters CDF to dominate the cheaters CDF as in Figure 3(a).

Figure 11 plots the results of our experiments. Each subplot represents only the cheaters discovered during the corresponding time interval, and an equal number of randomly sampled non-cheaters (statistical significance confirmed by KS test at 5% significance level, $p < 0.01$ in all cases, and D = 0.215, 0.2172, 0.1963, and 0.2911 for the fraction of cheater friends distributions of each interval, respectively). From the plots we see evidence that users with both a higher absolute number of cheater friends, as well as those with proportionally more cheater friends are more likely to become cheaters themselves.

To understand if the number of cheater friends has any predictive power on the state of a player, we used the framework developed by Backstrom et al. [5]. We first created a dataset of fair players with at least one cheater friend at the time of our first crawl. We then labeled the players from this set who were marked as cheaters by the time of our second crawl. We consider the new cheaters as joining a group in the sense of the Backstrom study to estimate the probability of a fair player becoming a cheater. We then applied the decision tree technique, using the number of cheater friends as the single feature, and achieved a ROCA of .61. With perfect knowledge, that is, if every single cheater was marked in our dataset by a perfect and timely cheating detection system, we would expect this value to be higher. (In comparison, Backstrom et al. show a ROCA value of .6456 for the DBLP communities and .69244 for LiveJournal.)

While these results show that the number of cheater friends has predictive power for the transition of a player from fair to cheater, it is difficult to ascertain the specifics of how the behavior propagates. Specifically, it is difficult to distinguish between homophily and contagion by simply observing the network properties and without strong assumptions about the social process [4, 26]. However, an individual's propensity towards unethical behavior, and cheating in particular, has been shown to be dependent on social norms, including the saliency of the behavior, and whether or not the behavior is observed from in- or out-group members [17]. Our large-scale findings further corroborate this theory and suggest that cheating behavior is not preponderantly dependent on a personal strategy that takes into account player-local information only, but spreads through a contagion process.

5.2 Are Cheaters in Positions of Influence?

The position of a node in the social network affects the potential influence the node has on close or even remote parts of the network. For example, a high degree centrality node—one with many direct neighbors—can directly influence more nodes than a low degree centrality node. High betweenness centrality nodes, on the other hand, mediate the traffic along the shortest paths between many pairs of nodes, thus having influence on remote parts of the network. A high betweenness centrality cheater, for example, could facilitate the propagation of cheats and other deviant behavior to distant parts of the gamers network.

To understand the relative importance of cheaters in the network, we study their potential for influence in the Steam Community by computing their degree centrality and node betweenness centrality. The degree centrality is simply the degree of the node in the network, and is thus a local metric. Betweenness centrality, however, is a global graph measure and consequently computationally expensive, requiring the calculation of the shortest paths between all pairs of nodes in the network. Due to the scale of our graph, we approximate betweenness centrality using *κ–path centrality*, a betweenness approximation method proposed in [2].

We observe a high correlation of 0.9731 between degree and betweenness centrality scores of the gamers. This high correlation remains consistent when we differentiate on the player type: 0.9817 for cheaters, and 0.9726 for non-cheaters. Consequently, if a player has many friends in the Steam Community network, (that is, high degree centrality), not only can she influence many players directly, but she can also mediate the information flow between remote players due to her likely high betweenness centrality.

We focus only on the most central players in the network and study how many of them are cheaters. Table 3 demonstrates that cheaters are under-represented among the most central players, despite the fact that they have about the same degree distribution as the fair players, as shown earlier in Figure 2(a). Over 7% of the entire player population in our dataset are cheaters, but they make up less than 7% of the top 1% most central players, and are not adequately represented until we consider the top 5% to top 10% most central players. Earlier results from Section 4.3 could provide an explanation for this. There seems to be social mechanisms that retard the growth of cheaters' social neighborhoods which could be preventing them from entering the top 1% central players in the social network.

Top-N%	0.1	0.5	1.0	5.0	10.0
DC	3.25	4.46	5.11	7.06	8.20
BC	5.16	5.95	6.35	7.86	8.58

Table 3: Percentage of cheaters found in top-N% of high degree centrality (DC) and betweenness centrality (BC) users in the Steam Community.

6. SUMMARY AND DISCUSSION

Online gaming has recently become the largest revenue-generating segment of the entertainment industry, with millions of geographically dispersed players engaging each other within the confines of virtual worlds. An ethical system is created along with the rules that govern the games. Just like in the real world, some players make the decision to circumvent the established rules to gain an unfair advantage, a practice actively discouraged by the industry and frowned upon by gamers themselves. This paper examined characteristics of these unethical actors in a large online gaming social network.

Due to the scale of our dataset, the majority of our computations used the MapReduce framework via the `python mrjob` interface for Hadoop on Amazon Elastic MapReduce. Our MapReduce stages involved graph pre-processing, game-play statistics computations, geographical data processing, computing degree distributions, intersections of sets, and geo-social metrics. Each solution included several MapReduce pipelines (chains of map tasks and reduce tasks) of smaller subtasks.

At a high level, viewed from the perspective of global network metrics, cheaters are well embedded in the social network, largely indistinguishable from fair players. This is not entirely unexpected. Cheaters are still gamers, and even though they are permanently marked, they remain members of the community. We observed evidence of this by examining both the social network and interaction logs from a multiplayer gaming server, where cheaters were not targeted or treated overly different from non-cheaters.

However, when we examine the transition from fair player to cheater, we observe the effects of the cheating brand. First, cheating behavior appears to spread through a social mechanism, where the presence and the number of cheater friends of a fair player is correlated with the likelihood of her becoming a cheater in the future. Consequently, cheaters end up having more cheater friends than the non-cheaters have. Second, we observed that cheaters are likely to switch to more restrictive privacy settings once they are caught, a sign that they might be uncomfortable with the VAC ban. Finally, we found that cheaters lose friends over time compared to non-cheaters, an indication that there is a social penalty involved with cheating.

Cheater distribution does not follow geographical, real-world population density. The fact that some regions have higher percentages of cheaters to the player population suggests that cheating behavior may be related to differences between specific geo-social cultures. Such cheating-prone communities might be the target of more scrutiny, or the result of higher tolerance to cheating behavior, both in the legislature and in the gaming population.

Our study has consequences for gaming in particular, but also for other online social networks with unethical members. In the case of gaming, individual servers can evaluate the cheating risk of a new player by looking at a combination of attributes inferred from the player's profile that include structural features. In the case of general online social networks, the findings of our study can be used to better understand the effects of countermeasures to deal with anti-social behavior. For example, the profiles of users who abuse the available communication tools for political activism or personal marketing, or who appear to automate their actions could be publicly tagged. Our study gives a preliminary indication that, over time, the reaction of fair users to such information will make it harder to benefit from forms of anti-social behaviors that attempt to harness network effects. The fair users tend to have a vested interest in maintaining the quality of the shared social space and will limit the connectivity of the abusing profiles.

Acknowledgements

We thank Y. Boshmaf, F. Figueiredo, A. Sotirkopoulos, S. Al-Kiswany, and C. Shawn of University of British Columbia for their feedback on an earlier version of this work. We also thank the anonymous reviewers for their input. This research was supported by NSF under Grants No. CNS 0952420 and CNS 0831785.

7. REFERENCES

[1] M. A. Ahmad, B. Keegan, S. Sullivan, D. Williams, J. Srivastava, and N. Contractor. Illicit Bits: Detecting and Analyzing Contraband Networks in Massively Multiplayer Online Games. *SocialCom'11*, 2011.

[2] T. Alahakoon, R. Tripathi, N. Kourtellis, R. Simha, and A. Iamnitchi. K-path centrality: A new centrality measure in social networks. In *SNS 11*, 2011.

[3] APB Reloaded Dev Blog. `http://goo.gl/m9klr`, 2011.

[4] S. Aral, L. Muchnik, and Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Science*, 2009.

[5] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. 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*, 2006.

[6] M. Balint, V. Posea, A. Dimitriu, and A. Iosup. An Analysis of Social Gaming Networks in Online and Face to Face Bridge Communities. In *LSAP 11*, 2011.

[7] C. Chambers, W. Feng, S. Sahu, D. Saha, and D. Brandt. Characterizing online games. *IEEE/ACM TON*, 2010.

[8] M. Claypool, D. LaPoint, and J. Winslow. Network analysis of counter-strike and starcraft. In *IPCCC 03*, 2003.

[9] M. Consalvo. *Cheating: Gaining Advantage in Videogames*. MIT Press, Cambridge, Massachusetts, 1st, edition, 2007.

[10] N. Ducheneaut, N. Yee, E. Nickell, and R. Moore. The life and death of online gaming communities: a look at guilds in world of warcraft. In *CHI 07*, pages 839–848, 2007.

[11] H. B.-L. Duh and V. H. H. Chen. Cheating Behaviors in Online Gaming. *Online Communities and Social Computing*, 2009.

[12] D. D. Dumitrica. An exploration of cheating in a virtual gaming world. *Journal of Gaming & Virtual Worlds*, 2011.

[13] W. Feng, F. Chang, W. Feng, and J. Walpole. Provisioning on-line games: a traffic analysis of a busy counter-strike server. In *IMW 02*, pages 151–156, 2002.

[14] W. Feng, F. Chang, W. Feng, and J. Walpole. A traffic characterization of popular on-line games. *IEEE/ACM TON*, 13:488–500, 2005.

[15] W. Feng and W. Feng. On the Geographic Distribution of On-line Game Servers and Players. In *NetGames 03*, 2003.

[16] T. Fritsch, B. Voigt, and J. Schiller. Distribution of online hardcore player behavior: (how hardcore are you?). In *NetGames 06*, 2006.

[17] F. Gino, S. Ayal, and D. Ariely. Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science*, 2009.

[18] J. Huizinga. *Home Ludens: A Study of the Play-Element in Culture*. Beacon Press, Boston, Massachusetts, reprint edition, 1950.

[19] B. Keegan, M. Ahmed, D. Williams, J. Srivastava, and N. Contractor. Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games. In *SocialCom 10*, pages 201–208, 2010.

[20] Y. Ku, Y. Chen, K. Wu, and C. Chiu. An empirical analysis of online gaming crime characteristics from 2002 to 2004. In *PAISI 07*, pages 34–45, 2007.

[21] A. Nazir, S. Raza, C. Chuah, and B. Schipper. Ghostbusting facebook: detecting and characterizing phantom profiles in online social gaming applications. In *WOSN 10*, 2010.

[22] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[23] D. Parfit. *Reasons and Persons*. Oxford University Press, 1984.

[24] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance matters: geo-social metrics for online social networks. In *WOSN 10*, 2010.

[25] T. Senior. Steam controls up to 70% of the pc download market and is "tremendously profitable". `http://goo.gl/MjgA3`, 2011.

[26] C. R. Shalizi and A. C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, 2011.

[27] M. Szell and S. Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32:313–329, 2010.

[28] X. Wei, J. Yang, L. Adamic, R. Araújo, and M. Rekhi. Diffusion dynamics of games on online social networks. In *WOSN 10*, 2010.

[29] Y. Xu, X. Cao, A. Sellen, R. Herbrich, and T. Graepel. Sociable killers: understanding social relationships in an online first-person shooter game. In *CSCW 11*, pages 197–206, 2011.