Article:
schiller, michael, Wallner, Guenter, Schinnerl, Christopher et al. (Accepted: 2018) Inside the Group: Investigating Social Structures in Player Groups and Their Influence on Activity. IEEE Transactions on Games. ISSN: 2475-1510 (In Press)

Inside the Group: Investigating Social Structures in Player Groups and Their Influence on Activity

Michael Schiller, Günter Wallner, Christopher Schinnerl, Alexander Monte Calvo, Johanna Pirker, Rafet Sifa and Anders Drachen

***Abstract*—Social features, matchmaking, and grouping functions are key elements of online multi-player experiences. Understanding how social connections form in and around games and their relationship to in-game activity offers insights for building and maintaining player bases and for improving engagement and retention. This paper presents an analysis of the groups formed by users of the `the100.io` – a social matchmaking website for different commercial titles, including *Destiny* on which we focus in this paper. Groups formed on `the100.io` can be described across a range of social network related metrics. Also, the social network formed within a group is evaluated in combination with user-provided demographic and preference data. Archetypal analysis is used to classify groups into archetypes and a correlation analysis is presented covering the effect of group characteristics on in-game-activity. Finally, weekly activity profiles are described. Our results indicate that group size as well as the number of moderators within a group and their connectedness to other team members influences a group's activity. We also identified four prototypical types of groups with different characteristics concerning composition, social cohesion, and activity.**

***Index Terms*—Social Networks, Matchmaking, Game Analytics, Destiny.**

I. Introduction

Social relationships formed within and through online multi-player games influence the engagement and user experience of players [1], [2]. Moreover, social relationships in games are essential drivers of retention and monetization in games [3], [4]. The facilitation and management of player communities and the connections between players is an important part of maintaining a healthy player base for a game and is vital for the survival of online multi-player games, which rely on a persistent presence [3], [4], [5], [6]. Building an understanding of how social connections are formed across such platforms – whether they are provided as part of a game or game distribution network or have grown around a game – and how connections can foster engagement, retention, or promote particular behaviors (e.g., to reduce toxicity among community members [7]), can thus offer actionable insights for companies to achieve this goal.

The importance of social connections in games means that massively multiplayer online (MMO) games – irrespective of hardware platform – routinely provide dedicated matchmaking or group-generation features in order to make it as easy as possible for players to find similarly skilled teammates, solve group quests, participate in raids, find opponents, a clan, or guild to join, etc. However, while many games such as *World of Warcraft* or *Starcraft* support the in-game formation of friendships, guilds, or groups [8], not all games (including *Destiny*) include features to build in-game communities. This has created an opportunity for external solutions such as online player-grouping websites and matchmaking services of various kinds. They actively seek to assist players in finding like-minded people to play with and thus in building and maintaining long-term social relationships in and around games.

Social networks of grouping or matchmaking features, third-party services, or similar can be analyzed through the adoption of social network analysis (SNA) techniques combined with machine learning of contextual data such as demographic, self-report, and behavioral telemetry (e.g., [4], [6], [9], [10], [11], [12], [13]). SNA can hence be used as a foundation for investigating player interactions and relationships. In practice, however, SNA in games is an underexplored topic across network analysis and games user research [11]. Furthermore, the combination of social network data and contextual data is even rarer, Rattinger et al. [4] forming a notable exception. There is, thus, a general gap in existing work regarding the knowledge about how network behavior in games relates to the behaviors of a player or the group the player is part of, the psychological aspects of the player (e.g., motivation, preference, personality), or the in-game behavior of the player [6]. In addition, work so far has focused on groups formed within a game itself (e.g., [14], [15], [16], [17]) and not on groups formed on external looking-for-group facilities.

In this paper, the focus is on taking a step toward addressing the current situation by combining the social network with self-report information from the social matchmaking service `the100.io` across tens of thousands of players of the game *Destiny* [18] – a hybrid online first-person shooter and multi-player/massively multi-player game. The work presented here extends previous efforts by not only considering a player-established community but also by integrating demographic and preference data.

We present a series of analyses targeting the problem of characterizing player groups and developing metrics to describe them, and investigate correlations between group characteristics and their activity level in *Destiny*. Specifically, we present a correlation analysis aiming at identifying the effect of group characteristics on group activity. Results show that the number of moderators, their connectedness, and the group size correlate with group activity. Categories of player groups developed via archetypal analysis [19], [20], [21]

G. Wallner is with the University of Applied Arts Vienna

M. Schiller, C. Schinnerl, and J. Pirker are with the Graz University of Technology

A. Monte Calvo is with Bungie, Inc.

R. Sifa is with Fraunhofer IAIS

A. Drachen is with the University of York

across a series of group features show the presence of four types of player groups with varying degrees of social cohesion, moderator activity, activity levels, character level, etc. Group activity is also presented as a function of weekdays to investigate when `the100.io` players schedule activities, across groups comprised of casual or serious players.

The metrics used to generate these results are based on factors and behaviors that are common across a wide range of online games and can thus likely be transferred to social behavior analysis in other titles as well. The archetypes presented provide a means for distinguishing different types of groups in games communities and thus give community managers and game designers concise information to act on to facilitate their needs.

II. Related Work

Analyzing and understanding social interactions and connections between players in online multiplayer games is crucial for obtaining a deeper understanding of in-game behavior, player experience, and player retention [4]. Thus, it is important to understand how these games function as entertainment communities and social platforms and how groups within and outside games are formed and structured.

In the following, we discuss related work in the fields of: (1) groups and communities in games, (2) social networks analysis in games, and (3) behavioral profiles and archetypes in games.

A. Groups and Communities in Games

Collaboration and competition have always been crucial elements of gaming and playing. Players always tended to form interest groups, with play communities existing long before modern multi-player online games [22]. Identification and analysis of such groups, social aspects, communicative strategies, and different interaction forms are relevant strategies to improve game design and to gain insights into social and communicative behaviors. Thus, understanding social behavior, groups, and communities in large-scale and popular multiplayer online titles is an essential step toward an improved understanding of player behavior. For example, Manninen [23] investigates interaction forms and communicative actions in multiplayer games and illustrates a social theory framework of interaction forms as a tool for designing and analyzing games.

Ducheneaut et al. [24] investigate and discuss social dynamics and social experiences in the large-scale gaming community of *World of Warcraft* and show that in-world grouping (e.g., through joint quests) is less important socially compared to player associations such as guilds. Guilds, player groups, and player communities, however, have a significant impact on player patterns. Ducheneaut et al. [15] explored structural properties of guilds which may contribute to the success or failure of the guild. The social network is approximated by relying on the locations of characters in the game world. Thurau and Bauckhage [25] performed a categorization of different guilds of players in *World of Warcraft* using matrix factorization in order to analyze the development of guilds over time. Poor [17], also focusing on *World of Warcraft*, studied the relationship between guild membership and character leveling, finding that guild membership does not significantly support leveling. Mason and Clauset [14] combined data on ad-hoc teams formed in *Halo: Reach* with survey data to investigate the influence of friendships on collaborative and competitive performance. In comparison to our work, players had to select their friends from a list compiled based on their game history while in our case this information was directly accessible. Goh and Wasko [26] used a mixed-methods approach, including affiliation networks, to identify characteristics of potential guild leaders. Chen et al. [16] looked into guild dynamics, focusing on guild-joining behavior, guild participation, and movement between guilds. Contrary to all these works which concentrate on in-game groups we are focusing on an external service aimed at facilitating play in the first place.

Unfortunately, identifying and analyzing meaningful in-game groups and communities often poses a challenge as the social network cannot be readily deduced as explicit information about connections is not available or accessible. However, implicit social connects as formed, for instance, through player matches have shown to be an important aspect for player engagement and player performance [4] and can be used to recommend teams and match-partners [27].

B. Social Network Analysis

Social Network Analysis (SNA) has been shown to be a valuable method to analyze social communities formed within traditional organizations [28] or in modern online platforms such as *Facebook* or *Twitter* [29]. It has become a significant tool in fields such as sociology, information science, political science, economics, or organizational studies (e.g., [30], [31]). However, its application for investigating gaming communities is comparatively new. As a consequence there are still relatively few studies that use SNA to analyze player behavior and structures. However, existing work so far has shown the potential of this graph-based approach for investigating social structures, match-partner recommender systems [27], and for identifying potential cheaters [32]. While most authors explored networks formed through friendships or groups, only a few looked at indirect connections, for example, formed through in-game behavior (e.g., [33]). Moreover, the state-of-the-art focuses on typical social network metrics to investigate social gameplay and does not include behavioral features or preference data. Recently, however, Rattinger et al. [4] explored social networks formed through matches in the hybrid shooter *Destiny* and combined it with behavioral profiles. The authors show correlations between such implicit social structures and in-game behavior, engagement, and performance.

C. Behavioral Profiling in Games

The availability of large-scale game behavioral data has led to a tremendous amount of attention to behavioral analytics in game development and research. The analysis of player behavior has rapidly emerged to become an integrated component of game development [6], [34], [35]. One critical challenge in game analytics is pattern finding and the development of actionable models of behavior based on such patterns and any contextual data. Behavioral profiling provides an opportunity for condensing highly varied and high-frequency user telemetry into condensed, actionable profiles. These can be used to inform design, assist matchmaking, build user prediction models, track problems, etc., similar to the application of profiling in areas such as web analytics [21], [36].

While a complete review of the previous work in behavioral profiling in games is out of scope of the current paper it is important to note that the application of behavioral profiling to digital games is relatively new, arising with the introduction of large-scale user behavior data through hosting of games on social media platforms and with the introduction of mobile platforms [3]. One of the first publications addressing the problem of developing actionable behavioral profiles from behavioral telemetry in games was Drachen et al. [37] who worked with self-organizing networks to develop profiles characterizing player behavior in the major commercial title *Tomb Raider: Underworld*. Since then a substantial amount of research on the topic has been released, including Thawonmas and Iizuka [38] who used multi-dimensional scaling to characterize behavior in the game *Shen Zhou Online*. Evaluating the fitness of simplex volume maximization and k-means on behavioral data from *Tera: Online* and *Battlefield 2*, Drachen et al. [20] noted the different strengths and weaknesses of centroid-seeking vs. convex hull-seeking clustering models. Normoyle and Jensen [39] introduced Bayesian Clustering to behavioral profiling in games, drawing on data from *Battlefield 3*. Bauckhage et al. [40] introduced spatiotemporal clustering and developed waypoint graphs that permitted behavioral-based partitioning of game maps. Drachen et al. [41] developed behavioral profiles for *Destiny*, comparing four different cluster models. In general, cluster analysis has become the primary machine learning tool used for profiling purposes. As a flexible unsupervised learning method, clustering is useful for pattern exploration and permits condensation of multi-variate space [21]. Reviews of clustering models and their application in digital games are provided by Bauckhage et al. [21] and Drachen et al. [41]. Archetypal analysis (AA) [19], [42] is repeatedly mentioned in this literature as a scalable model for developing plainly isolated and logical profiles in games and is therefore adopted here. An introduction to AA is provided in Section V.

III. Destiny and the100.io

Destiny [18] is an online multiplayer shooter set in a science fiction-themed world where players take on the role of *Guardians* to defend the Earth against alien aggressors to save mankind from extinction. Players can play as one of three character classes which can be leveled up to unlock new abilities and become more powerful. The game offers a wealth of weapons, armor, and other equipment with most of these being modifiable as well. The game blends shooter mechanics with elements of role-playing games. The gameplay mainly revolves around individual and small team combat. Toward this end, *Destiny* offers various player vs. player and player vs. environment game modes. Multi-player is often performed by assembling players into fireteams which work together to achieve a common objective or take on against each other.

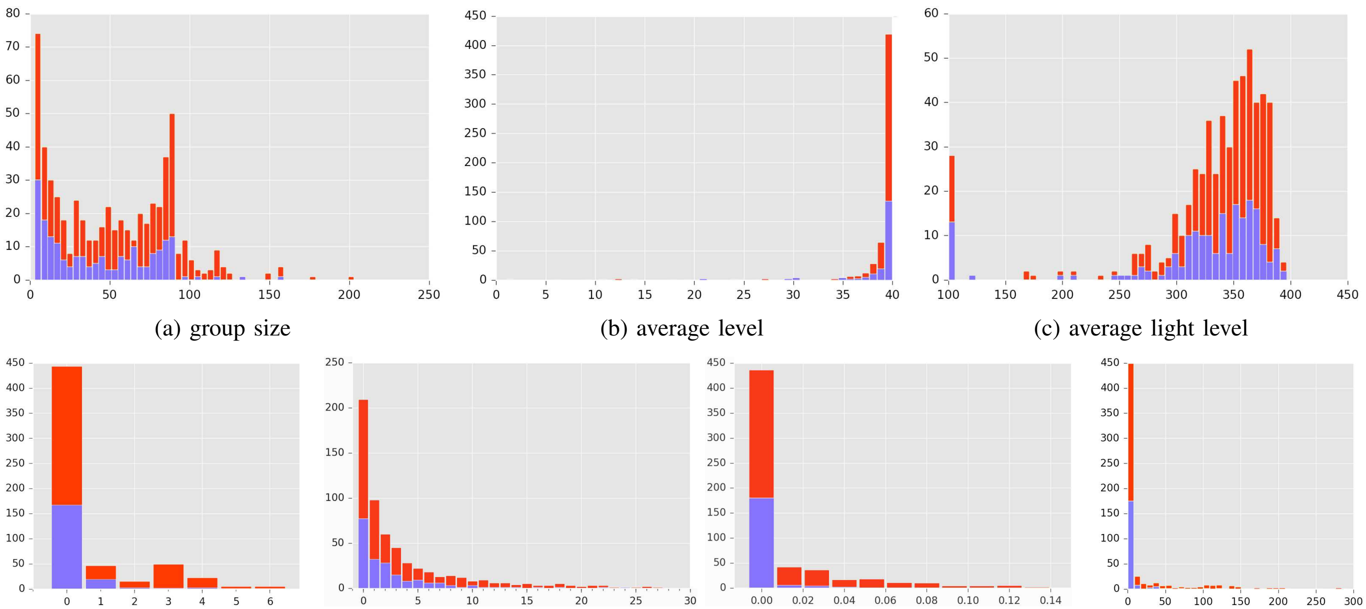
However, *Destiny* itself does not provide any in-game matchmaking facilities for most activities such as raids to help players to connect with each other. In lack thereof, so-called *Looking for Group* (LFG) websites emerged which assist players in finding team mates. `the100.io` is a group matchmaking service that helps players to find a permanent group of like-minded people while other LFG websites focus on temporary groups for instant matches. Users of the `the100.io` need to create a profile providing different information such as preferred platform and preferred time of the day for playing, time zone, character level, and light level (see Section IV). Based on the entered preferences the `the100.io` automatically assigns the player to a group of similar players. However, players can also join other groups apart from the one they get assigned to. Groups also have different properties such as play style, platform, typical time of day for playing, and the number of members. Furthermore, as *Destiny* does not support cross-platform play, groups are specific to a certain platform. Also, each group can have moderators and sherpas. The latter are players who act as guides for inexperienced players. Besides that, the website allows players to add friends and to schedule and sign-up for *Destiny* related activities. For instance, a user can schedule a game for 9 PM CET and allow other members to sign up for it.

IV. Data collection and preprocessing

Information about users, groups, and games are listed within pages on the `the100.io` and was collected through a Python script as of December 16, 2016. The collected data set contains information about 218,214 players registered on `the100.io` that scheduled a total of 637,823 unique games and form 2,468 groups. Since `the100.io` allows for scheduling games for different video games, groups that did not report playing *Destiny*, games scheduled for games other than *Destiny*[1], and games that had no group information attached were removed. Groups composed of fewer than three players and groups with missing activity score information were excluded as well. Furthermore, user data was checked for invalid and missing values in the self-reported variables such as character level and light level and these users were not taken into consideration for further analyses.

After cleaning the data 586 groups remained of which 196 groups were designated as serious and the remaining 390 as casual groups (see below). In terms of platform, 252 groups are dedicated to PS4, 42 to PS3, 216 groups are playing on Xbox One, 42 on Xbox 360, and the remaining 34 are PC groups. Visual inspection of the variables of interest did not indicate any remarkable differences among the different platforms for which reason we did not distinguish among platforms for this first investigation. In total 26,317 players distribute across these groups, having played a total of 1,493,599 games at the time of data collection. While `the100.io` requires that

[1] We have focused on a single game in this study as we also included game-specific performance measures which are difficult to compare across different games.



(a) group size (b) average level (c) average light level

(d) number of moderators (e) number of sherpas (f) density (g) activity score

Fig. 1: Histograms of distributions of group related characteristics according to serious (blue) and casual groups (red).

friendships need to be confirmed by both parties before being considered friends, we also considered friendships if only confirmed by one party as at least one user expressed interest in the connection.

For this paper the variables of interest for groups are:

- **play style** either casual or serious. Serious groups are groups which are intended for players with a serious and competitive play style. Casual groups are for people who play on a more leisurely basis. Here, it is important to note that the coding was not performed by ourselves. Rather the distinction is made by the `the100.io` itself and users get initially assigned to either a serious or casual group based on their self-reported play style.
- **group size** (N_g), i.e., the number of members of a group
- **number of moderators**
- **number of sherpas**
- **density**, as a measure of interconnectedness of the group members, calculated as the number of actual friendships divided by the number of potential connections, that is, $N_g \cdot (N_g - 1)/2$
- **global clustering coefficient** (**C**) as defined by Newman [43] as a measure of the overall clustering of a group (given by 3 $\times$ the number of triangles in a network divided by the number of connected triplets of nodes)
- **average degree centrality** (**$\overline{dc}$**) **of sherpas**: Besides the number of sherpas the connectivity of sherpas in the group might play a role for activity as well. As such we calculated the degree centrality (i.e,. number of friendships / $(N_g - 1)$) for each sherpa and averaged it over all sherpas of the group.
- **average degree centrality** (**$\overline{dc}$**) **of moderators**: As above, but for moderators.
- **activity score**, `the100.io` assigns a score to each group as an indicator of how active the group is. It equates to the number of confirmed sessions of a group (each time a member of the group joins a gaming session) over the past week.[2]
- **active games**, that is, the number of recent and upcoming games as listed on a group's profile page as a snapshot of activity. Hence, it reflects the current activity level at the time of sampling while the activity score indicates confirmed activity over a week.

In addition to these group-level variables, we derived information about the groups based on the individual members to derive measures of the group members' experience, in particular:

- **average level** ($\overline{\text{level}}$): the maximum level of a character of a player[3], averaged over all group members
- **average light level** ($\overline{\text{light level}}$): the light level is a rating of a character's equipped gear, for example, weapons. A higher light level corresponds to better equipment and results in better offensive and defensive abilities.

Please note that both character level and light level are user-reported variables. However, as `the100.io` assigns players based on their provided data we assume that players mostly report their actual data as otherwise, they may end up in groups not fitting their play style or experience.

V. Results

A. Basic Data Description

Figure 1 shows histograms of the distribution of the basic group-related properties. First of all, group size (Figure 1a) varies mostly between 0 and 100 members with peaks both near 0 and near 100. The sudden drop after that can be explained by the fact that the `the100.io` forms groups of 100

[2] https://www.reddit.com/r/the100website/comments/3d28u4/what_does_my_groups_activity_score_mean/ Accessed: February 2018

[3] In *Destiny* a player can create multiple characters. We used the maximum level as a reasonable indicator of a player's in-game experience.

TABLE I: Spearman rank correlations between different group-related characteristics. Correlations with $|\rho| > .5$ are written in bold face.

	activity score	active games	group size	no. of modera-tors	no. of sherpas	density	C	$\overline{\text{dc}}$ of modera-tors	$\overline{\text{dc}}$ of sherpas	$\overline{\text{level}}$	$\overline{\text{light level}}$
activity score	1										
active games	**.958***	1									
group size	**.687***	**.676***	1								
no. of moderators	**.748***	**.746***	**.606***	1							
no. of sherpas	**.695***	**.698***	**.751***	**.676***	1						
density	**.610***	**.613***	**.591***	**.683***	**.682***	1					
C	.405*	.424*	.309*	.465*	.425*	**.589***	1				
$\overline{\text{dc}}$ of moderators	**.721***	**.723***	**.570***	**.927***	**.648***	**.715***	**.522***	1			
$\overline{\text{dc}}$ of sherpas	**.672***	**.679***	**.578***	**.763***	**.724***	**.798***	**.573***	**.789***	1		
$\overline{\text{level}}$	-.085	-.084	-.234*	-.065	-.150*	-.073	-.091	-.080	-.059	1	
$\overline{\text{light level}}$	**.587***	**.580***	.462*	**.556***	**.533***	.440*	.344*	**.531***	.491*	.174*	1

dc = degree centrality, C = clustering coefficient, averaged values denoted by overlines, $^*p < .00091$ (Bonferroni adjusted) -1 0 1

TABLE II: Multiple linear regression of group characteristics on group activity.

Predictor	**Estimate**	**Std. Error**	β	**t-value**
no. of moderators*	11.10	1.47	0.31	7.57
group size*	0.16	0.04	0.12	3.68
no. of sherpas	0.57	0.38	0.06	1.50
$\overline{\text{dc}}$ of sherpas	-164008.24	100514.08	-0.07	-1.63
$\overline{\text{dc}}$ of moderators*	534756.40	51001.41	0.50	10.49
density	33.51	34.37	0.03	0.98

$^*p < .001$, adjusted R^2 = 0.7267

players. However, due to people getting invited to groups, groups may also get larger. Players in our dataset are all in all very advanced concerning their character level (Figure 1b) with most players having a level of or near 40 (the maximum possible character level). Light level (Figure 1c) although a little bit more dispersed is also quite high with most players having a light level between 300 and the maximum of 400. In terms of the number of moderators (Figure 1d) the majority of groups have none, similarly, most groups also do not have any sherpas (Figure 1e). Overall, however, groups have more sherpas than moderators. Most groups are also not at all or only loosely connected as reflected by the very low density values for most groups (Figure 1f). Lastly, it is also noticeable that a large portion of the groups only has very low activity scores. While groups with activity scores of up to 200 are still quite common, groups with activity scores larger than that are rare.

B. Correlations

Table I shows the results of a Spearman rank correlation (chosen because of non-normally distributed variables, see, e.g., [44]), relating the variables outlined in Section IV except play style (due to being a dichotomous variable). Please note that some individual correlations are based on a slightly smaller number of groups (573) as some groups were excluded because of missing data for the respective correlations and that the global clustering coefficient has only been calculated for groups with at least one triad (a group of three connected users), that is, 270 groups as otherwise the coefficient would be undefined. To account for multiple comparisons, a Bonferroni corrected (cf. [44]) α-level of .00091 was used to determine statistical significance. In the following discussion, we will restrict ourselves mainly to correlations with $|\rho| > .5$ in relation to the activity related measures.

First, we should note that the number of moderators and sherpas is highly correlated with group size. Since density is measured relative to the network size, it is also worth noting that density also increases with group size, i.e., players in larger groups establish relatively more friendships. The average level of the players in a group, however, did not result in any noteworthy correlations which, very likely, is a direct consequence of the level cumulating at the maximum level of 40. However, the light level did show an influence but also has been more widely distributed. Both, activity score and the number of active games show similar correlations with the other metrics and as such we will not distinguish between them in the remainder of this section. Concerning group composition and connectedness, group size, number of moderators and sherpas, average connectedness of moderator and sherpas, density and to a smaller extent the clustering coefficient all are positively correlated with activity. As such we also conducted a multiple linear regression to better understand the influence of the individual factors and to develop a model for predicting group activity from the number of moderators, the number of group members, the number of sherpas, the connectivity of sherpas and moderators, as well as density. Basic regression coefficients are shown in Table II. Three of the six predictor variables have a significant ($p < .001$) zero-order correlation with group activity, namely group size, number of moderators, and the average degree centrality of moderators. The three predictor model was able to account for 72.67% of the variance in group activity, $F(4, 574) = 254.5$, $p < .001$, with an adjusted R-squared of 0.7267.

C. Weekly Activity

As pointed out above we collected data on 1,493,599 games. These games were scheduled from January 1, 2011 to December 31, 2017 (games can be scheduled in advance) with the large majority of them taking place during 2015 (883,695)

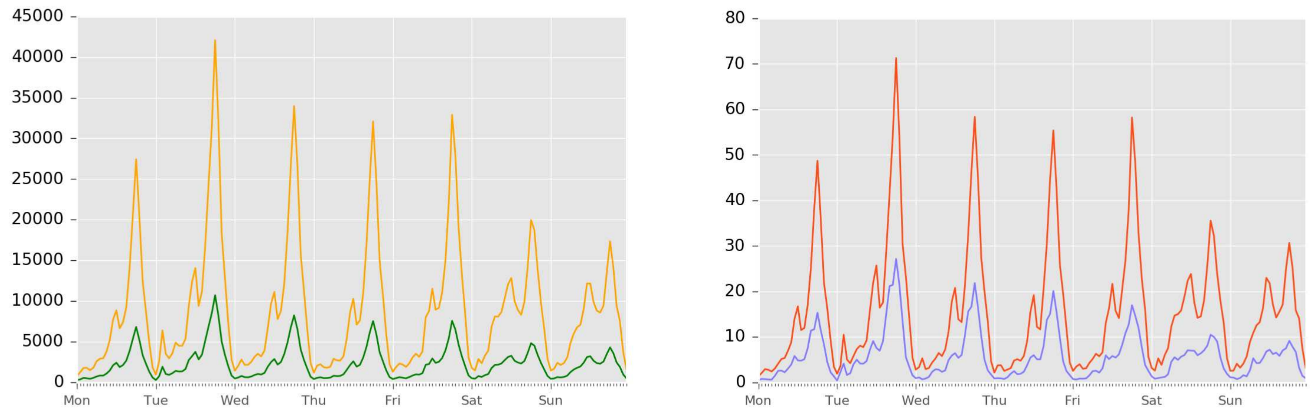


Fig. 2: Activity on weekdays and time of day. Left: Number of scheduled games (green) and number of players (orange) signing up for these games. Right: Average number of scheduled games across casual (red) and serious groups (purple).

and 2016 (587,247). Figure 2 (left) shows the number of games scheduled and the number of players signing up for these games on weekdays and time of day. Surprisingly, and contrary to what we would have expected, Saturday and Sunday have the lowest number of games while activity peaks on Tuesdays. Activity across the other four days is roughly constant with activity considerably increasing toward the evening of each day. These patterns are also evident if we split the scheduled games by serious and casual groups (see Figure 2, right), i.e., the weekly behavior is consistent for casual and serious groups. The low activity on weekends may indicate that players have less use for the `the100.io` during weekends, possibly because their regular playing groups have no trouble to coordinate on weekends. However, this is currently only speculative and further research would be necessary to verify this assumption.

D. Archetypal Analysis

While commonly employed to detect patterns in behavioral analyses in games, interpretations of clusters from the perspective of applicability of the developed profiles can be difficult [20], [37], [45]. This is notably the case for the perhaps most widely adopted unsupervised clustering algorithm, k-means, which is theoretically suited for behavioral analytics. However, as it is focused on retrieving compact cluster regions, results can be hard to interpret in practice, as discussed by Bauckhage et al. [21], [46].

The soft clustering based analysis in this work is performed by utilizing archetypal analysis (AA). AA was introduced by Cutler and Breiman [19], and more recently extended to be applicable to large-scale datasets [21], [46], [47]. Formally, as a constrained two matrix factorization technique, AA allows us to arrive at compact and interpretable data representations via extreme representative points that are called archetypes and the stochastic coefficients that indicate belongingness ratios to the corresponding archetypes. Formally, considering a column data matrix $\mathbf{X} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$ defined $\mathbf{X} = [\tilde{\mathbf{x}}_\mathbf{1}, \tilde{\mathbf{x}}_\mathbf{2}, \dots, \tilde{\mathbf{x}}_\mathbf{n}]$, archetypal analysis deals with finding $\mathbf{X} \approx \mathbf{ZA}$ where the two column matrices $\mathbf{Z} \in \mathbb{R}^{\mathbf{m}\times\mathbf{k}}$ and $\mathbf{A} \in \mathbb{R}^{\mathbf{k}\times\mathbf{n}}$ represent the archetypal matrix and the column stochastic coefficient matrix. Each column of $\mathbf{Z}$ is an archetype living in the data convex hull, whereas, each column of $\mathbf{A}$ lives in a $(k-1)$ simplex and is used to represent each data point as a convex mixture of the columns of $\mathbf{Z}$. It is important to note that since every data point $\mathbf{x_i}$ in $\mathbf{X}$ has a corresponding vector $\vec{a}_i$ with lower dimensionality (i.e., $k \ll m$), AA also allows for dimensionality reduction.

The parameter selection is done by applying an alternating least squares procedure where each iteration requires the solution of several constrained quadratic optimization problems. AA has become attractive to behavioral analytics in games because it permits detection of *special* player behaviors, such as elite players, people adopting cheats, or players who struggle to progress in the game, as it is focused on finding extremes in the dataset [20], [46], [48]. Specifically, what AA does is that it automatically detects a combination of features that leads, when being locked in pairs, to a similar but more complex segmentation as k-means without requiring any user intervention (e.g., in determining the value of k). Where k-means produces cluster centroids, AA is different in that it is not looking for commonalities between players, but rather archetypal (extreme) profiles that do not reside in dense cluster regions but at the edges of the multidimensional space. In game analytics, archetypal data representations were previously used for profiling player behavior [20], [45], [49], analyzing population interest [50], building game recommender systems [48], and analyzing behavioral structures in social multiplayer online games [4]. For more detailed applications of AA for behavioral profiling we refer to [20], [21], [45].

In our case, we used AA to find prototypical groups of the `the100.io`. To facilitate interpretability of the resulting clusters we kept the number of included features low while ensuring to include group characteristics (group size, density, number of moderators and sherpas), activity related measures (activity score), as well as factors reflecting the experience of groups (average level and average light level). Before running the AA, we excluded groups that contained invalid group features – such as invalid average level and average light level. This yielded a total number of 573 groups remaining for the archetypal analysis.

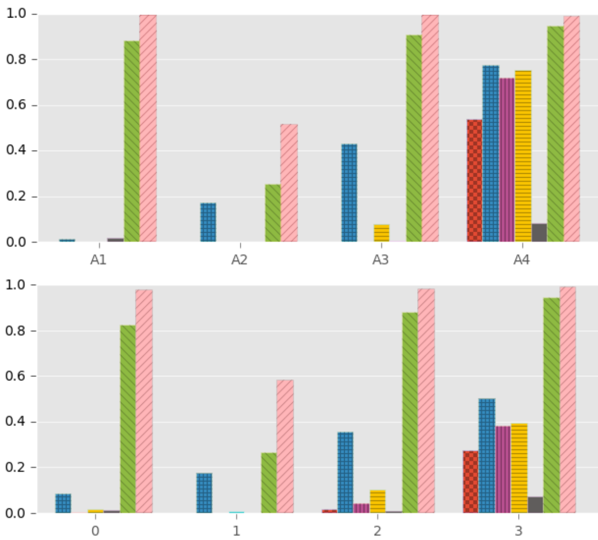


Fig. 3: Profiles of the four clusters (top: AA, bottom: k-means, = activity score, = group size, = no. of moderators, = no. of sherpas, = density, = avg. light level, = avg. level).

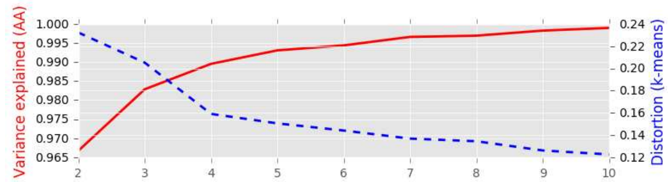


Fig. 4: Scree plot for AA (red) and k-means (blue, dashed) clustering for 2 to 10 clusters.

We then run the AA for two to ten clusters ($k = 2 - 10$) and after inspection of the percentage of variance using the elbow method (following [20], [21], [51]) and the clusters for interpretability (following [20], [37]) we ended up with a four cluster solution of which the profiles are shown in Figure 3. While the scree plot (Figure 4) showed no major elbow, higher number of clusters such as five or six clusters mainly resulted in $A2$ being split into smaller fragments. Please note that due to AA being a soft clustering approach a group belongs to the different archetypes ($A1$-$A4$) with varying degrees [21]. $A4$ are highly active, large, and densely connected groups with a fairly large number of moderators and sherpas while at the other end of the spectrum groups in $A1$ can be characterized by being small and inactive (and members may thus be in danger of leaving again). Between these two extremes, $A2$ covers groups which are already larger than the ones in $A1$ but have lower experience scores (level and light level) while groups associated with $A3$ are fairly large, have an increasing number of sherpas and also show an increase in the number of friendships.

To illustrate the structural characteristics of these groups, Figure 5 shows prototypical groups for each of the four archetypes (belongingness coefficient of each group $> .98$). Each sector of the chord diagram represents one group member with the sector being colored according to the member's role in the group. Edges connecting members indicate friendships. The inner circle is color-coded to reflect the activity score of the group. As can be seen from Figure 5a, the prototypical group belonging to $A1$ has just four members of which none is friends with each other. Probably as a result of the very small group size the group is also not very active. These are probably groups which have been newly formed on the `the100.io` and thus are still waiting for more members. The group serving as an example for $A2$ (cf. Figure 5b) has already more members including one serving as sherpa but members have not established connections so far with one exception. This starts to change with $A3$ (Figure 5c). In the chosen representative group, members already have more connections, the group size itself has approximately doubled compared to the prototypical group of $A2$, and some members have already taken the role of moderators and sherpas. The last group (Figure 5d) belonging to $A4$ again roughly doubled in group size with members being much more connected. The group in question also has a considerable number of moderators and sherpas, and some members even take the role of both.

As noted above, AA provides the option for both soft and hard clustering. Each has distinct advantages and disadvantages. In the current analysis, soft clustering was used, i.e. a group does not belong exclusively to one of these four archetypes but can be expressed as a combination of them, which provides the ability to evaluate cluster affiliations in a more nuanced fashion than hard clustering [50]. Hard clustering does not provide affiliation information across clusters, but has the advantage of providing clearer output. Table **??** shows the result of a hard clustering of the groups based on the highest membership value together with descriptive statistics for each cluster. In order to accentuate the AA-developed profiles, k-means clustering was also applied to the group dataset. K-means is a centroid-seeking cluster model – covered in detail in [51] – and thus works differently than the convex-hull seeking AA [20], [19], [36]. As for AA, k-means was run for k=2 to k=10 clusters. Similar as for AA, elbow plot indicates a k=4 solution (see Figure 4) with a, however, much more distinct elbow. Despite the two models having different search parameters, the resulting profiles are quite similar to the AA profiles and of similar size (n=240,31,217,85, similar ordering as in Table **??**, see also Figure 3) adding support to these. As the k-means results support the AA results, we are not covering them in greater detail here.

VI. Discussion

First, it is noticeable that the `the100.io` attracts high level players with high light levels, across casual and serious groups. Even players considering themselves as casual can thus be viewed as engaged and dedicated players. Most of the groups, however, are not very active and are not very connected as reflected by the overall low density values. However, the results of this study indicate that activity increases with group size. While this may not seem very surprising it may warrant further discussion in light of the work of anthropologists such as Dunbar [52] who stipulated that there is an upper limit of

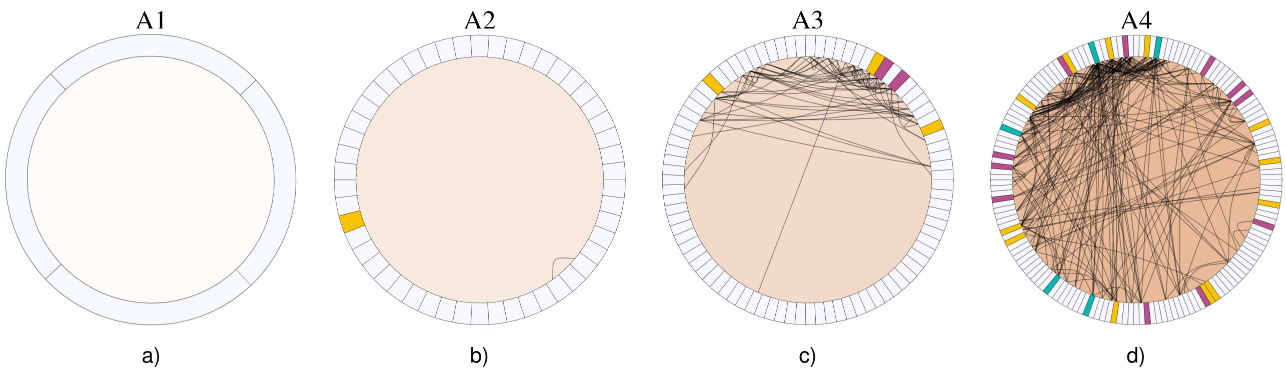


Fig. 5: Prototypical groups for each of the four archetypes. All groups have a belongingness coefficient greater than 0.98 with respect to the archetype in question. Each sector represents one group member colored with respect to the role in the group (= moderator, = sherpa, = sherpa & moderator). Friends are connected by lines. The background color of the inner circle reflects the group's activity score (0 352).

	A1				**A2**				**A3**				**A4**			
	min	max	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std
activity score	0	141	3.4	16.9	0	4	0.19	0.74	0	214	11.4	25.4	11	418	130.5	72.1
group size	3	85	29.8	17.6	4	83	35.9	22.2	46	155	74.9	16.4	75	203	106.7	25.5
no. of moderators	0	4	0.22	0.74	0	0	0	0	0	4	0.42	0.84	1	9	3.8	1.2
no. of sherpas	0	15	1.26	2.30	0	3	0.25	0.67	0	21	3.64	3.44	0	34	15.2	6.3
density	0	0.67	0.01	0.046	0	0.007	0.0004	0.001	0	0.102	0.006	0.015	0.011	0.135	0.055	0.031
$\overline{\text{light level}}$	168.3	396.5	331	38.1	100	261.1	110.8	34.4	261.2	387.5	329.3	22.9	350.71	389.53	375.5	9.8
$\overline{\text{level}}$	25	40	39.2	1.9	1	40	23.7	12.3	34.8	40	39.4	0.8	38.1	40	39.6	0.4
count	258				32				223				60			

stable relationships a human can maintain. For example, work on *Twitter* [53] found evidence of such an upper limit, finding that users can entertain up to 100–200 stable relationships. In the context of games this number has to be found to be much smaller. Ducheneaut et al. [15] found that most guilds in *World of Warcraft* have 35 or fewer members. Chen et al. [16] found the average guild size to be between 50 and 60 members and noted that the greatest instability seems to occur at a group size of around 60 with an average level of 60. Our results do not indicate such a size limit. This is probably due to guilds requiring more communication and organization than groups on player matching services which are intended to easily allow players to find playmates. Building larger groups seems thus to be more desirable in this context as it offers players more potential playmates. However, we should note that the maximum group size was 203 in our dataset. Williams et al. [54], also looking into *World of Warcraft* guilds noted that smaller groups tend to be more focused on social bonds whereas in our case density increases with group size as well. Again, this might be a consequence of the different purposes of guilds and the `the100.io`. With increasing group size usually more effort has to be spent to maintain cohesion among the group members. Our results show that as groups get larger, they are getting more organized with larger groups having a larger number of sherpas and moderators or even people who are taking on both roles. Indeed, the number of moderators together with the connectedness of the moderators seems to be the largest predictor of group activity. Moderators seem to act as a sort of facilitator of play within the groups. In this sense, it might thus be worthwhile to ensure that moderators are also present in smaller groups. At that point it might also be worthwhile to reemphasize that while the connectedness of moderators predicts activity we did not find evidence that the same also holds true for the overall interconnectedness of the group (i.e. density).

The results of the archetypal analysis give us an impression of the overall distribution of the groups. In some way, the identified archetypes can be viewed as reflecting the evolution of groups on the `the100.io` with groups starting small and then developing into larger, better connected and organized groups. Viewed from an activity perspective, at the top end of the activity spectrum we have groups characterized by having many quite well connected members, many moderators and sherpas ($A4$). This reflects the results from the correlation and regression analysis that moderators seem to serve as a facilitator of activity. However, assigning groups to the archetype based on their highest membership values shows us that only about 10.5% of the groups in our dataset fall within this high-activity cluster. At the lower end we have groups with a small number of members with none or only a small numbers of moderators and sherpas (e.g., groups mainly belonging to $A1$ or $A2$). These are most likely groups recently created on the `the100.io`. In terms of activity, a large number of groups falls within these two extrema but still have rather low activity with an average activity score of 11.4 if assigned to their dominant archetype (cf. Table **??**, $A3$). Providing means for such groups to reach the characteristics of $A4$ may help foster activity. For example, a LFG platform could provide recommendations to a group's founder for promoting group members (preferably highly active and well connected ones) to moderators. A k-means cluster analysis led to four clusters with similar characteristics adding support for the AA solution.

In terms of activity over time we witnessed lower activity on weekends than on weekdays with activity peaking on Tuesday evenings, irrespective of casual or serious groups. This peak seems to coincide with the weekly reset time where many activities and rewards are reset by *Bungie*, which as of the time of data collection took place at 2:00 AM Pacific time (see [55]). In general, afternoons and evenings are the preferred gaming times for all days of the week. As such it seems advisable to encourage events on weekday evenings where they are likely getting more attention.

In terms of the limitations of the current study we should note that we focused on one specific game – *Destiny* – in this study. However, metrics which have been specific to *Destiny* – level and light level – did not lead to any relevant conclusions. Since the other metrics used are mainly independent of the actual game, we believe that our results may also appear when looking at groups playing other, similar, games. However, we need further investigation to confirm this assumption. Furthermore, while we were able to obtain data on scheduled games we could not verify if these games really took place or how many players have participated in these games in the end. As we did not have access to these data, we could also not assess how long the games lasted or if they fell apart immediately after starting. To take this factors into account, one would need to be able to relate the games scheduled on the `the100.io` with the actual instances in *Destiny*. Despite this, we believe our paper contributes to the study of general MMO matchmaking and player behavior through the lens of *Destiny* and the `the100.io`.

Having said that, there are also several interesting avenues for further research. Among others, as we only looked at a snapshot of time it might be worthwhile to investigate how groups develop over time. Moreover, given that *Destiny* exposes in-game data through a publicly available API it might be interesting to observe how group structure correlates with in-game behavior of the groups or vice versa. Both directions could lead to further interesting insights on how groups need to be organized to stay healthy and active.

Lastly, while the work presented here is focused on the matchmaking service `the100.io`, it constitutes part of more considerable interdisciplinary challenges around how to handle group formation, group maintenance, and service, as well as overall community management in online environments [22], [29], [30], [31]. These are challenges that cut across domains such as information systems, human-computer interaction, social science, media, psychology, and application design. This provides a strong motivation to investigate social connections in and around games further.

VII. Conclusions

Online multiplayer and massively multiplayer games such as *Destiny* depend on players being able to find other people to play with [1], [4], [8], [12], [22], [27], [56]. Being able to analyze, categorize, and understand social structures in player communities, therefore, provides not only insights into online behavior but can also be leveraged by game companies to enhance matchmaking, player grouping, tune in-game activities, and events to the behavioral patterns of groups, as well as improve engagement via promoting group types that facilitate the requirements of the players. Being able to analyze online player communities at both the group level and the individual level can thus directly contribute to a more sophisticated user experience in multi-player games. As an example, identifying players that are not socially active and thus in danger of leaving, or groups who are not active or of sufficient size to foster activity, can be of great value in order to counteract negative development, for example, by providing the kind of help needed, incentives, or even adapting game content. In essence, understanding how to establish a thriving community which is well-aligned with the particular needs of a particular game can be a valuable asset for ensuring long-time engagement and retention.

The results presented in this paper contribute to the understanding of online player communities [4], [11], [22], [27], [32], [56]. A large-scale analysis of an online social player community has been presented, covering tens of thousands of players and integrating data about their social connections as well as self-report data about playing preferences. While social networks in games have seen some research in the past, such work has almost exclusively relied on implicit in-game "friends" connections (Jia et al. [11] being a notable exception), rather than communities established explicitly by players and formed around one or more games, where repeated shared activities permit evaluation of the strength of connections between players and groups.

Analyses have been presented that investigate correlations between group characteristics and activity level, showing that the size of the group, as well as the number of moderators and how well connected they are with the other group members correlate with activity. The influence of sherpas on activity has not been as high as we would have expected. Categories of groups have been generated using archetypal analysis, indicating four distinct types of player groups each with their own characteristics. Group activity was also presented as a function of weekdays to investigate when `the100.io` players schedule games, across groups comprised of self-reported casual or serious players. Finally, we here take an applied angle, describing and defining a series of metrics, for example, group characteristics, and models, such as archetypal analysis, which can be employed by game developers and community managers to gain insights into their communities whether formed through or around a game.

References

[1] B. Gajadhar, Y. de Kort, and W. IJsselsteijn, "Influence of social setting on player experience of digital games," in *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2008, pp. 3099–3104.

[2] N. Yee, "Motivations for play in online games," *CyberPsychology & Behavior*, vol. 9, no. 6, pp. 772–775, 2007.

[3] A. Alsén, J. Runge, A. Drachen, and D. Klapper, "Play with me? understanding and measuring the social aspect of casual gaming," in *Proc. of AAAI AIIDE Player Analytics Workshop*, 2016, pp. 115–121.

[4] A. Rattinger, G. Wallner, A. Drachen, J. Pirker, and R. Sifa, "Integrating and inspecting combined behavioral profiling and social network models in Destiny," in *Proc. Entertainment Computing - ICEC 2016*, G. Wallner,

S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, and H.-S. Yang, Eds. Springer, 2016, pp. 77–89.

[5] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, ""alone together?": Exploring the social dynamics of massively multiplayer online games," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2006, pp. 407–416.

[6] M. Seif El-Nasr, A. Drachen, and A. Canossa, *Game Analytics - Maximizing the Value of Player Data*. Springer, 2013.

[7] B. Maher, "Can a video game company tame toxic behaviour?" 2016.

[8] Y. M. Kow and T. Young, "Media technologies and learning in the Starcraft esport community," in *Proc. of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 387–398.

[9] N. Ducheneaut and R. J. Moore, "The social side of gaming: A study of interaction patterns in a massively multiplayer online game," in *Proc. of the 2004 ACM Conference on Computer Supported Cooperative Work*. ACM, 2004, pp. 360–369.

[10] A. Iosup, R. van de Bovenkamp, S. Shen, A. L. Jia, and F. Kuipers, "Analyzing implicit social networks in multiplayer online games," *IEEE Internet Computing*, vol. 18, no. 3, pp. 36–44, 2014.

[11] A. L. Jia, S. Shen, R. V. D. Bovenkamp, A. Iosup, F. Kuipers, and D. H. J. Epema, "Socializing by gaming: Revealing social relationships in multiplayer online games," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 2, pp. 11:1–11:29, 2015.

[12] A. Salovaara, M. Johnson, K. Toiskallio, S. Tiitta, and M. Turpeinen, "Playmakers in multiplayer game communities: Their importance and motivations for participation," in *Proc. of the 2005 International Conference on Advances in Computer Entertainment Technology*. ACM, 2005, pp. 334–337.

[13] G. Stafford, H. Luong, J. Gauch, S. Gauch, and J. Eno, "Social network analysis of virtual worlds," in *Proc. of the 8th International Conference Active Media Technology*, R. Huang, A. A. Ghorbani, G. Pasi, T. Yamaguchi, N. Y. Yen, and B. Jin, Eds. Springer, 2012, pp. 411–422.

[14] W. Mason and A. Clauset, "Friends ftw! friendship and competition in Halo: Reach," in *Proc. of the 2013 Conference on Computer Supported Cooperative Work*, ser. CSCW '13. ACM, 2013, pp. 375–386.

[15] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, "The life and death of online gaming communities: A look at guilds in World of Warcraft," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2007, pp. 839–848.

[16] C.-H. Chen, C.-T. Sun, and J. Hsieh, "Player guild dynamics and evolution in massively multiplayer online games," *CyberPsychology & Behavior*, vol. 11, pp. 293–301, 2008.

[17] N. Poor, "What MMO communities dont do: A longitudinal study of guilds and character leveling, or not," 2015.

[18] Bungie, "Destiny," Game [PlayStation 3, PlayStation 4, Xbox 360, Xbox One], Santa Monica, CA, USA, September 2014.

[19] A. Cutler and L. Breiman, "Archetypal analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, 1994.

[20] A. Drachen, R. Sifa, C. Bauckhage, and C. Thurau, "Guns, swords and data: Clustering of player behavior in computer games in the wild," in *IEEE Conference on Computational Intelligence and Games*. IEEE, 2012, pp. 163–170.

[21] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 266–278, 2015.

[22] C. Pearce, T. Boellstorff, and B. A. Nardi, *Communities of play: Emergent cultures in multiplayer games and virtual worlds*. MIT Press, 2011.

[23] T. Manninen, "Interaction forms and communicative actions in multiplayer games," *Game studies*, vol. 3, no. 1, p. 2003, 2003.

[24] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, "Alone together?: exploring the social dynamics of massively multiplayer online games," in *Proc. of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 407–416.

[25] C. Thurau and C. Bauckhage, "Analyzing the evolution of social groups in World of Warcraft," in *Proc. of the 2010 IEEE Symposium on Computational Intelligence and Games*. IEEE, 2010.

[26] S. H. Goh and M. M. Wasko, "Leadership in MMOGS: Emergent and transformational leadership candidates," in *AMCIS 2009 Proceedings*, 2009.

[27] A. Losup, R. Van De Bovenkamp, S. Shen, A. L. Jia, and F. Kuipers, "Analyzing implicit social networks in multiplayer online games," *IEEE Internet Computing*, vol. 18, no. 3, pp. 36–44, 2014.

[28] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Social network analysis for organizations," *Academy of management review*, vol. 4, no. 4, pp. 507–519, 1979.

[29] M. Grandjean, "A social network analysis of Twitter: Mapping the digital humanities community," *Cogent Arts and Humanities*, vol. 3, no. 1, p. 1171458, 2016.

[30] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *science*, vol. 323, no. 5916, pp. 892–895, 2009.

[31] J. Scott, *Social network analysis*. SAGE, 2012.

[32] J. Blackburn, N. Kourtellis, J. Skvoretz, M. Ripeanu, and A. Iamnitchi, "Cheating in online games: A social network perspective," *ACM Transactions on Internet Technology (TOIT)*, vol. 13, no. 3, p. 9, 2014.

[33] J. Pirker, A. Rattinger, A. Drachen, and R. Sifa, "Analyzing player networks in Destiny," *Entertainment Computing*, 2017.

[34] J. Bohannon, "Game-miners grapple with massive data," *Science*, vol. 330, no. 6000, pp. 30–31, 2010.

[35] A. Drachen, E. Lunquist, Y. Kung, P. Rao, D. Klabjan, R. Sifa, and J. Runge, "Rapid prediction of player retention in free-to-play mobile games," in *Proc. of AAAI AIIDE*, 2016.

[36] R. Sifa, A. Drachen, and C. Bauckhage, *Profiling in Games: Understanding Behavior from Telemetry*, 2017.

[37] A. Drachen, A. Canossa, and G. N. Yannakakis, "Player modeling using self-organization in Tomb Raider: Underworld," in *IEEE Symposium on Computational Intelligence and Games*. IEEE, 2009, pp. 1–8.

[38] R. Thawonmas and K. Iizuka, "Visualization of online-game players based on their action behaviors," *International Journal of Computuer Games Technology*, vol. 2008, pp. 5:1–5:9, 2008.

[39] A. Normoyle and S. T. Jensen, "Bayesian clustering of player styles for multiplayer games," in *Proc. Artificial Intelligence and Interactive Digital Entertainment International Conference*. AAAI Press, 2015, pp. 163–169.

[40] C. Bauckhage, R. Sifa, A. Drachen, C. Thurau, and F. Hadiji, "Beyond heatmaps: Spatio-temporal clustering using behavior-based partitioning of game levels," in *IEEE Conference on Computational Intelligence and Games*. IEEE, 2014, pp. 1–8.

[41] A. Drachen, J. Green, C. Gray, E. Harik, P. Lu, R. Sifa, and D. Klabjan, "Guns and guardians: Comparative cluster analysis and behavioral profiling in Destiny," in *IEEE Conference on Computational Intelligence and Games*. IEEE, 2016, pp. 1–8.

[42] A. Drachen, C. Thurau, S. R., and B. C., "A Comparison of Methods for Player Clustering via Behavioral Telemetry," in *Proc. of FDG*, 2013.

[43] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[44] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. Sage Publications Ltd., 2013.

[45] R. Sifa, A. Drachen, C. Bauckhage, C. Thurau, and A. Canossa, "Behavior evolution in Tomb Raider Underworld," in *IEEE Conference on Computational Intelligence in Games*. IEEE, 2013, pp. 1–8.

[46] C. Bauckhage and R. Sifa, "k-Maxoids Clustering," in *Proc. of KDML-LWA*, 2015.

[47] C. Thurau, K. Kersting, and C. Bauckhage, "Yes we can: Simplex volume maximization for descriptive web-scale matrix factorization," in *Proc. of the 19th ACM International Conference on Information and Knowledge Management*. ACM, 2010, pp. 1785–1788.

[48] R. Sifa, C. Bauckhage, and A. Drachen, "Archetypal game recommender systems," in *Proc. of the 16th LWA Workshops: KDML, IR and FGWM*, 2014, pp. 45–56.

[49] C.-U. Lim and D. F. Harrell, "Revealing social identity phenomena in videogames with archetypal analysis," in *Proc. of the 6th International AISB Symposium on AI and Games*, 2015.

[50] R. Sifa, C. Bauckhage, and A. Drachen, "The Playtime Principle: Large-scale Cross-games Interest Modeling," in *Proc. of IEEE CIG*, 2014.

[51] C. Aggarwal and C. Reddy, *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, 2013.

[52] R. I. M. Dunbar, "Coevolution of neocortical size, group size and language in humans," *Behavioral and Brain Sciences*, vol. 16, no. 4, pp. 681–694, 1993.

[53] B. Goncalves, N. Perra, and A. Vespignani, "Modeling users' activity on Twitter networks: Validation of dunbar's number," *PloS one*, vol. 6, no. 8, 2011.

[54] D. Williams, N. Ducheneaut, L. Xiong, Y. Zhang, N. Yee, and E. Nickell, "From tree house to barracks the social life of guilds in World of Warcraft," *Games and culture*, vol. 1, no. 4, p. 338361, 2006.

[55] Bungie, "Destiny: Daily and weekly reset times," 2017, https://www.bungie.net/en/Help/Article/45956, Accessed: November, 2017.

[56] M. Szell and S. Thurner, "Measuring social dynamics in a massive multiplayer online game," *Social networks*, vol. 32, no. 4, pp. 313–329, 2010.