2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Guilt by Association? Network Based Propagation Approaches for Gold Farmer Detection

Muhammad Aurangzeb Ahmad
Department of Computer Science
University of Minnesota
Email: mahmad@cs.umn.edu

Brian Keegan
Department of Political Science
Northeastern University
Email: bkeegan@northeastern.edu

Atanu Roy
Department of Computer Science
University of Minnesota
Email: atanu@cs.umn.edu

Dmitri Williams
Annenberg School of Communication
University of Southern California
Email: dmitri.williams@usc.edu

Jaideep Srivastava
Department of Computer Science
University of Minnesota
Email: srivastav@cs.umn.edu

Noshir Contractor
School of Communication
Northwestern University
Email: nosh@northwestern.edu

***Abstract*—The term 'Gold Farmer' refers to a class of players in massive online games (MOGs) involved in a set of inter-related activities which are considered to be deviant activities. Consequently these gold farmers are actively banned by game administrators.The task of gold farmer detection is to identify gold farmers in a population of players but just like other clandestine actors they not labeled as such. In this paper the problem of extending the label of gold farmers to players which are not labeled as such is considered. Two main classes of techniques are described and evaluated: Network-based approaches and similarity based approaches. It is also explored how dividing the problem further by relabeling the data based on behavioral patterns can further improve the results**

I. Introduction

Gold farming and real money trading refer to practices that involve the sale of virtual in-game resources for real-world money via exchanges outside of the game itself. The name stems from a variety of repetitive routines (farming) which are employed to accumulate virtual wealth (gold) which is sold to other players who lack the time or desire to accumulate their own in-game capital [10], [9]. Gold buyers are regular players who purchase this virtual capital to obtain more power weapons, armor, and abilities for their characters, accelerating them to higher levels and allowing them to explore larger parts of the game world and confront more interesting and challenging enemies [6]. The phenomenon of Gold farming in these games is analogous to criminal activity in the real world [11]. Researchers have argued that to the extent that individuals in online virtual worlds engage in similar psychological, social, and economic behavior as they do in the real world, virtual world research can potentially be mapped backwards and employed to understand real world phenomenon [14].

Massively-multiplayer online games such as World of Warcraft, EverQuest II, and Lord of the Rings Online are examples of virtual worlds in which players accumulate experience, armor, spells, and weapons to improve their power during encounters with non-player characters (NPCs) and player-versus-player combat (PvP). Virtual goods like in-game currency, scarce commodities, and powerful weapons require substantial investments of time to accumulate, but these can also be obtained from other players within the game through trade and exchange. Just as these in-game economies exhibit similar macroeconomic characteristics observed in real-world economies [6] virtual worlds also contain illicit markets for acquiring goods and skills. Such opportunities form the basis of the Gold Farming phenomenon.

Gold farming has been constructed as a deviant activity by both the game developers as well as the player communities for a variety of reasons. First, in-game economies are designed with carefully-calibrated activities and products that serve as sinks to remove money from circulation. Because gold farmers and buyers inject currency into the economy, they create inflationary pressure, unintended arbitrage opportunities, and other perverse incentives for market agents. Second, farmers activities often overtly affect other players experiences by excluding them from shared game environments, employing anti-social computer scripts (bots) to automate the farming process, and engaging in the outright theft of account and financial information from their customers [10]. Third, the game developers are risk-averse to the legal implications (such as property rights, taxation, and torts) of sanctioning a multinational industry estimated to generate between 100 million and 1 billion dollars in revenue annually [6], [12] while lacking legal jurisdiction, precedent, or regulation [11]. Finally, farming upsets the meritocratic and fantasy-based nature of the game in which some players can buy rather than earn accomplishments, thus potentially driving legitimate players away [8]. For all these reasons, game developers actively and visibly ban accounts engaged in gold farming [1].

While it is possible to identify a set of features for detecting Gold Farmers based on domain knowledge or other criteria like information gain [1]), the task of detecting Gold Farmers is not straightforward because not all the players who are Gold Farmers are identified as Gold Farmers and many seemingly misclassified instances of Gold Farming may actually be Gold Farmers. Thus it is necessary to augment the task of classification with techniques like network analysis and label propagation. In this paper we study the problem of automatically detecting Gold Farmers by augmenting the classification task with label propagation. The main idea is to relabel a subset of the instances of the majority class which may have been incorrectly labeled in the first place or instances

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

which may have evaded detection. While label propagation is a widely studied problem, given the unique nature of the Gold Farming dataset we address the following two sub-problems in this domain. (i) Label propagation when the membership of one minority class is known for certain while the membership of the majority class is unknown (ii) Label propagation when the classes have multiple labels but only one of the labels is the correct one. We use data from an MMORPG called EverQuest II (EQ2) for the Gold Farmer prediction and label propagation task.

II. Related Work

There have a number of studies on virtual property [6], economic impact [11] and legal issues [4] related to real money transactions in the virtual worlds. Research has also inferred indirectly the scale of the activity based upon proxy measures of price level stabilization and price similarity across agents [12]. The work which is most closely related to this paper is the work on automatic detection of Gold farmers as machine learning problem by Ahmad et al [1] and the problem of false negatives in Gold Farmer detection by Atanu et al [3]. We note that the current work goes beyond the previous work by taking into account the networked nature of Gold Farming and addressing the problem of potentially incorrect labels in the data which may negatively affect the classification results and disambiguating labels in the dataset. Bot detection techniques are not really relevant tor work since bots are not as prevalent in EQ2 as reported in previous work on this subject [1].

Machine learning and data mining have been extensively used in the context of detecting and combating cybercrime [1]. Other studies in criminal analysis have employed social network analysis, entity detection, and anomaly detection techniques have been used extensively in this context [17]. Label Propagation is a widely studied problem in semi-supervised learning [17]. Blum et al address the problem of semi-supervised learning as a graph mincut problem where the positive labels act as sources and negative labels act as sinks [5]. Zhu and Ghahramani [16] proposed a method based on Markov Random fields, Szummer and Jaakkola [13] proposed a method based on a t-step Markov random walk on the graph where the influence of one example to another example is proportional to how easy the random walk goes from one to the other. An extensive survey of related literature on the related problem of semi-supervised learning in Graphs is given by Zhu et al [17]. While there are similarities between the various problems described here, it should be noted that none of them address the problem that we are addressing in this paper as the setting of the problem is fundamentally different where a subset of the labels are known and another subset of the majority class are uncertain to some extent and there are additional sub-classes in the minority class.

III. Problems with Gold Farmer Detection

Gold Farming activities may include but are not limited to Killing NPCs, creating/acquiring high-value items, market arbitrage, etc. Players sell farmed gold to in-game buyers who use the currency to purchase more powerful items and abilities to open new areas of the game. There have been some efforts with respect to systemizing the problem of catching Gold Farmers [1][3] to make it more of a science than an art.

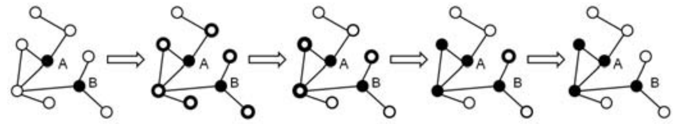

Fig. 1. Label Propagation

Traditionally most gold farmers are caught by: Other players reporting in-game suspicious behavior that may be associated with Gold Farmers, Farmers solicitations and spamming, organized sting operations to lure and catch Gold Farmers, administrator heuristics about the behavior of Gold Farmers etc [1]. Farming operations employ highly-specialized operations that have to balance practices to efficiently accumulate gold with practices to avoid detection [10].

Ahmad et al [1] identified four classes of gold farmers. While a Gold Farmer can take on multiple roles simultaneously, in some cases the overlap is likely to be zero e.g., if the Gold Farmer is playing the character with a high frequency of play then the character is most certainly will be a gatherer with almost no overlap between the other Gold Farming types. It should be emphasized that while techniques for automatically identifying Gold Farmers [1] like the ones described in this paper can be used to automatically label Gold Farmers, the labeled cases have to be investigated by a game administrator before banning accounts for Gold Farming because of the high cost associated with banning an account which is a legitimate account.

IV. Label Propagation

While the players who are labeled as Gold Farmers are most certainly Gold Farmers, the same cannot be stated with certainty for the players who are not labeled as Gold Farmers. Since Gold Farmer detection is most cases is heuristic it is always the case that a large number of Gold Farmers evade detection [1], [3]. Hence classification techniques can only go so far [1], [3]. An additional problem with classifying players as Gold Farmers is that there are players who are labeled as one of the Gold Farmers sub-classes but have a high degree of overlap between different types of behaviors associated with other classes.

A. Problem Description

In the current domain there are two main classes of labels: Golf Farmers and non-Gold Farmers. The Gold Farming class itself can be divided into multiple sub-classes as described in section III. We first describe the more general problem of label propagation and then refine it for our context. The problem of label propagation can be described as follows: Given a set of instances $X = \{x_1, x_2, ..., x_n\}$ and a set of labels $L = \{l_1, l_2, ..., l_m\}$ such that a subset of the instances $X_k = \{x_1, x_2, ..., x_k\}$ where $k < n$ are labeled and the rest of the instances are not labeled, the problem of label propagation is to predict the labels of the unlabeled instances.

B. Gold Farmer Label Propagation Problem

The problem of label propagation in the Gold Farming domain is slightly different from how it is posed in traditional

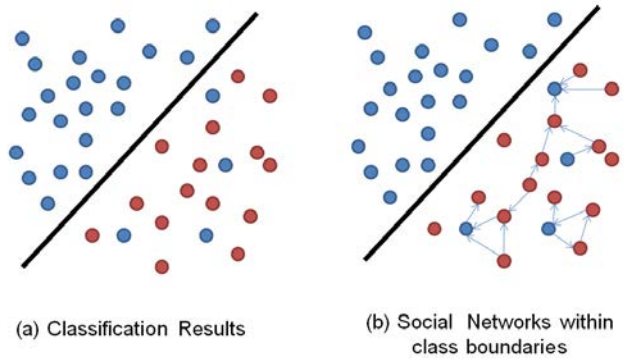

Fig. 2. Classification Boundaries and Social Networks of Gold Farmers

label propagation domains. The main difference is that given two classes of labels, the membership of minority class is known for certain for a subset of cases while the membership of the majority class is not known for certain. Formally this problem can be described as follows: Given a set of instances $X = \{x_1, x_2, ..., x_n\}$ and a set of labels $L = \{l_1, l_2\}$ such that a subset of the instances $X_k = \{x_1, x_2, .., x_k\}$ where $k << n$ are labeled with l_1 and $X_U = X - X_K$ is the set of unlabeled instances, the problem of Gold Farming class label propagation is the problem of predicting the labels in X_U. Based on the problem description the problem of label propagation for Gold Farmers can be be addressed by two different approaches: (i) In terms of similarity between individual instances (ii) In terms of label propagation in a social network.

1) Similarity Matrix-based Method: Similarity based approaches look at the space of features for the labeled cases and compute similarity between them and the unlabeled cases. Based on a given criteria or a cut-off instances are re-labeled as the class of interest. We use the following scheme, modified from Zhou et al's [15] label propagation scheme, for this class of methods.

1) Compute the affinity matrix W as follows:
   $W_{ij} = exp(-||x_i - x_j||^2/2\sigma^2),\ \text{if}\ i \neq j,$
   $W_{ij} = 0,\ \text{if}\ i = j,$
2) Construct the matrix S as follows:
   $S = D^{-1/2}WD^{1/2}$
   Where D is a diagonal matrix such that
   $D_{ii} = \Sigma_(j = 1)^n W_{ij}$
3) Compute $F(t)$ until convergence as follows:
   $F(t+1) = \alpha SF(t) + (1-\alpha)\gamma$, where $\alpha \in (0,1)$
4) If F^* is the limit of the sequence $\{F(t)\}$ then each x_i is labeled as follows:
   $y_i = argmax_{j \leq c} F^*_{ij}$

The space of features that is used is described in section V-B.

2) Network Based Method: The network based approach is based on the premise that for certain types of activitie,s players are likely to engage in those types of activities with other players with whom they have a strong relationship as compared to others. In general graph based techniques for label propagation assume that the nodes which are connected to one another with high edge weights are likely to have similar labels [17]. Given the high class imbalance in the dataset traditional graph based approaches are not really applicable to this context [17]. To address this issue instead we propose a model inspired from the spreading activation model [7]. The main idea behind the activation model is that one begins with a set of already labeled nodes called source nodes with weights or "activation." One then iteratively propagates that activation out to other nodes linked to the source nodes. The modified model can be described as follows:

Input:

1) A set of nodes $N = n_1, n_2, , n_m$ with a subset N_L of the nodes $|N_L| << |N|$ labeled with class C and a union of graphs $G = G_1 \cup G_2 \cup ... \cup G_P$ on the same nodeset N. Each edge e_{ij} on the graph has an associated weight w_{ij}.
2) An activation value $a_i \in (0,1)$ associated with each node n_i.
3) Firing Threshold $F \in (0,1)$ and the decay factor $D \in (0,1)$

Procedure:

1) Consider a set $R = \phi$ and initialize the graph G by setting the activation value as follows:
   $a_i > F$, if $L(n_i) = C$
   $a_i = 0$, otherwise
2) Do until $|A| > 0$, where $A = \{n_i | a_i > F\} - R$
   a) Fire the nodes in A i.e., for each node connected to a node $n_i \in A$, adjust the activation value as follows:
      $a_j = a_j + a_{ij}.w_{ij}.D$
   b) Update A as follows: Remove all the nodes which have fired and add them to set R.

Network based methods are useful for cases where traditional approaches misclassify Gold Farmers but there is sufficient secondary evidence that the Gold Farmers mostly socialize with other Gold Farmers. Figure 1 illustrates the main idea behind this method. We start with an initial set of labeled Gold Farmer nodes A and B and when the stopping criteria is reached the pool of Gold Farmers is increased.

C. Label Disambiguation Problem

While not Gold Farmers have their type specified in a subset of the Gold Farming labels have labels for some subclasses. However there is an overlap in the labels of many of these instances to a great extent as described in section II, even though the overlap between their corresponding behavioral signatures is almost non-existent. Consequently these labels have to be disambiguated since there is not much overlap between the behaviors of the various sub-types of Gold Farmers under consideration. Formally the label disambiguation problem can be described as follows: Given a set of instances $X = \{x_1, x_2, ..., x_n\}$ and a set of labels $L = \{l_1, l_2\}$ such that a subset of the instances $X_k = \{x_1, x_2, ..., x_k\}$ where $k < n$ are labeled with both l_1 and l_2, the problem of label disambiguation is the problem of assigning the instances of X_k with either the label l_1 or l_2. We employ a modified version of the similarity based approach described previously to address this problem. We start with the assumption that the instances which have multiple labels are analogous to the unlabeled cases and treat them as such examples. After propagating the labels based on similarity we force a labeling in these instances based on their similarities. For comparison we also employed the network based approach in this case as well.

V. Dataset

We used Gold Farming data from the MMO game EQ2. There are a total of 2,122,781 distinct characters and 675,296 distinct accounts in the game logs. It is possible for an account to have multiple characters associated with it where an account corresponds to a real person. The data spans from January 2006 to September 11, 2006. The dataset consisted of 6,651 Gold Farmers. This also implies that the percentage of Gold Farmers is less 0.31 and thus the Gold Farming class is a rare class.

A. Network Data

We note that the scheme described in section IV-B2 does not depend upon the choice of network used. Since multiple networks are present in the EQ2 dataset we employ them for the label propagation scheme from IV-B2. We use three different in-game networks for this purpose: Mentoring Network, Housing-Trust Network and the Trade Network. The semantics relating to the strength of relationship in each network is described in detail in [2]: A subset of Gold Farmers occupy very prominent in the trade network such that they have an extremely high out-degree as compared to most other nodes in the network. This is illustrated in Figure 3 which shows a snapshot of the Trade Network. The known Gold Farmers are shown in red in contrast to other nodes in the network which are shown in blue. The high degree of connectivity of the Gold Farming nodes is readily evident. The opposite is observed in the Housing-Trust and the Mentoring network [3] where the Gold Farmers are not prominent but rather try to keep to themselves away from other Gold Farmers.

B. Featureset

In order to do a fair comparison between the proposed techniques and what has been reported in literature, we use the same featureset that was used by Ahmad et al [1]. Due to limitations in space the reader is referred to [1] for details regarding individual features, however we note that the featuresets can be categorized into the following categories:

- **Player Character Features:** The attributes of a player's character e.g., gender, class, race, faction etc. It should be noted that the character's attribute are not the same as the player's real world attributes.
- **Aggregate Features:** The aggregate of various in-game features like sum of in-game metal, experience points, trade-skill points etc for all the characters associated with the account.
- **Demographic Features:** Demographic features about the player like gender, country, language etc.
- **Temporal Features:** Features constructed from in-game temporal behaviors like monster kills, instances of experience points gained etc.

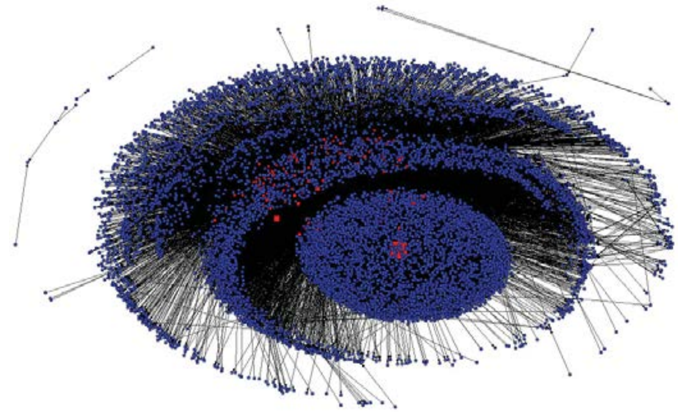

Fig. 3. A snapshot of the trade network

VI. Experiments

We address this problem as a binary classification problem where the two classes refer to the Gold Farmer and the non-Gold Farmer class. 10 fold cross validation is used and we use the same standard set of seven classifiers for classifying Gold Farmers that was used by Ahmad et al [1]: Naive Bayes, Bayes Net, KNN, Logistical Regression, AdaBoost, and JRIP. Due to limitations of space we only report the results from comparing the results of the best models and classifiers reported in previous work.

A subset of the non-gold farming instances are relabeled based on the two techniques described previously. In the case of similarity based method we relabel these instances if they are sufficiently similar to the instances in the Gold Farmer class. A different approach is used for the network propagation approach, instances are relabeled based on if the labels converge until the stopping criteria is met. In both the cases once relabeling has done we address the problem as a normal classification problem i.e., the relabeled instances are considered to be instances of the Gold Farmer class. Using the same set of features but with newly relabeled instances as well as the original instances of Gold Farmers we consider the classification task. The rest of the experimental setup including the classifiers considered remains the same.

The classifiers are reevaluated on the relabeled instances and the main purpose of reevaluation is to determine if relabeling makes a difference with respect to the class boundaries in the previously poorly labeled data. The main idea being that if a significant improvement can be demonstrated then the relabeling approaches can be used as a viable Gold Farmer detection method. This still of course does not completely address the problem of false positives; in section VI-C we analyze the classification rules obtained from JRIP using the relabeled data to illustrate the regularities that are observed between the originally labeled data and the relabeled data.

A. Classification Results

The results of our experiments are summarized in Table VI-A where the precision, recall and F-score for each classifier is given. Original results refer to the results that are reported in [1]. Sim Based refers to the results obtained from using the similarity based method and Network Prop refers to the method

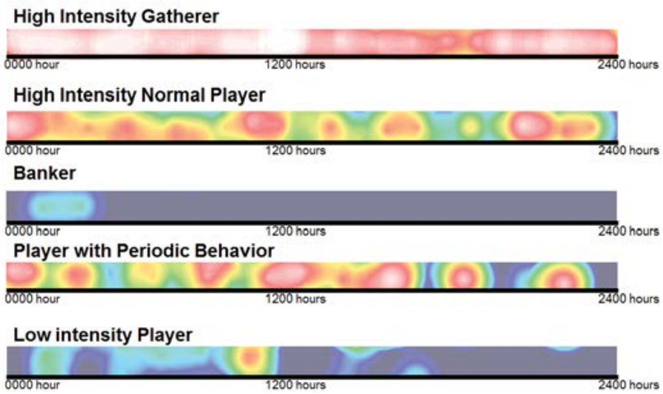

Fig. 4. Heat Signatures of Gold Farmers and Other Player Types

that is based on the network propagation based approach. The change in performance as well as the life is also given. The column for the original results corresponds to Model 4 of the results from [1] which yielded the best results. While the data is available for all three networks we report the best results in Table VI-A because of limitations of space we report the best results which in this case are obtained from using the mentoring network.

First consider the similarity based approaches, the best results are obtained for JRIP and KNN. In most cases substantial Lift from the original results ins obtained. Even in the case of the KNN the results improve from the previous case but the improvement for JRIP is quite substantial. The best classification results are obtained for the JRIP algorithm for the network propagation approach. It is interesting to note that JRIP does quite poorly when only the originally labeled dataset is used. We discuss the implications of some of these results in section VI-C. One can also compare the results from the similarity based approach as well as the network based approach where the network based approach gives better results in almost all of the cases.

Since the best results for the network based are obtained from the mentoring networks, we speculate that this is because mentoring relations are most often between Gold Farmers even though not all players which are mentored by Gold Farmers are labeled as Gold Farmers. This makes sense from the perspective of transaction costs associated with forming mentoring edges. In contrast to other types of edges i.e., social relationships mentoring does require substantial effort in terms of time for the parties involved to form that relationship [2]. Thus it would make sense that Gold Farmers would form mentoring relationships with other Gold Farmers.

B. Analysis of Sub-classes

The second task that we described was the disambiguation of sub-classes of Gold Farmers. We note that while most of the data uses the generic Gold Farmer labels to refer to the Gold Farmers, a subset of the labels do indicate the type of the Gold Farmer one is dealing with.These are given in Table VI-B. The main problem with this table is that it does not use the standard categorization used for Gold Farmers and the labels that are given are not exclusive in the case of 'Spammers and Bots.' A large percentage of Bots in this case refer to Gatherers

TABLE II. LABEL SUB-CLASSES OF GOLD FARMERS

GF Class	Instances
Spammer	3,296
Spammers and Bots	1,350
Spammer Only	1,946
e-bay Related	22

given the type of activities that they perform, however even this categorization is not exclusive. This makes the task of disambiguation quite complicated. To partially circumnavigate this problem we focus on two sub-classes of Gold-Farmers whose behaviors are more well known [1][2] and other metrics can be used for labeling them.

We consider the level of activities of players over the course of a day and visualize it as a heatmap of behaviors in Figure 4. The x-axis represents the time of the day over the 24 hour period. The y-axis represents days where we consider data over the course of three months. The standard coloring for heatmaps is used i.e., from blue to red, thus red represents high levels of activity and blue represents low levels of activity. The activities are averaged over the subset of Gold Farmers and randomly sampled players.

From the visualization it is clear that the behaviors of Gathers and Bankers are quite distinct from one another and from other classes that are considered. We caution that this behavioral signature is not exclusive to Gold Farmers but a subset of 'normal' players also exhibit such behaviors. However if one already has a set of labeled Gold Farmers one could use these signatures to identify different sub-classes. The high intensity of activity exhibited by certain Gold Farmers (Gatherers) is different from other high intensity players. For comparison we also show players which exhibit periodic behaviors as well as players with a low intensity level of play. The main lesson from this exercise is that at least for some sub-classes of Gold Farmers one can be separated based on their intensity of play.

C. Gold Farmers related Classification Rules

In this section we examine some of the classification rules which are obtained from the JRIP classifier. We emphasize that these rules are obtained from the original instances that were labeled as Gold Farmers as well as the instances that are relabeled as Gold Farmers from the network-propagation methods. These JRIP rules are given as follows:

- The player speaks Chinese, has less than 794 XP, has less than 19 trade transactions and is more than 27 years old.
- The player speaks Chinese, has less than 200 XP, has less than 19 trade transactions and is between 25 and 31 years old.
- The player has more than 1,799 XP and the age of the player is *exactly* 37.59 years. (245 instances)
- The player speaks Chinese, the age is between 20 and 23 year and has more than 15 trade related transactions.
- The player speaks Chinese, is between 27 and 33 years old and has between 40 and 52 economic transactions.

TABLE I. COMPARISON OF CLASSIFICATION PERFORMANCE ACROSS CLASSIFIERS AND TECHNIQUES

Classifier	Metric	Original Results	Sim. Based	Change In Performance	Lift	Network Prop	Change In Performance	Lift
Bayes Net	Precision	0.291	0.170	-0.121	0.584	0.189	-0.102	0.649
	Recall	0.513	0.834	0.321	1.626	0.819	0.306	1.596
	F-Score	0.371	0.282	-0.089	0.760	0.307	-0.064	0.827
J48	Precision	0	0.494	0.494	N/A	0.62	0.62	N/A
	Recall	0	0.189	0.189	N/A	0.337	0.337	N/A
	F-Score	0	0.273	0.273	N/A	0.437	0.437	N/A
J Rip	Precision	0.526	0.495	-0.031	0.941	0.537	0.011	1.021
	Recall	0.056	0.462	0.406	8.250	0.462	0.406	8.250
	F-Score	0.102	0.478	0.376	4.686	0.497	0.395	4.873
KNN	Precision	0.345	0.436	0.091	1.264	0.46	0.115	1.333
	Recall	0.361	0.396	0.035	1.097	0.428	0.067	1.186
	F-Score	0.353	0.415	0.062	1.176	0.443	0.09	1.255
Logistic Regression	Precision	0.333	0.455	0.122	1.366	0.534	0.201	1.604
	Recall	0.020	0.189	0.169	9.450	0.271	0.251	13.550
	F-Score	0.038	0.267	0.229	7.026	0.360	0.322	9.474
Nave Bayes	Precision	0.204	0.146	-0.058	0.716	0.142	-0.062	0.696
	Recall	0.223	0.538	0.315	2.413	0.502	0.279	2.251
	F-Score	0.213	0.230	0.017	1.080	0.221	0.008	1.038
Adaboost w/ DT	Precision	0.271	0.405	0.134	1.494	0.471	0.2	1.738
	Recall	0.183	0.105	-0.078	0.574	0.080	-0.103	0.437
	F-Score	0.218	0.167	-0.051	0.766	0.137	-0.081	0.628

Here XP refers to experience points. The main observation to note here is that most of the Gold Farmers are Chinese speakers which is in line with previous observations [10]. The third rule is quite interesting which gives an exact age for Gold Farmers and the rule holds for a large number of instances (245). The reason that this is observed is because the Gold Farmer accounts are usually created by the same person or the same group of people. In this instance the accounts that were created have the same date of birth. It is extremely unlikely for such a large number of accounts to not only have the same data of birth and also socialize primarily with Gold Farmers. This was not discovered by the game admin who initially caught the Gold Farmers. The rest of the rules deal with the experience points gained by the players as well as their economic activity in the game.

VII. CONCLUSION

In this paper we addressed the problem of automatically detecting Gold Farmers by employing network analysis, label propagation and machine learning in a unified framework. The problem of automatically detecting Gold Farmers is not a straightforward classification problem as there are a number of issues involved in the process i.e., not all instances of Gold Farmers are labeled as Gold Farmers so that many of the cases of such labeling have to be indirectly inferred. We employed two different techniques for labeling propagation which showed significant improvement over previously reported results. An additional issue is multiple labeling for exclusive classes which can only be partially addressed because of labeling issues. In future work we plan to expand the framework by taking into account the temporal aspects of the networks that Gold Farmers are involved in as well as the behavioral signatures of farmers described in this paper.

ACKNOWLEDGMENT

This work was sponsored by the Army Research Laboratory and DARPA under Cooperative Agreement Numbers W911NF-09-2-0053 and W911NF-12-C-0028. The data was provided by the Sony Online Entertainment. The findings presented do not in any way represent, either directly or through implication, the policies of any of the sponsors.

REFERENCES

[1] Muhammad Aurangzeb Ahmad, Brian Keegan, Jaideep Srivastava, Dmitri Williams, Noshir Contractor, *Mining for Gold Farmers: Automatic Detection of Deviant Players in MMOGs*, cse, vol. 4, pp.340-345, 2009 Intl.l Conference on Computational Science and Engineering, 2009

[2] Muhammad Aurangzeb Ahmad, Zoheb Borbora, Jaideep Srivastava, Noshir Contractor *Link Prediction Across Multiple Social Networks* Domain Driven Data Mining Workshop (DDDM2010), ICDM 2010 Sydney, Australia.

[3] Atanu Roy, Muhammad Aurangzeb Ahmad, Chandrima Sarkar, Brian Keegan and Jaideep Srivastava *The Ones That Got Away: False Negative Estimation Based Approaches for Gold Farmer Detection* IEEE SocialCom 2012 Amsterdam, Netherlands September 3-5, 2012

[4] J. Balkin, *Virtual Liberty: Freedom to Design and Freedom to Play in Virtual Worlds.*, Virginia Law Review, vol. 90, no. 8, 2004.

[5] A. Blum, S. Chawla, *Learning from labeled and unlabeled data using graph mincuts*. In Proc. 18th Int. Conf. Machine Learning, 1926. (2001)

[6] T. Castronova, D. Williams, C. Shen, Y. Huang, B. Keegan, L. Xiong, R. Ratan *As real as real? Macroeconomic behavior in a large-scale virtual world*. New Media and Society. 2009.

[7] Collins, Allan M.; Loftus, Elizabeth F., *A spreading-activation theory of semantic processing*, Psychological Review. 1975 Nov Vol 82(6) 407-428

[8] Mike Consalvo, *Cheating: Gaining advantage in videogames*, The MIT Press, 2007.

[9] J. Dibbell *Owned!: Intellectual property in the age of dupers, gold farmers, eBayers, and other enemies of the virtual state* The State of Play: Law, Games, and Virtual Worlds, J.M. Balkin, and B.S. Noveck eds., New York University Press, 2003.

[10] Richard Heeks *Analysis Current Analysis and Future Research Agenda on gold farming* Informatics Group IDPM, SED, University of Manchester, UK 2008

[11] G. Lastowka, *ID theft, RMT and Lineage*, Terra Nova 2006 http://terranova.blogs.com/terra_nova/2006/07/id_theft_rmt_nc.html.

[12] T. Lehtiniemi *How big is the RMT market anyway*, Virtual Economy Research Network, March 2, 2007.

[13] M. Szummer, T. Jaakkola *Partially labeled classification with Markov random walks*. NIPS 14. 2001.

[14] Dmitri Williams *The mapping principle and a research framework for virtual worlds*, 2008.

[15] D. Zhou , O. Bousquet, T.N. Lal, J. Weston and B. Scholkopf *Learning with Local and Global Consistency*. NIPS, 2004

[16] X. Zhu, Z. Ghahramani, J. Lafferty *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*. ICML 20: 912919. (2003)

[17] Xiaojin Zhu. 2005. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison. http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf