2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Identifying Power Elites in Massively Multiplayer Online Games by Applying Machine Learning to Communication and Support Networks

Siegfried Müller
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
siegfried.mueller@tum.de

Raji Ghawi
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
raji.ghawi@tum.de

Jürgen Pfeffer
School of Social Sciences and Technology
Technical University of Munich
Munich, Germany
juergen.pfeffer@tum.de

***Abstract*—The aim of this paper is to show how machine learning can predict whether an individual is more powerful than others in the group. The crucial point here is to consider the structural position of the actors in the social networks in which they are embedded. The approach we have taken for constructing these intra-group networks is the aggregation of communication and support interactions. Our research is based on longitutional data from the Massively Multiplayer Online Game (MMOG) Travian that was collected over a 12-month period. The data includes 202,764 communication and 96,913 support interactions between players that we applied for the construction of interaction networks. We also had access to status information on a daily basis for 21,431 individual players who were members of 4,758 alliances. Methodically, we applied 10 established metrics from SNA-based team research in combination with the Random Forstest classification algorithm. Our results show that interaction networks are well suited to assign members into two groups of powerful (elite) and non-powerful (non-elite) players. It turned out that the identification of non-elite members was much easier to accomplish than that of elite members. Regarding the application of multiplex networks, we could not confirm a higher explanatory power by using combined networks. In summary, we can say that the network patterns of elite members are clearly different from those of non-elite members. In this way, we were able to predict affiliation to each category with an accuracy (F1) of 0.88 for communication networks and 0.83 for support networks.**

***Index Terms*—Virtual Teams, Social Network Analysis, Elite Subgroups, Leadership, Community Detection, Machine Learning, Communication Network, Support Network, Multiplexity, Massively Multiplayer Online Game**

I. Introduction

"Elites are subgroups of individuals within a society that have the ability and means to influence, lead, govern, and shape societies" [10]. The ability to identify these elite subgroupsis one of the prerequisites for gaining a better understanding of how social groups function. The theoretical literature on elites considers various aspects such as influence on others, role in society, wealth, experience, and fame, among others [7]. A common characteristic that distinguishes elite members from non-elite members is that this subgroup is in a position that grants more influence and power [17]. Raven attributes the de facto power that individuals have to the existence of power bases [21]. These include: reward resources, coercive resources, legitimacy, reference, and the diminishing power of third parties. Traditionally, one would approach the question of whether a person has access to these power bases by looking at their formal position (formal power) within their organization [7]. From the perspective of social network analysis, power can alternatively be viewed "as a property of interorganizational ties, which can be described in terms of resource networks" [20]. In our study, we aim to bring both perspectives together.

Here, we used data from the German-language version of the *Massively Multiplayer Online Game (MMOG)* Travian[1]. In this game, which is organized in rounds that last about a year, hundreds of alliances compete against each other to be the first to build a monument in the end [4]. Just like in the real world, not all players are equal in Travian. Within the alliances there exist elite subgroups of powerful players who have certain *privileges* (power bases) that allow them to reign over ordinary players. Privileges are originally granted by the founder of the alliance, who gives or takes away formal power from the privileged members. The three most important privileges a player can hold are: (1) to grant and revoke privileges, (2) to dismiss members, and (3) to bring in new members. Each of these privileges puts a player in a position to exert significant influence (reward or sanction) over the other team members, thus making them a member of the *elite subgroup*.

The research question we wish to address in this study is whether the structural positions within players' communication and/or support networks allow for successful prediction of membership in these elite subgroups. We proceeded as follows: for each observation period, we created the team's communication and support networks. Using these networks, we were able to identify 10 network metrics for each player (our independent variables). As our dependent variable, we determined whether a player possessed one or more of the above privileges and was thus defined as a member of the elite subgroup. Methodically, we tested different machine learning techniques. It turned out that the "random forest algorithm" that we finally applied provided the best results.

[1] https://www.travian.com/

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

The contributions of this paper are as follows:

1) We show how patterns derived from interaction networks (communication and support) can be used to reveal membership in elite subgroups.
2) We demonstrate that the traditional features, centralization (betweenness and closeness) combined with in-degree and density provide the best results.
3) We achieve up to 0.88 classification accuracy (F1) for communication networks and up to 0.83 for support networks.
4) Finally, we contribute an application example to the relatively new research direction of MMOGs as research environments.

The rest of our paper is organized as follows: section 2 presents related work in this area; section 3 provides a brief description of the game context and its features; further, we describe the data and the steps we took to construct the variables used; section 4 explains those applied network attributes as well as the underlying formulas; section 5 provides an overview of our data and the computed results for the network metrics used; section 6 addresses the prediction task and describes our applied methodology; finally, section 7 and section 8 present results and limitations of our work and provide an outlook for future work.

II. Related Work

MMOGs offer numerous opportunities for the research community [5]. In recent years, researchers have turned significant attention exploring the potentials that those virtual worlds offer [6], [8]. Assmann et al. examined the opportunities that these new research environments provide and saw the wide availability of data and high motivation to participate as a major advantage [4]. Ross and Castronova argued, for example, that if virtual world behavior can be viewed as a model for human behavior in general, the virtual world becomes a powerful tool for empirical social science [22]. Further, they stated that those game environments "free the researcher from the burden of data collection and take advantage of large-scale databases and the computational power of virtual worlds to provide huge datasets that can be generalized to the real world" [22]. With regard to interaction networks, Williams showed that "communication scholars could unobtrusively test theories of group interactions, organizational theory, health communication, communication modalities, social capital, interpersonal behaviors, networks, and countless others" [27]. Regarding the influence of network patterns in communication networks and their impact on team performance, Cummings and Cross did early work [11]. A theoretical framework on the role of communication networks in team's fuctioning was presented by Monge and Contractor [18]. Mehra et al., in turn, addressed the question of the extent to which leaders' network patterns within and outside their group affect team performance and reputation. Besides traditional work environments, for instance, researchers such as Hossain and Zhu examined whether these network patterns also play a role in distributed teams [15]. In the context of MMOGs, Williams and colleagues explored what methodological, measurement, and organizational challenges might be expected [28]. Further, Williams introduced a framework (mapping principle) with the intention to ensure that results from MMOG studies are comparable [27]. These theoretical approaches were combined with machine learning-based predictive models to predict team performance [24], [23], [19]. Szell and Thurner, in turn, tested a number of social dynamics hypotheses to examine whether the patterns found in MMOGs resemble those found in non-virtual human groups [25]. In the area of elite subgroup detection, Corminas-Murtra et al. drew on MMOG data [10]. Others used this interaction data to examine the influence of trust on shared leadership [12]. MMOGs were also used as research environments in the field of research on the emergence of inequality [13]. Other areas of interest focused on group formation processes, where researchers examined parallels between the behavior of gamers in online games and that of street gangs in the real world [1], [16]. In addition to group formation, the exploration of the potentials of community detection represents another growing research stream in this area [3], [2], [14].

III. Data

A. The World of Travian

Travian is a browser-based MMOG that initially was published in 2004. The original data collection for this study took place in 2009/10. At this time the game was operated in 53 countries. With around 150 million players signed up to date, the game is one of the most popular in the strategy games sector. The game is organized in game rounds (servers) on which up to 20,000 players play for the duration of about a year. Travian is highly competitive and can be won only in cooperation. Players start with a village where they grow resources, develop their infrastructure and raise armies to protect their territories. In addition to building and expanding their own villages, one of the most important success factors of the game is being a member of an alliance. Cooperation and coordination within an alliance are crucial factors for success within the game. In consequence, there is usually a clear hierarchy and distribution of tasks within the alliances. This hierarchical structure has its origins in the alliances' formation process. Alliances are established by their founder, who initially holds all the power. New members can subsequently join the alliance at the invitation of the founder. The founder now decides whether to share his power with the new members. This is done by granting subordinates special rights that distinguish them from regular members. This is shown as an example in Figure 1. In addition to a number of functional assignments (e3, e5, e6, e7), there are three privileges (e1, e2, e4) which are associated with a high level of power: (1) the right to grant and to withdraw privileges, (2) the right to dismiss members, and (3) the right to invite new members. The conferral of one of these rights has the effect of giving a member of the alliance a level of power that distinguishes him/her from ordinary members, thus making him/her a member of the elite subgroup.

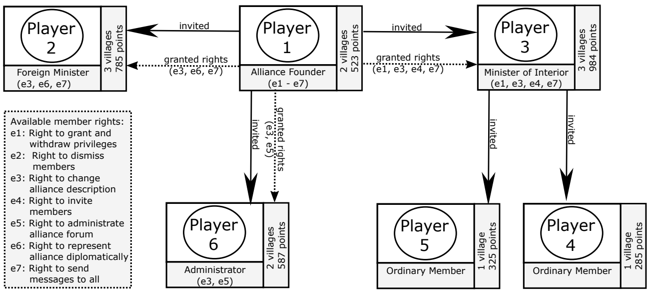


Fig. 1: Exemplary process of formation of an alliance and the origin of power bases.

B. Dataset

The initial data collection that took place in 2009/10 was part of a larger research program on collaboration in teams [4], [12], [26]. For this purpose, the operator Travian Games GmbH granted access to its game databases, which allowed for extensive data collection. The data was provided by the game operator through a daily download of a cleaned version of the game database (MySQL). As for demographic data, some information is known from a general survey [12]. Most of the players were from the German-speaking countries: Germany, Austria and Switzerland. 77% of the respondents were male and on average 30.3 years old. 62% had a permanent job. To ensure privacy, the operator removed all personal information and communication content before transferring the data. The duration of data collection was 51 weeks (358 days). Using this raw data we extracted the following three datasets:

1) Alliance Members: This dataset indicates the membership of players in alliances on a daily basis. Each data row associates the alliance ID (`aid`) with the player ID (`uid`) and the day (relative to the game start date). In addition, each instance contains 7 flags (`e1` to `e7`) that indicate what privileges (power bases) the player holds within the associated alliance on that day.

2) Communication Dataset: This dataset indicates intra-alliance communications among players as expressed on a daily basis. Each data row associates the IDs of two players: the sender and the receiver of a message, with the alliance ID (of which the sender and receiver are members) and the day (during which the message was sent).

3) Support Dataset: This dataset indicates intra-alliance support (resources or troops) among players on a daily basis. Each data row associates the IDs of two players: the sender and the receiver of the support, with the ID of the alliance (of which the sender and receiver are members) and the day (during which the support was sent).

C. Preprocessing

Based on the Alliance Members dataset, we constructed two additional datasets, *Alliance Cases* and *Player Cases*, which we used later as the basis for the construction of communication networks and the prediction of membership in the elite subgroup.

TABLE I: Datasets Summary

Dataset	Alliance Members	Communication	Support
#data rows	2,276,042	831,203	277,234
#unique alliances	4,758	2,074	1,626
#days	358	358	357
#unique users	21,431	18,056	13,389
#alliance-user pairs	42,755	na	na
#player-player pairs	na	202,764	96,913

As a next step, we excluded alliances that have fewer than 5 members. Such alliances do not actually represent teams. In addition, we excluded such cases where the alliance network consists of more than one connected component.

1) Preparation of Alliance Cases: Then, for each alliance, we split its lifespan into overlapping periods. This led to our study cases, i.e., each case is an alliance-period pair. We then used a moving window to split the lifespan, with window length of 60 days, and window step of 30 days. That is, each case comprised alliance ID, period ID (integer starting from 1), period start day and end day, and the actual length of the period (which can be less than 60 for the last period of the alliance). With this step, we obtained 4,972 alliance cases (alliance-period pairs). The period length was 60 days in 3,824 cases, and less than 60 in 1,148 cases. The average period length was 56.44 days. On average, an alliance was associated with 4.33 periods.

2) Preparation of Player Cases: For each alliance case, we extracted the set of players who were members of the alliance during the associated period. For each of the players, we then found the number of actual days when the player was member (it could be less than the period length if the player joined the alliance after the period start day or left it before the period ended). Moreover, we aggregated the flags of the players privileges, `e1`, ..., `e7`, over the associated period. To do so, for each such flag, we first took the average of the daily values (over the period), then we rounded that value to a binary variable (which means either 0 or 1).

The result of this step was a set of player cases, each of which comprised an alliance, a period, and a player, as well as the actual number of days and the 7 aggregated privileges for that player as member of the alliance during the period.

D. Target Variable

Our goal was to use communication and support networks to predict membership in the elite subgroup of an alliance. Although there are seven different privileges in the game Travian, few of them represent power, namely `e1`, `e2` and `e4`. Therefore, we could consider a player who had at least one of these rights as a member of the elite subgroup of the alliance.

The target (i.e., dependent) variable was thus constructed for each player case as a binary flag whose value was:

- 1 : if the player had one of the rights `e1`, `e2` or `e4`, and
- 0 : otherwise (i.e., if the player had other rights or no rights at all).

In the dataset of player cases, there were 17,067 cases (16.65%) with target=1 (i.e., an elite leader), while in the remaining 85,436 cases (83.35%) the target was 0 (other players).

IV. Communication and Support Networks

A. Communication Networks

Based on the communication dataset and the constructed dataset of alliance cases, we built the communication network for each alliance-period pair as a directed unweighted graph. Then, for each alliance-period pair $\langle a, p \rangle$, we took a subset of the communication dataset that comprises communications that occur during the period p between pairs of players u and v who were both members of alliance a. Hence, in this type of network, the nodes are the alliance members, and an edge connects a node u to another v whenever the member represented by u sends one or more messages to another player represented by v, i.e., whenever there was an entry in the communication dataset that associates u to v with the corresponding alliance and period.

In this sense, each network corresponds to an data row in the alliance cases dataset, and each node corresponds to an instance in the player cases dataset.

We then extracted various metrics of those networks, at both network-level and node-level, that we later used as the features (i.e., predictors, or independent variables) for the prediction task.

Network level features

- n: number of nodes (i.e., players).
- m: number of edges (i.e., player-player communications).
- $density$: fraction of actual edges to possible edges:

$$density = \frac{m}{n(n-1)}$$

Node level features

- deg_{in}: In-degree, i.e., number of incoming edges.
- lcc: Local clustering coefficient, i.e., the fraction of pairs of the node's neighbors that are neighbors with each other.
- cc: Closeness centrality of a node is the inverse of the sum of distances to all other nodes.

$$cc(u) = \frac{1}{\sum_{v} d(u,v)}$$

  where $d(u, v)$ is the distance (shortest path length) between nodes u and v. We normalized closeness centrality by multiplying it by $n - 1$.
- bc: Betweenness centrality of a node u is the fraction of shortest paths between any two nodes s and t via u to all shortest paths between s and t:

$$cb(u) = \sum_{s \neq u \neq t} \frac{\sigma_{s,t}(u)}{\sigma_{s,t}}$$

  where $\sigma_{s,t}$ is the total number of shortest paths from node s to node t, and $\sigma_{s,t}(u)$ is the number of those paths that pass through u. Further, we normalized betweenness centrality by multiplying it by $\frac{2}{(n-1)(n-2)}$
- ecc: Eccentricity of a node u is the largest distance between u and all other nodes. A prerequisite for the calculation of eccentricity is that the network consists of only one component.

$$ecc(u) = \max_{v} d(u, v)$$

- f_{kcore}: A flag indicating whether the node belongs to the network k-core, which is the maximum subgraph that contains nodes of degree k or more (we consider the largest k for which the network has a k-core).
- f_{center}: A flag indicating whether the node belongs to the network center, which is the set of nodes that have eccentricity equal to the minimum eccentricity.

Closeness centrality assesses how a node is close to other nodes in the network, while betweenness centrality assesses the extend to which a node lies on the shortest path between other nodes, hence it is a way of detecting the amount of influence a node has over the flow of information in a network

Out of the 4,972 alliance cases, we obtained 3,390 valid communication networks (68%), where the remaining 1,582 networks (32%) were excluded because they were either empty (have no edges) or disconnected (multiple components).

Out of the original 102,503 player cases, we obtained 81,462 valid ones (79.5%) along with their network features, whereas the remaining 21,041 cases (20.5%) were excluded, either because their alliance cases were excluded, or because they were isolated (have no connections).

B. Support Networks

We applied the same steps on the support dataset to construct support networks and extracted their features. Out of the original 4,972 alliance cases, we obtained 2,371 valid support networks (48%), that comprised 42,435 valid player cases (41.4%) out of the original 102,503 player cases, along with their support network features.

V. Analysis

To this end, we obtained two datasets of network features of players (communication and support). Table II shows a summary of those datasets. For instance, in both datasets we can see that four out of five nodes do not belong to the elite.

TABLE II: Summary of Datasets

	Communication		Support	
#alliance-periods	3,390		2,371	
#unique alliances	1,105		885	
#cases	81,462		42,435	
target=0	68,538	(84%)	34,954	(82%)
target=1	12,924	(16%)	7,481	(18%)
#unique players	16,568		10,626	

In order to understand how the different network features are related to the target variable, we first calculated the Pearson correlation, as shown in Table III. We can see that some of the communication networks features have a moderate positive correlation with the target variable; including deg_{in}, f_{kcore} and f_{center}, as well as closeness centrality (cc) and betweenness centrality (bc) which is the feature with the strongest correlation (0.477). On the other hand, other features, such as n, m, lcc and ecc have moderate negative correlation with the target variable.

TABLE III: Pearson correlation of features (*network level* and *node level*) with target variable in both types of networks

	n	m	density
Communication	−0.218	−0.187	0.200
Support	−0.264	−0.210	0.257

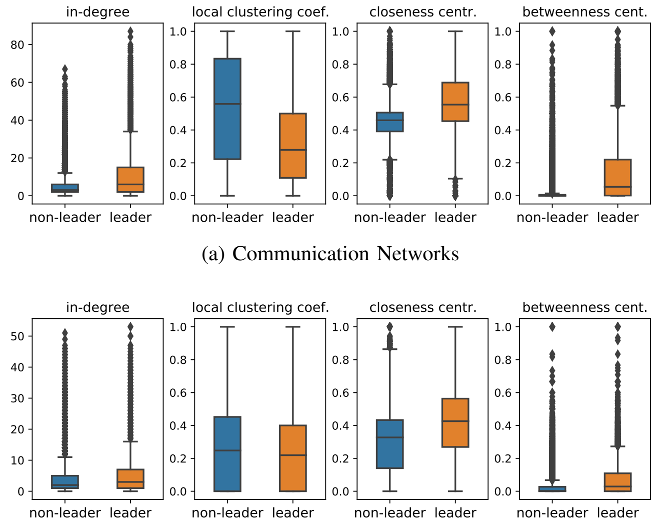
	deg_{in}	lcc	bc	cc	ecc
Communication	0.329	−0.191	0.477	0.326	−0.249
Support	0.110	−0.031	0.267	0.207	−0.257

Having a moderate, positive or negative, correlation for most of the features is a good indication of the ability of those features to predict the target variable, and hence to distinguish between leaders and non-leaders.

We split the data (player cases) into two groups based on the target variable: leaders (target=1) and non-leaders (target=0), and then compared them in terms of the different network features. We find that the leaders group has a higher mean than the non-leaders for four features, namely density, in-degree, closeness and betweenness centrality. In contrast, for the number of nodes n and edges m and the local clustering coefficient lcc, the mean of the leaders group is lower than that of the non-leaders (see Fig. 2).

In order to confirm these findings, we performed one-sided independent t-tests to compare the means of the two groups (leaders and non-leaders) and to check whether the observed differences are significant. In those tests, the null hypothesis is that the two means are equal $H_0 : \mu_1 = \mu_0$; whereas the alternative hypothesis varies from one feature to another. For the number of nodes n and edges m and the local clustering coefficient, the alternative hypothesis is that the mean of leaders group is *less* than non-leaders $H_a : \mu_1 < \mu_0$; whereas for density, in-degree, closeness and betweenness centrality, the alternative hypothesis is that the mean of leaders group is *greater* than non-leaders $H_a : \mu_1 > \mu_0$. The results of the t-tests show that in all cases (all features, and for both communication and support networks) the p-value is smaller than 0.001, which indicates that the results are very significant.

For categorical features, including f_{kcore}, f_{center} and ecc, we show in Table IV contingency tables of those features with the target variable. Using those tables, we can find joint probabilities of the features with the target variable. For instance, with respect to communication networks, among the player cases with target=1 (elite members), 72% have $f_{kcore} = 1$, i.e., belong to the k-core of the network, while only 28% of elite members are not core nodes. On the other hand, 50% of the members of the elite subgroup have $f_{center} = 1$, i.e., are in the network center, and 50% are not. The members of the elite subgroup have a mean eccentricity 2.49 and median 2, while non-elites members have a mean of 3.03 and median 3. That is, elite member players tend to have a lower eccentricity than other players, which means that they tend to be more central



(a) Communication Networks

(b) Support Networks

Fig. 2: Comparison of elite and non-elite players in terms of network features

and closer to the rest of players in the network. With respect to support networks, 70% of the elite members are in the k-core, while 44% of them are in the center.

Moreover, since for these categorical features t-tests are not applicable, we instead performed chi-squared tests (χ^2) to check whether those features are independent from the target variable. As a result of those χ^2 tests for the three variables, we obtain a p-value smaller than 0.001, which means a significant dependence of the target variable on the variables f_{kcore}, f_{center} and ecc.

TABLE IV: Contingency tables of the target variable with the two categorical features: f_{kcore}, f_{center}

		f_{kcore}		f_{center}	
	total	0	1	0	1
target=0	68,538	51,353	17,185	55,037	13,501
target=1	12,924	3,591	9,333	6,503	6,421
sum	81,462	54,944	26,518	61,540	19,922

VI. Elite Member Prediction

In this section, we address the classification of players into elite members (who have privileges e1, e2, or e4) and non-elite members (who have none of these privileges) based on the characteristics of their communication and support networks.

Data: As mentioned earlier, the data instances are player cases (i.e., a player as a member of an alliance during certain period). The target (aka. class) is a binary variable indicating whether an instance (player) is an elite member (target=1) or not (target=0). Our dataset consists of two parts:

- Communication data that comprises 81,462 instances (84% belong to class 0, and 16% to class 1).
- Support data that comprises 42,435 instances (82% belong to class 0, and 18% to class 1).

Features: For both types of networks, we have 10 features (independent variables), three of which are *group-level features*: number of nodes (n) and edges (m), and density; whereas the other seven features are *node-level features*: in-degree (deg_{in}), local clustering coefficient (lcc), closeness- (cc) and betweenness centrality (bc), eccentricity (ecc), as well as f_{kcore} and f_{center} which are two flags indicating whether the node belongs to the network k-core and the network center, respectively. Although most of those features have their values in the range [0,1], some of them have a larger scale, such as n, m and ecc. Thus, in order to put all features on the same scale, we apply min-max normalization on all features.

Training-Testing Split: We split the dataset into two subsets: one of them is used for training (80%), while the other is used for testing (20%), i.e., evaluating the performance of the classifier. Moreover, since the classes are not balanced, i.e., one out of five players is an elite member, we perform the splitting in a *stratified* manner. Stratification means to split the dataset into training and testing sets in such a way that each set contains approximately the same percentage of samples of each target class as the complete set. In this way, the ratio of classes is preserved in both training and testing subsets.

As we split, for the same player, some cases could be put in the training set and the rest in the test set. However, this does not have any effect since each case only comprises network features, without any features about the player's identity.

Algorithm: We tried several classification algorithms, including *k-nearest neighbors (kNN)*, *random forests*, and *logistic regression*. We found however that random forests algorithm outperforms the others, therefore we stick with this algorithm in the reminder of this paper[2].

Results: Table V shows the results of the classification task of elite membership for both communication and support datasets. We can see that accuracy for communication dataset is 0.88, meaning that 88% of the instances in the test set were correctly classified. However, for class 0 (non-elites) the precision and recall, and hence the F1 score, are high above 0.9; whereas for class 1 (elites) those metrics are lower, i.e., precision is 0.67 and recall is 0.49, hence the F1 score is only 0.57. However, the overall weighted average of F1 scores for both classes is still good (0.87). On the other hand, for the support dataset the accuracy is 0.83, which is bit lower compared to the communication dataset. Moreover, precision and recall, and hence the F1 score, are good for class 0, but relatively low for class 1; that is, the F1 of class 0 is 0.90 but for class 1 it is 0.36 only. The overall weighted average of F1 scores for both classes is acceptable (0.80).

These results mean that communication data is better to classify elite members than support data. Moreover, the classification algorithm is more capable of identifying non-elite members than identifying the elite members. More precisely, the bottleneck is about the recall of class 1, i.e., the fraction of true positives, which are the cases correctly classified as positive to all actual positive cases.

[2]We use python implementation as provided by scikit-learn library with default parameters, i.e., no. of estimators=100, criterion='gini', etc.

TABLE V: Classification Results

	Communication			Support		
	precision	recall	F1	precision	recall	F1
class 0	0.91	0.96	0.93	0.86	0.94	0.90
class 1	0.67	0.49	0.57	0.51	0.28	0.36
accuracy			0.88			0.83
weighted avg	0.87	0.88	0.87	0.80	0.83	0.80

In order to get a better understanding of the importance of features and how they contribute to the prediction of elite membership, we repeated the classification tasks several times using different combinations of the features[3]. That is, each time, we use a subset of the features including one, two, or three features as predictors. The results are shown in Table VI where we can observe a very good predictability of class 0 (non-elites) with a F1 score higher than 0.90, whereas class 1 (elites) is much harder to predict. Hence, in the following we focus on the F1 of class 1.

TABLE VI: Classification results when subsets of features are used (communication networks)

Features	Accuracy	F1		
		class 0	class 1	w. avg.
bc	0.85	0.91	0.44	0.84
cc	0.86	0.92	0.42	0.84
lcc	0.85	0.92	0.34	0.82
density, bc	0.87	0.92	0.50	0.86
density, cc	0.86	0.92	0.50	0.85
m, bc	0.87	0.92	0.49	0.86
deg_{in}, density, m	0.88	0.93	0.57	0.87
deg_{in}, m, n	0.87	0.93	0.56	0.87
cc, deg_{in}, m	0.87	0.93	0.56	0.87

If only one feature is used, the best performance can be obtained using betweenness centrality (with F1 = 0.44), followed by closeness centrality (F1=0.42) and local clustering coefficient (F1=0.34). On the other hand, if we use two features, we get the best performance when we combine betweenness or closeness centrality with density (F1=0.50 for both combinations) or when we combine them with the number of the edges m. However, if we use three features, we obtain the best performance using in-degree, density and the number of edges m (F1=0.57). Moreover, the scores we obtain using this combination are almost the same scores obtained using the complete set of features. This means that, for our data, we can achieve the best performance if we use those three features only, and in that case, the rest of features would not significantly improve the performance.

VII. Discussion

In this section, we recap the essential aspects of the elite membership classification.

Elites vs Non-Elites: We found that the scores of class 0 (non-elites) are always higher than the scores of class 1 (elites), meaning that the classification algorithm is more capable of identifying non-elites than identifying the elites. In fact, this manifests in a low recall of class 1, i.e., a low fraction of true

[3]For sake of brevity, we do so for communication data only.

positives, with a high number of false negatives. This can be attributed to two reasons.

First, the inability of the features to clearly distinguish between the two classes. As shown in Fig. 2, most of the features exhibit similar distributions of their values for the two classes. Moreover, the f_{center} feature is useful to identify the instances of class 0 (non-elites), where 80% of such instances have the value 0 (not in the center) for this feature and the remaining 20% have the value 1 (in the center). However this feature is unable to identify the instances of class 1 (elites) as 50% of them have value 0, and the rest 5% have value 1.

Secondly, the low recall of class 1 is mainly an outcome of being an imbalanced class. To mitigate the problem of imbalanced data, we opted to apply a stratified splitting of the dataset into training and testing sets, such that each set contains approximately the same percentage of samples of each target class as the complete set. Other ways to mitigate this problem include resampling the dataset, either by under-sampling the majority class (non-elites) or over-sampling the minority class (elites), as well as generating synthetic samples to randomly sample the attributes from instances in the minority class using SMOTE technique (Synthetic Minority Over-sampling Technique) [9]. We delegated applying such methods to our data to future work.

Feature Importance: We found that centrality features, betweenness and closeness, have the most impact on the predictability of elites, in particular when used alone or combined with density or number of edges. Other important node-level features include in-degree and the local clustering coefficient, whereas eccentricity, f_{kcore} and f_{center} are less important.

On the other hand, we found that although network-level features, particularly density and number of edges, are not useful alone as predictors, they become important when combined with node-level features, particularly centrality features or in-degree. Overall, with a combination of in-degree, density, and number of edges we obtained the best possible performance.

In future work, we could let the data splitting have a validation dataset, such that the selection of features combination is based on the performance of that validation dataset.

Communication vs Support Data: The results discussed so far pertain to applying the classification task independently, i.e., using communication data only, and using support data only. In particular, we found that communication data is better than support data for classifying elite players. However, a natural question arises about the possibility of combining the communication and support data to perform the classification. We attempt to do such a combination in two different ways: network-wise and feature-wise.

In network-wise combination, we merged the communication network with the support network (at alliance-period level) obtaining one network for both, then we calculated the features of the combined network. This merging was done as:

- overlap: we connected two nodes if they were connected in both networks.
- union: we connected two nodes if they were connected in either network.

The overlap network comprised 28,846 instances, whereas the union network comprised 83,050 data rows. Using the features of those networks as predictors, we obtained a weighted average F1 score of 0.80 for the overlap network, and 0.87 for the union network.

In feature-wise combination, we merged the features of the communication networks and the support network, obtaining a single dataset with 20 features (10 about communication and 10 about support network). The combined dataset comprised 39,915 rows (slightly fewer than the instances of the support dataset), and the resulting weighted F1-score is 0.88, which is slightly better than using the communication dataset only.

Effects of Other Factors: We considered two factors and examined how they affect the classification performance.

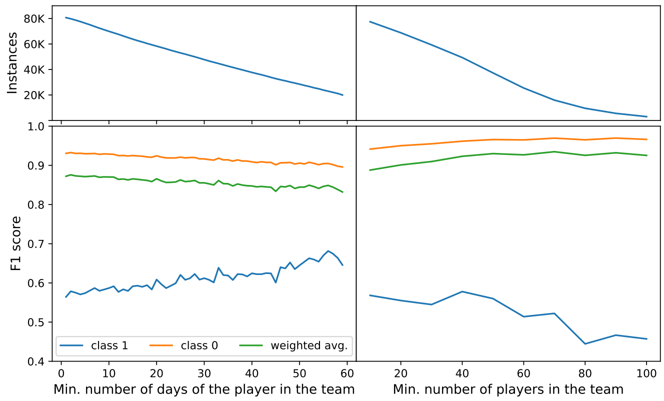


Fig. 3: Effects of number of days, and number of players on the classification performance

The first factor is the number of days that a player spends as a member of the team during the studied period. We used this factor to trim the dataset, i.e., we excluded player cases that have a number of days less than a minimum threshold. As we increased the threshold, the number of data rows (player cases) would obviously decrease (Fig. 3 left), but we observed that the F1 score of class 1 (elites) increased. For instance, if we excluded the data rows with less than 56 days, we obtained an F1 score of 0.68 for class 1, which is higher than the original score (with all data rows), however, in that case, the number of data rows was $\sim$ 23K only, which is about the quarter of all data rows. Moreover, the F1 score of class 0 (non-elites) decreased, and so did the weighted average F1 score.

The second factor is the team size, i.e., the number of players in the team during the studied period. We also used this factor to trim the dataset by excluding team cases that have a number of players less than a minimum threshold. As we increased the threshold, the number of team cases and the number of data rows (player cases) obviously decreased (Fig. 3 right). Here we observe that the F1 score of class 0 (non-elites) is almost stable (it increases very slightly at the beginning), and so does the weighted average F1 score. However, in this case the F1 score for class 1 decreases.

Overall, we found that both factors have a slight impact on the classification results, with some trade-off between the two classes of players, elites and non-elites. Thus, it does not seem necessary to use either of them in an actual trimming of the datasets, as this would reduce the data rows without a significant improvement in the classification performance.

VIII. Conclusion

In this work, we wanted to explore whether elite membership can be predicted through the use of interaction networks. Our results clearly show that this is possible. We were able to show that elite members exhibit different patterns of behavior in their interaction networks (i.e., communication, support) than non-elite members of a team. Here, we observed that, while identifying non-elite members was relatively easy, identifying elite players required a much more sophisticated approach. One reason for this could be our definition of elite membership. This definition is based on the assumption that formal power (the possession of privileges) reflects the actual reality in teams. This conceptual approach might be the most critical assumption of our current paper. For example, if a certain percentage of elite members are inactive during the observation period, our classification approach will fail because these members leave no trace of their interactions. In our discussion, we addressed this issue in part by looking at team size and the number of days a player is a member of the team. The results presented there clearly show that (at the cost of a reduced data sample) the problem of identifying elite members (class 1) decreases when inactivity (numbers of days active) can be tracked. Future work should therefore focus more on the extent to which the data can represent reality. In addition, our results indicate that group size deserves more attention. This should also deserve additional attention in future work. Given these limitations and the need for future research, we believe that this work has allowed us to show that the combined use of interaction networks and machine learning offers good opportunities to learn more about the formation and detection of elite subgroups in teams.

References

[1] Muhammad Aurangzeb Ahmad, Zoheb Borbora, Cuihua Shen, Jaideep Srivastava, and Dmitri Williams. Guild Play in MMOGs: Rethinking Common Group Dynamic Models. In Anwitaman Datta, Stuart Shulman, Baihua Zheng, Shou-De Lin, Aixin Sun, and Ee-Peng Lim, editors, *International Conference on Social Informatics*, volume 6984 of *Lecture Notes in Computer Science*, pages 145–152, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[2] Hamidreza Alvari, Alireza Hajibagheri, Gita Sukthankar, and Kiran Lakkaraju. Identifying community structures in dynamic networks. *Social Network Analysis and Mining*, 6(1):1–13, 2016.

[3] Hamidreza Alvari, Kiran Lakkaraju, Gita Sukthankar, and Jon Whetzel. Predicting Guild Membership in Massively Multiplayer Online Games. In William G. Kennedy, Nitin Agarwal, and Shanchieh Jay Yang, editors, *Social Computing, Behavioral-Cultural Modeling and Prediction. SBP 2014. Lecture Notes in Computer Science*. Springer, Cham, 2014.

[4] Jakob J. Assmann, Marcus Drescher, Julia V. Gallenkamp, Arnold Picot, Isabell M. Welpe, and Rolf T. Wigand. MMOGs as emerging opportunities for research on virtual organizations and teams. *16th Americas Conference on Information Systems, AMCIS 2010*, 2010.

[5] William Sims Bainbridge. The scientific research potential of virtual worlds. *Science*, 317(5837):472–476, 2007.

[6] John C. Beck and Mitchel Wade. *Got Game: How the Gamer Generation is Reshaping Business Forever*. Professional development collection. Harvard Business School Press, 2004.

[7] Tom Bottomore. *Elites and Society*. Taylor & Francis, 1993.

[8] Edward Castronova. *Synthetic Worlds*. The University of Chicago Press, Chicago, 1 edition, 2005.

[9] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, jun 2002.

[10] Bernat Corominas-Murtra, Benedikt Fuchs, and Stefan Thurner. Detection of the elite structure in a virtual multiplex social system by means of a generalised K-core. *PLOS ONE*, 9(12):1–19, 2014.

[11] Jonathon N. Cummings and Rob Cross. Structural properties of work groups and their consequences for performance. *Social Networks*, 25(3):197–210, 2003.

[12] Marcus Drescher, Audrey Korsgaard, Isabell Welpe, Arnold Picot, and Rolf Wigand. The dynamics of shared leadership: Building trust and enhancing performance. *Journal of Applied Psychology*, 99(5):771–783, 2014.

[13] Benedikt Fuchs and Stefan Thurner. Behavioral and network origins of wealth inequality: Insights from a virtual world. *PLoS ONE*, 9(8), 2014.

[14] Alireza Hajibagheri, Gita Sukthankar, Kiran Lakkaraju, Hamidreza Alvari, Rolf T. Wigand, and Nitin Agarwal. Using Massively Multiplayer Online Game Data to Analyze the Dynamics of Social Interactions. In Kiran Lakkaraju, Gita Sukthankar, and Rolf T. Wigand, editors, *Social Interactions in Virtual Worlds*, chapter 14, pages 375–416. Cambridge University Press, Cambridge, 1 edition, 2018.

[15] Liaquat Hossain and David Zhu. Social networks and coordination performance of distributed software development teams. *Journal of High Technology Management Research*, 20(1):52–61, 2009.

[16] Neil F. Johnson, Chen Xu, Zhenyuan Zhao, Nicolas Ducheneaut, Nicholas Yee, George Tita, and Pak Ming Hui. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review E*, 79(6):1–12, 2009.

[17] Wright C. Mills. *The Power Elite*. Oxford University Press, New York, 1959.

[18] Peter R. Monge and Noshir S. Contractor. *Theories of Communication Networks*. Oxford University Press, Oxford, 1 edition, 2003.

[19] Siegfried Müller, Raji Ghawi, and Jürgen Pfeffer. Using Communication Networks to Predict Team Performance in Massively Multiplayer Online Games. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 353–360, The Hague, Netherlands (virtual), Dec 7-10, 2020.

[20] Robert Perrucci and Marc Pilisuk. Leaders and Ruling Elites: The Interorganizational Bases of Community Power. *American Sociological Review*, 35(6):1040–1057, 1970.

[21] Bertram H. Raven. The Bases of Power: Origins and Recent Developments. *Journal of Social Issues*, 49(4):227–251, 1993.

[22] Travis L. Ross, Edward Castronova, and Gert G. Wagner. Empirical Research Methods in Virtual Worlds. In Carlos Nunes Silva, editor, *Online Research Methods in Urban and Planning Studies: Design and Outcomes*, chapter 18, pages 299–311. IGI Global, 2012.

[23] Kyong Jin Shim, Samarth Damania, Colin Delong, and Jaideep Srivastava. Player and team performance in EverQuest II and Halo 3. *IEEE Potentials*, 30(5):21–26, 2011.

[24] Kyong Jin Shim, Richa Sharan, and Jaideep Srivastava. Player performance prediction in massively multiplayer online role-playing games (MMORPGs). In *Advances inKnowledge Discoveryand Data Mining*, volume 6119 LNAI, pages 71–80, 2010.

[25] Michael Szell and Stefan Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32(4):313–329, 2010.

[26] Rolf T. Wigand, Nitin Agarwal, O. Isaac Osesina, Winfried Hering, Audrey Korsgaard, Arnold Picot, and Marcus Drescher. Social network indices as performance predictors in a virtual organization. *Proceedings of the 2012 4th International Conference on Computational Aspects of Social Networks, CASoN 2012*, pages 144–149, 2012.

[27] Dmitri Williams. The mapping principle, and a research framework for virtual worlds. *Communication Theory*, 20(4):451–470, 2010.

[28] Dmitri Williams, Noshir Contractor, Marshall Scott Poole, Jaideep Srivastava, and Dora Cai. The Virtual Worlds Exploratorium: Using Large-Scale Data and Computational Techniques for Communication Research. *Communication Methods and Measures*, 5(2):163–180, 2011.