Using Communication Networks to Predict Team Performance in Massively Multiplayer Online Games

Siegfried Müller
Bavarian School of Public Policy
Technical University of Munich
Munich, Germany
siegfried.mueller@tum.de

Raji Ghawi
Bavarian School of Public Policy
Technical University of Munich
Munich, Germany
raji.ghawi@tum.de

Jürgen Pfeffer
Bavarian School of Public Policy
Technical University of Munich
Munich, Germany
juergen.pfeffer@tum.de

***Abstract*—Virtual teams are becoming increasingly important. Since they are digital in nature, their "trace data" enable a broad set of new research opportunities. Online Games are especially useful for studying social behavior patterns of collaborative teams. In our study we used longitudinal data from the Massively Multiplayer Online Game (MMOG) Travian collected over a 12-month period that included 4,753 teams with 18,056 individuals and their communication networks. For predicting team performance, we selected 13 SNA-based attributes frequently used in team and leadership research. Using machine learning algorithms, the added explanatory power derived from the patterns of the communication networks enabled us to achieve an adjusted $R^2 = 0.67$ in the best fitting performance prediction model and a prediction accuracy of up to 95.3% in the classification of top performing teams.**

***Index Terms*—Performance Prediction, Virtual Teams, Social Network Analysis, Communication Network, Machine Learning, Massively Multiplayer Online Game**

I. Introduction

Traditionally, teams have worked together at the same location, whereas today, virtual teams have become a reality in most organizations [21]. A study conducted by the Society for Human Resource Management stated that approximately 66% of multinational organizations utilize virtual teams [14]. These new ways of collaboration produce a vast amount of "digital exhaust," as Leonardi & Contractor call the electronic traces created by modern communication technologies (e.g., e-mail, messenger or VOIP) [20].

When studying teams and their functioning, "communication has always been viewed as a key element" [19]. Especially when team members have never met in person, their communication "is often the only visible artifact of the group's existence" [1]. "Relational theories have depicted leadership as socially constructed through communication exchanges" [9]. "Scholars adopting the social network approach further argue that by focusing on informal social contexts, i.e., social networks, researchers can examine 'how work really gets done in organizations' [8]" [17]. Cross and Parker claim that "one has to examine how people are connected to each other and to focus on the wider social environment rather than formal dyadic relations between a leader and her followers" [17].

Beside traditional work environments, a very promising field of research for studying teams is the analysis of Online Games. Within these virtual worlds, thousands of players are organized in virtual teams [6]. The Harvard Business Review stated: "Online game leaders operate in a context that may well foreshadow the business environment of the future" [26]. *Massively Multiplayer Online Games (MMOGs)* in particular allow and also require cooperation and competition on a large scale. Unlike in traditional experiments, the participants solve engaging problems and challenges. It is not necessary to incentivize participants since they are already highly motivated through the social interaction and the game design [2]. Castronova states that, even if "this place isn't 'real' by any means [...], it does feel real enough to the users that they can fairly easily immerse themselves in it for hours on end, month after month, year after year, in a sort of parallel existence" [5].

In our study we applied data from an international strategy game called *Travian*[1]. The game is organized in rounds lasting approximately one year. Therefore, repeated interaction between users plays an important role in the games' social ecosystem. Within the early phase, players team up with others to form alliances containing up to 60 members. These alliances are necessary to protect each other and to achieve the goal, which is to be the first to complete a monument at the end of the game. The game can only be won in cooperation and coordination. Therefore, intra-alliance communication plays an important role in succeeding. Communication takes place in an in-game messaging system that is part of the game server.

We conducted our research within the environment of an online simulation game. The motivation for our research was to show that communication networks (or the information they contain) can be applied as predictors for team performance. Therefore, we developed two distinct models. We applied a **baseline model** to enable our prediction task to cover the main effects originating from the game design. This baseline model includes the age, the time since formation of the team, and the group size (N), which plays an important role as the game favors alliances bigger in size. Secondly, we build a **network model** that extends the baseline model by including 13 network attributes, commonly used in team and leadership literature. To further improve our results we finally developed a **logarithmic model** extending our network model with the logarithm of the applied features (if available).

The contributions of this paper are as follows:

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

IEEE/ACM ASONAM 2020, December 7-10, 2020

1) We demonstrate how social network patterns in communication networks can be applied to predict team performance.
2) We provide an overview of the ability of different machine learning approaches to deliver accurate prediction outcomes.
3) We achieve an accuracy of 0.67 in prediction of team performance using our extended network model.
4) We achieve an accuracy of 0.95 in classification of top performing team.
5) Finally, we deliver insight into a set of general aspects to consider when tackling the world of MMOG datasets.

The rest of our paper is organized as follows: section 2 shows related work; section 3 describes the mechanics of the game, the dataset and our preprocessing steps; section 4 shows how we calculate network features; sections 5 through seven describe how we analyze, predict performance and carry out classifications; section 8 is the conclusion and describes limitations and future work.

II. Related Work

Online games are not limited to their potential as a laboratory where leadership and its outcomes can be studied. “Anthropologists see new cultures, entrepreneurs see new markets, lawyers see new precedence, and social and political experts see new pressures and looming crises” [5].

Given the “scientific research potential of virtual worlds” as discussed in the 2007 article in Science [3], the fields of application are wide—especially in the area of team research, where Massively Multiplayer Online Games (MMOGs) and Massively Multiplayer Role-Playing Games (MMORPGs) have been widely used [24], [6], [30]. Assmann et al. assessed the “opportunities to overcome some limitations of traditional research environments” [2]. They point out that they “offer a unique opportunity to study virtual organizational structures” [2]. In communication research for example, Gloor et al. [15] have been working on how online communication behavior can be optimized and how it is influencing individual and team performance. Williams et al. [36] conducted an interdisciplinary “study of behavior within a game and also game activities that parallel those in ‘real life’”, whereas Korsgaard et al. [18] worked in the area of “emergence and persistence of trust and cooperation, as well as [in the area of] the impact of different communication media for coordination and information management in virtual organizations”. Other work investigated the effect of shared leadership within groups and its relationship with group trust development [12]. Further, MMOGs have been applied as research frameworks for military training and education [4], [10]. Even combat activities within these games have been studied [16], [31].

Regarding the prediction of team performance, team processes and leadership behavior have been studied frequently. Pobiendina et al. [25] for example, have used role distribution, experience, the number of friends, and national diversity in Dota2 to study their influence on team performance. Prediction models on team performance have been applied and tested by Shim et al. [29], [28] using data from EverQuest and Halo3. Working with data from Travian, Wigand et al. [35] proposed using centrality measures from the game’s message network as performance indicators and for predictive modeling.

III. Data

A. The World of Travian

Travian is a commercial *Massively Multiplayer Online Game (MMOG)* operated in 53 countries around the world. Up to 20,000 users play at any one time in game worlds adapted to the local market. The game world used for this study (travian.de) is a version that has been localized for German-speaking countries. The players start with one village where they grow resources, level up their infrastructure and build armies to protect their kingdom. Troops can also be used to raid resources from other players, instead of producing those resources themselves, or to fight wars to conquer new territories.

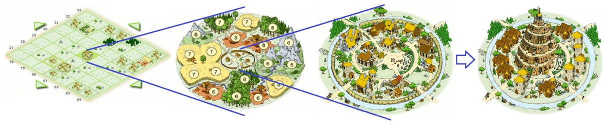


Fig. 1: Travian - Screenshots from the game indicate different zoom levels (map, fields, and village) and final monument

In addition to founding and developing new villages, the most important aspect of the game is to be part of an alliance. The environment of the game is highly competitive and only a high degree of cooperation allows a team to survive and achieve its goals. The alliance leaders are highly dependent on the contribution of every single member. Therefore, there is a great amount of social pressure to take things seriously and to invest a significant amount of time. Players who do not show a certain amount of commitment and/or performance (e.g., growth rate) are not invited to join alliances or are even dismissed. Alliance leaders face a trade-off when it comes to achieving a high ranking position. The easiest way to increase alliances ranking is to invite additional members to the alliance. But doing this comes at a price. Leading and coordinating bigger groups/organizations is challenging and evidence from the game shows that often a smaller team of highly experienced players is more effective in reaching their goals. Therefore, some—not all—top-ranked alliances opt to remain small in number rather than expand to include the maximum 60 members allowed.

To enable communication between players, the game provides an in-game messaging system and an (internal) forum that can only been accessed by a specific alliance. For our study, we used messages sent via the IGM, which means that our data collection has been non-obtrusive and not reactive. All players were informed by the game operator about their (anonymous) participation in a scientific research project to which they agreed by accepting the general terms and conditions. Completing additional surveys [2] has been voluntary and had no impact on regular participation in the game.

Alliances (as teams are called) can be established by players whose villages reach a certain threshold/number of inhabitants. An invitation is required for joining a team. The game tracks when this invitation has been sent and when it has been accepted. The same applies when members are leaving the team or have been dismissed. Teams can therefore be regarded as having clearly defined boundaries.

B. The Dataset

In 2009/10, the operator Travian Games GmbH granted access to its game databases, which enabled an extensive data collection for scientific research. The operator of the game provided a daily download of a *cleaned* version of the game database (MySQL). The majority of the players were from the German-speaking countries: Germany, Austria and Switzerland. Participants were 77% male, averaging 30.3 years old. 62% had a permanent employment. To comply with privacy protection, the operator removed all personal information and communication content before sharing the data with the researchers.

Alliance size ranged from 2 to 60 members, which is the maximum number of members that the game design allows. On average, these groups were sized 14.5 individuals. A total of 4,758 alliances have been formed during this particular game. The data collection period was 51 weeks (356 days). Using this raw data we extracted the following two datasets:

1) Performance Dataset: The game *Travian* uses specific rankings, also referred to as *alliance rankings*, to indicate alliance performance. Rankings are based on the sum of inhabitants each alliance member has. The number of inhabitants a player has under him increases each time the player's infrastructure is upgraded. The alliance with the most inhabitants is rated as number one, the alliance with the second-most inhabitants as number two and so on. Rankings within the game are calculated in real time. Since our raw data only contained one data point (MySQL snapshot) per day, we reverse engineered the ranking algorithm and adapted it via aggregation to a weekly measure. The decision for weekly aggregation was based on preliminary analysis of the data to avoid artifacts from the daily snapshots.

As the game proceeds, players' villages develop and the overall number of inhabitants increases constantly. Figure 2 shows how the number of inhabitants evolves over time. In order to use the number of inhabitants as a performance measure, we needed to normalize it in a way that makes it comparable across weeks since start of the game world. Thus, we used min-max normalization on a weekly basis. Let $H(a, w)$ denote the number of inhabitants of alliance a at week w, then: $H_{min}(w) = \min_a\{H(a, w)\}$ and $H_{max}(w) = \max_a\{H(a, w)\}$ are, respectively, the minimum and maximum number of inhabitants per alliance at week w. The performance $P(a, w)$ of alliance a at week w is then stated as:

$$P(a,w) = \frac{H(a,w) - H_{min}(w)}{H_{max}(w) - H_{min}(w)}$$

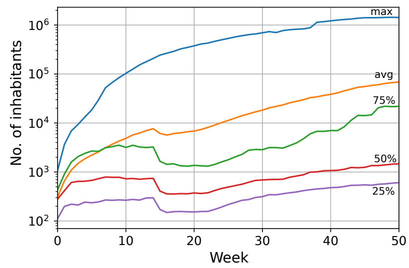


Fig. 2: Evolution of number of inhabitants per alliance over time

2) Communication Dataset: This dataset indicates intra-alliance communications among players as expressed on a weekly basis. Each entry 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 week ID (during which the message was sent).

Table I provides some statistics about both datasets, including the number of records, number of alliances, number of weeks, and number of alliance-week pairs.

TABLE I: Statistics of datasets

Dataset	Performance	Communication
No. of records	53,766	526,002
No. of alliances	4,753	2,074
No. of weeks	51	52
No. of alliance-week pairs	53,766	16,532

C. Pre-processing

From Table I, we observe that the two datasets have different numbers of alliances, weeks and alliance-week pairs; hence, there are some incompatible data entries. For instance, there are some data entries that appear in the `performance` dataset but not in the `communication` dataset, and vice versa. Moreover, in some alliance-week pairs, the number of alliance members in the `performance` dataset is different from the number in the `communication` dataset. To fix these issues, we performed the following pre-processing steps.

- Since we did not possess performance data within the first week of existence of some alliances, we opted to exclude this first week of *all* alliances.
- Since some alliances have missing communication information at the end of their lifespan, we opted to exclude the last week(s) of those alliances.
- To fix the discrepancy in the number of alliance members between the two datasets, we opted to use the maximum of these two numbers as the number of alliance members, for all alliance-week pairs.

Overall, as a result of pre-processing steps, we got rid of incompatible data. To this end, the `communication` dataset consists of (the remaining) 14,954 alliance-week pairs

[2] not part of this study

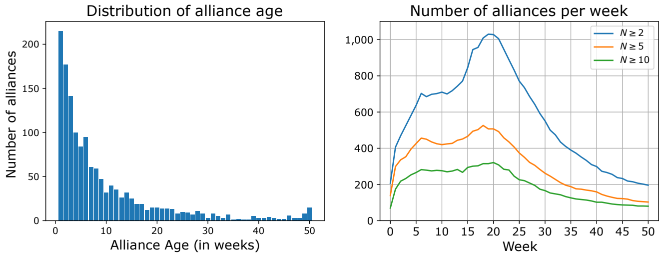


Fig. 3: Alliance distribution over time (N: alliance members)

(corresponding to 50 weeks, and 1,852 alliances). The number of remaining entries is reduced to 510,285 (97%).

Figure 3 gives an overview of the distribution of alliances over time. Figure 3-a shows a histogram of the alliance age (in weeks), where we observe a skewed relationship between the age and the number of alliances having that age (survived that number of weeks). Most alliances have a relatively short lifespan, whereas few alliances survived for an extended period of time. Figure 3-b shows how the number of alliances changes over the entire period of the game.

IV. Communication Networks

Based on the `communication` dataset, we constructed the communication network as a directed graph for each alliance-week pair. Since we are interested in network structure, not in communication frequency, we opted to use the unweighted version of the graph. In this type of network, the nodes are the alliance members, and an edge links 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 is an entry in the communication dataset that associates u to v with the corresponding alliance and week. Figure 4 and 5 show two examples.

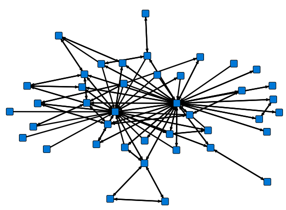


Fig. 4: Alliance 203, week 12

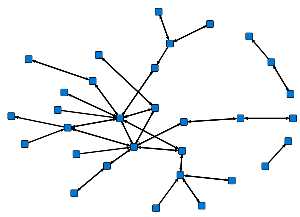


Fig. 5: Alliance 2, week 45

Overall, we have communication networks for 14,954 alliance-week pairs (corresponding to 1,852 alliances, and 50 weeks). In addition to the number of nodes, N, and the number of edges, E, we calculated several network metrics for each network, including: density, average in-degree, transitivity, reciprocity, centralization, and k-core:

- **Density**: the ratio of the number of actual edges to the number of possible edges: $density = 2E/N(N-1)$
- **Average in-degree** (avg_din).
- **Transitivity**: the fraction of present triangles to all possible triangles (triads).
- **Reciprocity**: the ratio of the number of edges pointing in both directions to the total number of edges.

A. Centralization

In network analysis, centrality is a node-level index of the structural importance of nodes. Many metrics have been developed in the literature to measure the centrality of nodes, including degree centrality, closeness centrality, and betweenness centrality [13], [34]. Let $c_1, \cdots, c_n$ be node-level centrality measures, where c_i is the centrality of node i by some metric. It is often useful to standardize the c_i 's by their maximum possible value: $\widetilde{c_i} = c_i/c_{max}$

While *centrality* is a node-level index, *centralization* is a group-level index that refers to how centralized the network is i.e., to what extent is there a small number of highly central nodes? Let $c^* = max\{c_1, \cdots, c_n\}$. Let $S = \sum_i [c^* - c_i]$. Then $S = 0$ if all nodes are equally central; S is large if one node is most central. Thus, network centralization is stated as:

$$C = \frac{\sum_i [c^* - c_i]}{\max \sum_i [c^* - c_i]}$$

where the "max" in the denominator is over all possible networks. With this formula, we get $0 \leq C \leq 1$. In particular, $C = 0$ when all nodes have the same centrality (e.g., cycle); whereas $C = 1$ if one actor has maximal centrality and all others have minimal (e.g., star). Table II shows the formulas for degree centralization, closeness centralization and betweenness centralization.

TABLE II: Centralization formulas

Degree centralization	$C^d = \frac{1}{2(N-1)(N-2)} \sum_i [c^{d*} - c_i^d]$
Closeness centralization	$C^c = \frac{2N-3}{3(N-1)(N-2)} \sum_i [\widetilde{c}^{c*} - \widetilde{c}_i^c]$
Betweenness centralization	$C^b = \frac{1}{N-1} \sum_i [\widetilde{c}^{b*} - \widetilde{c}_i^b]$

For our alliance communications networks, we actually calculated five centralization metrics: three versions of degree centralization using in-degree, out-degree, and degree, as well as closeness and betweenness centralization.

B. k-Core

A k-core is a maximum subgraph that contains nodes of degree k or more. In our networks, for each node we find the *core number*: $kcore(u), u \in G$, from which we then compute three network features:

- kcore k_{max}: $k_{max} = \max_{u \in G}\{kcore(u)\}$
- kcore size: number of nodes in the k-core, i.e., nodes whose core number is k_{max}.
- kcore relative size: fraction of nodes in the k-core to all nodes in the network.

To this end, we obtain a new dataset that summarizes the communication network of each of the $\langle$ alliance, week $\rangle$ pairs (14,954 pairs). Where each pair is associated with 14 attributes.

As a new attribute, we introduce the *age* of the alliance, as we assume a strong interdependence with the maturity of the group. The age of an alliance at a given week (stated in weeks) is the number of weeks elapsed since the alliance creation. Formally, the age of an alliance a at week w is stated as:

$$age(a, w) = w - firstweek(a) + 1$$

where $firstweek(a)$ indicates the first week within the lifespan of the alliance a, and is given by $firstweek(a) = \min\{w \mid (a, w) \in S\}$. We also excluded the cases with the number of alliance members $N \leq 3$. The result of this step was the removal of 1,542 alliance-week pairs. Hence, the dataset then consisted of the remaining 13,412 alliance-week pairs.

The last step is to join the `network attributes` dataset with the `performance` dataset, such that for each alliance-week pair, we have the network features along with the performance of the alliance during that week.

V. Analysis

Now, since we have completed our dataset, we start looking at the features we have at hand. First we look at the correlation of these features with the target feature, the performance of an alliance in a week. This can be seen in figure 6 (including alliance age). We observe that the features that are the most correlated with the performance are the number of nodes N (alliance members), and the number of edges E. Some other features also have a relatively high positive correlation with the performance, including the avg. in-degree, k-core k, k-core size, and age. There are also other features that show a relatively high negative correlation with the performance, including the closeness centralization (cntrz_cc), and k-core relative size (kcore_rel_size). The remaining features have weak, positive or negative correlation, such as density, centralization, transitivity and reciprocity.

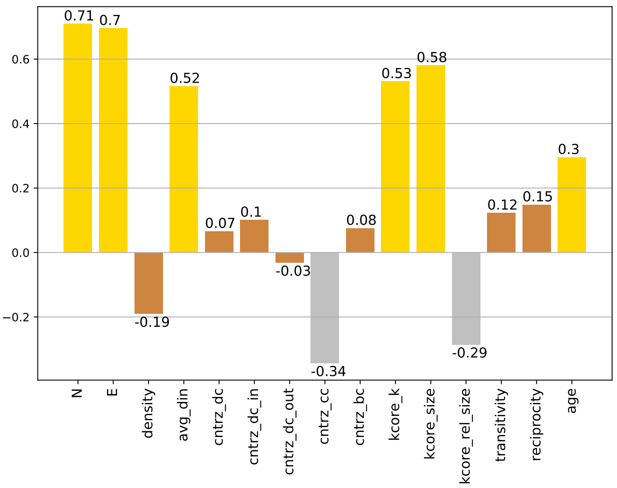
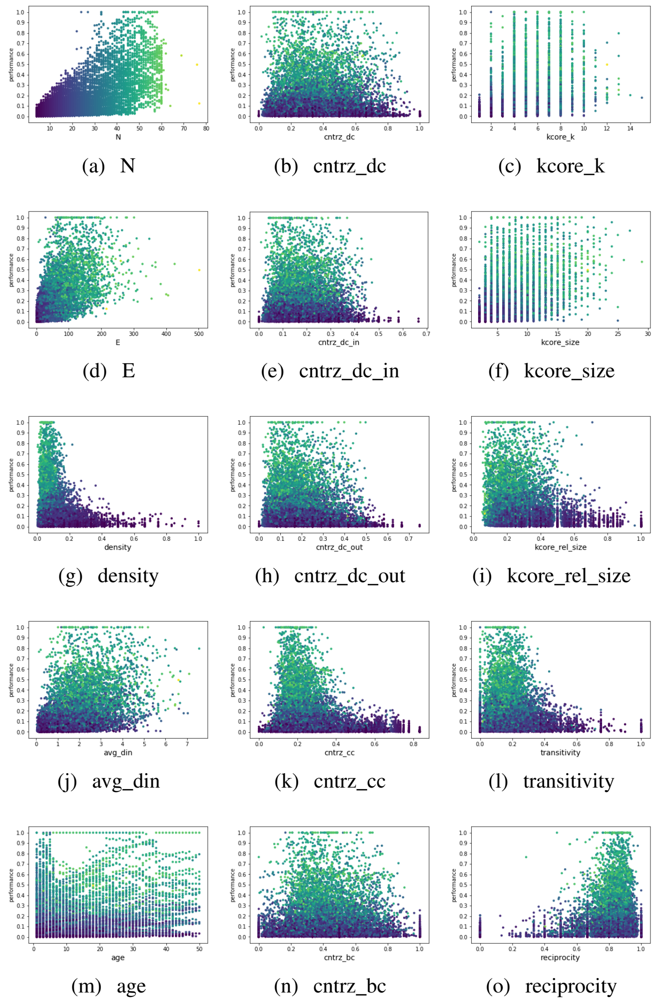


Fig. 6: Correlation of network attributes with the performance.

In order to have insight into how the performance is related to each feature, Figure 7 shows scatter plots of each feature with respect to the performance.



(a) N (b) cntrz_dc (c) kcore_k (d) E (e) cntrz_dc_in (f) kcore_size (g) density (h) cntrz_dc_out (i) kcore_rel_size (j) avg_din (k) cntrz_cc (l) transitivity (m) age (n) cntrz_bc (o) reciprocity

Fig. 7: Scatter plots of network features with performance. The data points are colored based on N, the number of nodes/members (lighter points indicate more members)

VI. Performance Prediction

In this section, we address the prediction of the alliance performance based on the network attributes. For this purpose, we used our final dataset, where we first split it into 80%, 20% training-test sets. As a prediction algorithm, we used the classic Linear Regression approach, as implemented in the LinearRegression module of python's scikit-learn library[3].

For the evaluation of the prediction accuracy, we use the coefficient of determination (R^2), which is stated as follows:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where y_i and $\hat{y}_i$ are the actual and predicted values of the target variable (performance). The problem with R^2 is that it automatically and spuriously increases when extra explanatory

[3]https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

variables are added to the model. As an attempt to take account of this phenomenon, an adjusted R^2 is typically used, that adjusts for the number of explanatory terms in a model relative to the number of data points [32].

Since linear regression assumes that the variables are distributed normally, we needed to check the normalcy of the outcome variable, which is the performance in our case. Figure 8 shows the histogram plot of the performance, which exhibits a skewed distribution. Therefore, we also considered alternatives such as the logarithm and the square root of the performance. In our experiments, we used these three outcome metrics, and compared the prediction of these metrics using several models.

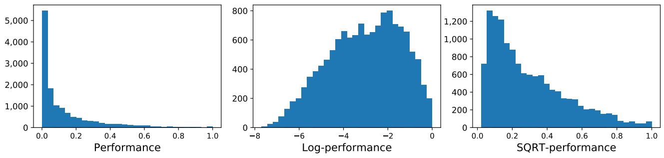


Fig. 8: Performance, Log-Performance, and SQRT-performance.

We developed three models for the prediction task:

- **Baseline model**: The purpose of this model is to cover the main effects originating from the game design, i.e., the features that are not related to the communication network, namely, the number of alliance members N, and the alliance *age*. Actually, having more alliance member automatically leads to a higher ranking position. Holding a certain limit in alliance members is required, but not sufficient for reaching a high ranking position. Therefore, we included *group size (N)* to capture this effect. Secondly, groups need time to form and to arrive at the *performing stage* [33]. Therefore, we included time since foundation of the alliance (*age*).
- **Network model**: As a second step, we extended our model by adding 13 network features derived from the intra-alliance communication networks. To capture (collective) leadership structures, we included *density*, *centralization* [11], [23] and *k-core* [27], [7]. We used *average in-degree* to track prestige [22]. Finally, we included *transitivity* and *reciprocity* to cover the most important structural tendencies [34].
- **Logarithmic model**: This model extends the network model further because it includes all the features mentioned above plus the logarithm of these features. Some features have 0 values, such as the five centralization features, transitivity, and reciprocity. Thus, the logarithm of those features is not available.

Table III shows these different models and the features included in each model.

We used the 80% training set to train the linear regression model, and the 20% test set is then used to test the trained model. The prediction results, in terms of adjusted R^2, for the three models (baseline, network, and logarithmic) and for the

TABLE III: Prediction models

	N	E	density	avg_din	cntrz_dc	cntrz_dc_in	cntrz_dc_out	cntrz_cc	cntrz_bc	kcore_k	kcore_size	kcore_rel_size	transitivity	reciprocity	age
Baseline model	×														×
Network model	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Logarithmic model	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
log(.)	×	×	×	×						×	×	×			×

three outcome variables (performance, log-perf. and sqrt-perf.) are shown in Figure 9.

	Base model	Network model	Log model
Performance	0.5195	0.5924	0.6229
Log-Perf.	0.5051	0.6192	0.6224
SQRT-Perf.	0.5759	0.6544	0.6740

Fig. 9: Prediction results

First let us compare the three outcome variables. We observed that for the three models, the sqrt-performance can be predicted more accurately than the performance or log-performance. The log-performance can be predicted more accurately than the performance using the network model, while it is as precise as the performance using the logarithmic model. However, using the baseline model, the performance can be predicted more precise than the log-performance.

We observed that the baseline model is better suited than the network model for predicting the performance, but the network model is better suited for predicting the log-perf. and sqrt-perf. Finally, we observed that the logarithmic model is the best suited model in terms of the three outcome variables. In particular, the best prediction result is achieved using the logarithmic model when predicting the sqrt-perf, namely, the adjusted $R^2 = 0.67$.

VII. Classification of Top Performing Alliances

In this section, we address the problem of classifying the top alliances based on their performance. First, we need to specify what the top performing alliances are. To do so, we choose a threshold α (e.g., 10%), and then for each week, find the $\tau = (1 - \alpha)\%$ quantile of the performance during that week (e.g., 90% quantile). We constructed a binary variable (target) that indicates whether an alliance is among top performing alliances. Each alliance having a performance greater than or equal to τ during that week is considered as top performing alliance, i.e., *target* = 1; otherwise, *target* = 0. Thus, the classification turns out into a binary classification task.

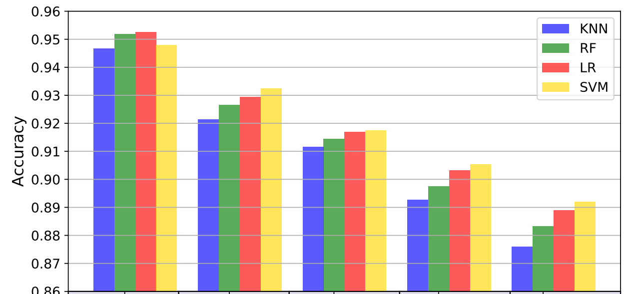
To address this binary classification task, we used four different classification approaches:

- **kNN**: k-Nearest Neighbors ($k = 45$).
- **RF**: Random Forest classifier (nr. estimators=100).
- **LR**: Logistic Regression.
- **SVM**: Support Vector Machine (*linear* kernel).

The features used for the classification are all 15 features in our final dataset (including N, E and *age*). Thus, this corresponds to the **network model** as mentioned earlier in the prediction section (no logarithm features are used). Moreover, all the features are transformed using min-max scaling, such that each feature falls within the [0,1] range.

The evaluation of the classification is tackled using the *accuracy* metric, which is the fraction of correctly classified instances to all instances (in the test set). In all the classification experiments, we used cross validation over 5 folds, where the reported accuracy is the average over the 5 fold classifications. The results of the classification, in terms of accuracy, for the five different thresholds (from top 5% to top 25%), and for the four classification approachs (kNN, RF, LR, and SVM), are shown in Figure 10.

First, we observe that, for any classification approach, the classification accuracy decreases as we increase the threshold of the top alliances. For instance, when we classify the top 5% alliances, the accuracy is about 95%, whereas when we classify the top 20% alliances, the accuracy is about 90%. Second, when we compare the different classification approaches, we observe that, in general, the best classifier is SVM, followed by Logistic Regression, followed by Random Forest, where kNN is the least accurate classifier.



	top 5%	top 10%	top 15%	top 20%	top 25%
KNN	0.9467	0.9213	0.9115	0.8926	0.8760
RF	0.9518	0.9266	0.9145	0.8974	0.8833
LR	0.9526	0.9293	0.9169	0.9031	0.8890
SVM	0.9480	0.9324	0.9175	0.9053	0.8920

Fig. 10: Classification Results

VIII. Conclusion

The goal of our study was to find out whether it is possible to predict alliance performance using SNA-features from *communication networks*. Further, we wanted to test the ability of *classification tasks* to identify the best performing alliances. In both cases, we have been able to show that it is possible to do this. Future research will help us to deepen our understanding of the underlying dynamics and enable us to apply our findings in a less specific context.

One major challenge we faced was the fact that we conducted our research within the environment of an online simulation game. We were able to track the interaction of 18,000 individuals, but we also had to learn that it is not an easy task to study these communication effects in isolation.

Despite the fact that the applied machine learning algorithms delivered excellent results in the classification tasks, we identified three effects that made it difficult to interpret the results outside the specific context: (1) the game's definition of performance, (2) the effect of group size on certain network attributes, and (3) the influence of past events.

Definition of team performance: We opted to define success in the same way the game does. By using a slight modification to the official alliance ranking, we were able to ensure that our definition of performance matched the players' incentives provided by the game design. With this clear advantage, we faced a challenging hurdle: the **ranking** is highly influenced by **group size (N)**. As described above, alliance leaders face a trade-off. One option is to add as many members as possible to the alliance. Having more members automatically leads to more inhabitants, which leads to a higher position in ranking. One the other hand, it is more difficult to coordinate a bigger group as opposed to a small team of highly experienced players. Evidence from the game shows that both strategies have been applied successfully for top performing teams. Nevertheless, there is a clear restriction. Figure 7 shows that alliances need to exceed a certain number of members (about $N > 35$) to be able to reach a top position in ranking. However, it is not sufficient to have many members in order to become a highly ranked team. We were able to show that the additional information extracted from the communication networks are able to make the difference. Applying these measures makes it possible to successfully forecast team performance.

Effect of group size on network attributes: One critical effect is that **N** is included in the formulas used to calculate certain **network attributes**. This leads to an unwanted dependency between these network attributes and N. Hence, the correlations of density, avg. in-degree, centralization and k-core show two different effects (the intended network effect *and* the indirect effect of N). Neither effect can be separated from the other.

The influence of past events: In our study, we used 13,412 alliance-week pairs. One assumption in our study is that all these data points are time independent of each other. It can be assumed that an alliance that was successful one week will also be successful the following week. The same logic applies to communication patterns. It will be important to address this issue in future studies.

Given these limitations, our future work will focus on eliminating these restrictions, which will make it possible to generalize our findings i.e., to better explain team dynamics in real world work teams.

One approach could be to (1) develop alternative measures for team performance that are either not or are less correlated with group size N. Further, it will be helpful to split the dataset to be able to (2) take group size into account (e.g., separate small and big teams). We also recommend (3) considering time by introducing dynamic analysis that can take into account past interactions and outcomes. Finally, we propose (4) refining the theoretic foundation. In this study, we have already implemented insights from team and leadership research. Additional theoretical models such as outside connectivity, core-periphery structure or the role of strong and weak ties can be expected to improve prediction results.

As we have shown in our paper, the opportunities for conducting research into online games are manifold. Researching online gaming is a very promising field, especially in view of the vast amount of data it can offer. We also demonstrated how important it is to oversee the effects coming from the special environment of these virtual worlds. The opportunities in this field are promising, and will be even more so once these very special frameworks and their limitations are better understood and mastered.

REFERENCES

[1] Manju Ahuja, Dennis Galletta, and Kathleen Carley. Individual centrality and performance in virtual r&d groups: An empirical study. *Management Science*, 49(1):21–38, 2003.

[2] Jakob Assmann, Marcus Drescher, Julia Gallenkamp, Arnold Picot, Isabell Welpe, and Rolf Wigand. MMOGs as emerging opportunities for research on virtual organizations and teams. *16th Americas Conference on Information Systems 2010*, 1:232–240, 2010.

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

[4] Curtis Bonk and Vanessa Dennen. Massive multiplayer online gaming: A research framework for military training and education. Technical report, Indiana University, Bloomington, 2005.

[5] Edward Castronova. *Synthetic worlds: The business and culture of online games*. University of Chicago press, 2008.

[6] Elaine Chan and Peter Vorderer. Massively multiplayer online games. In Peter Vorderer and JE Bryant, editors, *Playing video games: Motives, responses, and consequences*, pages 77–88. Lawrence Erlbaum Associates Publishers, 2006.

[7] Noshir Contractor, Leslie DeChurch, Jay Carson, Dorothy Carter, and Brian Keegan. The topology of collective leadership. *Leadership Quarterly*, 23(6):994–1011, 2012.

[8] Robert Cross and Andrew Parker. *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business Press, Boston, 2004.

[9] Kristin Cullen-Lester, Cynthia Maupin, and Dorothy Carter. Incorporating social networks into leadership development: A conceptual model and evaluation of research and practice. *Leadership Quarterly*, 28(1):130–152, 2017.

[10] Sara De Freitas and Mark Griffiths. Online gaming as an educational tool in learning and training. *British Journal of Educational Technology*, 38(3):535–537, 2007.

[11] Lauren D'Innocenzo, John Mathieu, and Michael Kukenberger. A Meta-Analysis of Different Forms of Shared Leadership–Team Performance Relations. *Journal of Management*, 42(7):1964–1991, 2016.

[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, 2014.

[13] Linton Freeman. Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3):215–239, 1978.

[14] Lucy Gilson, Travis Maynard, Nicole Jones Young, Matti Vartiainen, and Marko Hakonen. Virtual Teams Research: 10 Years, 10 Themes, and 10 Opportunities. *Journal of Management*, 41(5):1313–1337, 2015.

[15] Peter Gloor, Maria Paasivaara, Detlef Schoder, and Paul Willems. Finding collaborative innovation networks through correlating performance with social network structure. *International Journal of Production Research*, 46(5):1357–1371, 2008.

[16] Yun Huang, Mengxiao Zhu, Jing Wang, Nishith Pathak, Cuihua Shen, Brian Keegan, Dmitri Williams, and Noshir Contractor. The formation of task-oriented groups: Exploring combat activities in online games. In *2009 International Conference on Computational Science and Engineering*, volume 4, pages 122–127. IEEE, 2009.

[17] Markku Jokisaari. A Social Network Approach to Examining Leadership. In *Handbook of Methods in Leadership Research*, chapter 11, pages 1–56. Edward Elgar Publishing, 2016.

[18] Audrey Korsgaard, Arnold Picot, Rolf Wigand, Isabelle Welpe, and Jakob Assmann. Cooperation, coordination, and trust in virtual teams: Insights from virtual games. In *Online worlds: Convergence of the real and the virtual*, pages 253–264. Springer, 2010.

[19] David Krackhardt and Jeffrey Hanson. Informal networks: The company behind the chart. In Rob Cross, Andrew Parker, and Lisa Sasson, editors, *Networks in the Knowledge Economy*, pages 235–247. Oxford University Press, New York, 2003.

[20] Paul Leonardi and Noshir Contractor. Better people analytics. *Harvard Business Review*, November-December:1–13, 2018.

[21] Jessica Lipnack and Jeffrey Stamps. Virtual teams: The new way to work. *Strategy & Leadership*, 27(1):14–19, 1999.

[22] Jacob Moreno. The sociometric view of the community. *The Journal of Educational Sociology*, 19(9):540–545, 1946.

[23] Vias Nicolaides, Kate LaPort, Tiffani Chen, Alan Tomassetti, Eric. Weis, Stephen Zaccaro, and Jose Cortina. The shared leadership of teams: A meta-analysis of proximal, distal, and moderating relationships. *Leadership Quarterly*, 25(5):923–942, 2014.

[24] Celia Pearce. *Communities of play: Emergent cultures in multiplayer games and virtual worlds*. MIT press, 2011.

[25] Nataliia Pobiedina, Julia Neidhardt, Calatrava Moreno, Maria del Carmen, and Hannes Werthner. Ranking factors of team success. In *International World Wide Web Conference*, pages 1185–1194, Rio de Janeiro, Brazil, 2013. ACM.

[26] Byron Reeves, Thomas Malone, and Tony O'Driscoll. Leadership's online labs. *Harvard Business Review*, 86(5):58–66, 2008.

[27] Stephen Seidman. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983.

[28] 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.

[29] Kyong Jin Shim and Jaideep Srivastava. Team performance prediction in massively multiplayer online role-playing games (mmorpgs). In *2010 IEEE Second International Conference on Social Computing*, pages 128–136. IEEE, 2010.

[30] Theodoros Sourmelis, Andri Ioannou, and Panayiotis Zaphiris. Massively multiplayer online role playing games (mmorpgs) and the 21st century skills: A comprehensive research review from 2010 to 2016. *Computers in Human Behavior*, 67:41–48, 2017.

[31] Mirko Suznjevic, Ivana Stupar, and Maja Matijasevic. Mmorpg player behavior model based on player action categories. In *2011 10th Annual Workshop on Network and Systems Support for Games*, pages 1–6. IEEE, 2011.

[32] Henri Theil. *Economic Forecasts and Policy*. Contributions to economic analysis. North-Holland Pub. Co., 1961.

[33] Bruce Tuckman and Mary Ann Jensen. Stages of small-group development revisited. *Group & Organization Studies*, 2(4):419–427, 1977.

[34] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, MA, 1994.

[35] Rolf Wigand, Nitin Agarwal, Isaac Osesina, Winfried Hering, Audrey Korsgaard, Arnold Picot, and Marcus Drescher. Social network indices as performance predictors in a virtual organization. In *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)*, pages 144–149, Sao Carlos, Brazil, 2012. IEEE.

[36] 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.