

The Trails of Just Cause 2: Spatio-Temporal Player Profiling in Open-World Games

Myat Aung, Simon Demediuk
mta510@york.ac.uk, simon.demediuk@york.ac.uk
University of York
York, United Kingdom

Rafet Sifa
Rafet.Sifa@iais.fraunhofer.de
Fraunhofer IAIS
Sankt Augustin, Germany

Yuan Sun, Ye Tu, Yu Ang, Siva Nekkanti,
Shantanu Raghav, Diego Klabjan
yuansun2015, YeTu2014, YuAng2015, SivaNekkanti2015:
ShantanuRaghav,d-klabjan,@northwestern.edu
Northwestern University
Chicago, Illinois

Anders Drachen
anders.drachen@york.ac.uk
University of York
York, United Kingdom

ABSTRACT

Behavioral profiling of players in digital games is a key challenge in game analytics, representing a particular challenge in Open-World Games. These games are characterized by large virtual worlds and few restrictions on player affordances. In these games, incorporating the spatial and temporal dimensions of player behavior is necessary when profiling behavior, as these dimensions are important to the playing experience. We present analyses that apply cluster analysis and the DEDICOM decompositional model to profile the behavior of more than 5,000 players of the major commercial title *Just Cause 2* integrating both spatio-temporal trails and behavioral metrics. The application of DEDICOM to profile the spatio-temporal behavior of players is demonstrated for the purpose of analysing the entire play history of *Just Cause 2* players, but also for the more detailed analysis of a single mission. This showcases the applicability of spatio-temporal profiling to condense player behavior across large sample sizes, across different scales of investigation. The method presented here provides a means to build profiles of player activity in game environments with high degrees of freedom across different scales of analysis - from a small segment to the entire game.

CCS CONCEPTS

• Information systems → Clustering; • Applied computing → Computer games.

KEYWORDS

game analytics, behavioral profiling, game development, behavior mining, profiling, open-world game

ACM Reference Format:

Myat Aung, Simon Demediuk, Yuan Sun, Ye Tu, Yu Ang, Siva Nekkanti, Shantanu Raghav, Diego Klabjan, Rafet Sifa, and Anders Drachen. 2018. The Trails of Just Cause 2: Spatio-Temporal Player Profiling in Open-World Games. In *FDG '19: ACM Foundations of Digital Games, August 26–30, 2019, San Luis, CA*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3337722.3337765>

1 INTRODUCTION

Behavioral profiling forms one of the core challenges of game analytics because it condenses what can be very varied (high-dimensional), volatile and potentially high volume data about the behavior of players within the confines of a game into descriptions that highlight the patterns of player behavior [2–5, 7, 8, 10–12, 15, 19, 23, 25, 26, 29, 32, 33]. The purposes of profiling can be varied, from design evaluation, progression analysis, user experience evaluation, and even purely explorative. Jointly, profiling helps build an understanding of the users. However, behavioral profiling in digital games is not a straightforward task due to the shifting requirements of a profiling exercise, common high-dimensionality in the data, volatility and the lack of clear guidelines for which types of behavioral features to incorporate into profiles [2, 3, 10, 14].

These problems are notably present in games where players have wide degrees of freedom in how they want to approach and play the games, for example in some Massively Multiplayer Online Games (MMOGs) and some action-adventure games. This is especially the case for Open-World Games (OWGs) commonly referred to as "Sandbox" games due to their nature. Open World Games are characterized by featuring large virtual worlds that can span hundreds of square kilometres of virtual real estate, with very few restrictions on the freedom of a player to go where they please; and a corresponding range of affordances, e.g. in how to accomplish objectives. In addition, OWGs also typically span longer game lengths, with some extending into the hundreds of hours. Examples stretch back to *Ultima Online* (from 1981) which used a world design called "overworld" with an Open World design. *Elite* (from 1984) fielded a large open section of space for players to explore. In more recent times, games such as *The Elder Scrolls: Oblivion* (2006), *Skyrim* (2011), *Minecraft* (2011), *EVE Online* (2003) and the *Grand Theft Auto-series* (1997 onwards) have all been based on OWG mechanics.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
FDG '19, August 26–30, 2019, San Luis Obispo, CA, USA
© 2019 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-7217-6/19/08.
<https://doi.org/10.1145/3337722.3337765>

Subsequently, OWGs may have players move across their worlds several times before they are finished.

The spatial and temporal dimensions of play are important in any digital game, as they represent the dimensions that games are experienced through. Work across game analytics, game AI and other domains have generally incorporated one or both dimensions [14, 26, 38, 39]. Ignoring either time or space in behavioral analyses can lead to information loss and potentially misleading results, notably where the spatial dimension is important to the playing of the game [5, 9, 37]. Historically, simple visualizations of spatial player behavior have been utilized, notably heatmaps. However, heatmaps are limited in that they ignore directional and temporal information [9, 37].

A defining characteristic of OWG design is its facilitation of a range of player motivations and playstyles over prolonged periods of play. However, this also means that profiling of player behavior either for a section or the entire game must incorporate those same highly varied behaviors within the data and analytical techniques [2]. This requirement is a significant justification for the necessity of analyzing spatial and temporal data in player research. The work presented here focuses on developing profiling techniques that address the requirements of OWGs specifically, integrating spatio-temporal telemetry data.

This game genre has not been the subject of profiling work in the past. Since OWGs allow players to move freely instead of limiting their trajectory to set paths, this increases the complexity of making data-driven calculations as each trajectory can be theoretically unique. The profiling technique presented here significantly expands the applications of the Decomposition into Directional Components (DEDICOM) framework adapted to games by Bauckhage et al. [3], a spatial analysis technique that produces more detailed information than aggregate heatmaps. While Bauckhage et al. [3] focused on player trails, this study integrates other dimensions of player behavior, such as characterizing player types by the actions performed during play. Furthermore, the DEDICOM technique has only been applied to game worlds with linear levels and shorter periods of play (a *Quake III* demo and an *Unreal Tournament 2003* level), while the application is here focused on the unconstrained play of OWGs across the entire playing history of the players. This work also utilizes a large sample of real player telemetry of extended play duration.

In the solution presented here, trails are condensed to clusters of checkpoints rather than raw spatio-temporal data. We combine DEDICOM-generated profiles with visualization of player behavior, and apply these techniques across play histories as well as for specific segments of a game. As a case study, we employ a high-dimensional dataset from the major commercial OWG *Just Cause 2* (JC2). This dataset is then split into two parts, one for Early Dropouts, those stopping shortly after starting to play, and Committed Players who keep playing for a longer period. This split allows the development of behavioral profiles for specific segments of players, which in turn helps inform about differences in player experience. This also showcases how DEDICOM can be applied based on pre-defined parameters of interest. The goal is to develop and present a combinative method that can accommodate large data of OWGs and address their specific challenges.

2 CONTRIBUTION

Based on a dataset covering more than 5,000 players from the major commercial title *Just Cause 2*, analyses are presented that apply cluster analysis and the DEDICOM decompositional model to profile player behavior in the context of open-world games. While previous work has investigated non-spatio-temporal profiling or geographical trails independently [5, 12, 13, 21, 23, 37], the presented technique incorporates spatio-temporal dimensions. The results provide a means for building behavioral profiles in unrestricted game environments to interpret the player experience of specific components of a game world in its entirety.

3 RELATED WORK

Behavioral profiling based on telemetry data from digital games is a recent development within the larger domain of game analytics [14, 16, 20, 24], although profiling or segmentation of players in general has been utilized to guide game design and game AI for close to two decades [4, 7, 8, 13, 15, 20, 26, 38]. Behavioral profiling in games is today generally driven by gameplay telemetry data or from other interactions between players and game ecosystems such as social media, attribution models, psychographics and demographic information [28]. Such profiling in games is based on a variety of techniques ranging from descriptive statistics to machine learning utilized to address a variety of goals, including human behavior research, design evaluation, monetization, optimization, debugging and exploration [7, 10, 13, 24, 28, 29, 33, 35]. The current work on telemetry-based, behavioral profiling in games across academia and industry was reviewed recently by Sifa et al. [30].

In parallel with the work in profiling, the visualization and evaluation of player behavior has been adopted across industry and academia. For example, Hoobler et al. [18] visualized player trails and build heatmaps for *Return to Castle Wolfenstein: Enemy Territory*. Drachen & Schubert [9] provided a review of spatial analysis and visualization work in games. A few examples of visualization work include Wallner & Kriegelstein [37], which presented a visual analytics system that could also handle spatial data. Mirza-Babei et al. [22] discussed visualization of quantitative data in detail. Nguyen et al. [23] presented a visualization tool for puzzle games which also involved spatio-temporal operators. Miller and Crowcroft [21] investigated group movement in *World of Warcraft* using waypoint modelling. With specific regard to analyzing spatio-temporal data, Drachen et al. [12] clustered players according to spatio-temporal behavior and distance between players, while Rioult et al. [26] used topological measures to predict esports match outcomes. Bauckhage et al. [3] adopted the DEDICOM model [6, 31, 34] to cluster players of *Quake: Arena* and develop waypoint graphs for behavior-based partitioning. However, other behavioral information was not integrated in the analysis (further details of DEDICOM is provided below).

4 JUST CAUSE 2: GAMEPLAY

Just Cause 2 is an open-world (sandbox) single-player action adventure game released by Square Enix in 2010. In the game, players adopt the role of an agent whose aim is to disrupt a fictive tyrannical regime on the island nation of Panau. The modelled environment covers 1035.55 square kilometres. The player is free to roam the

game's open world, with the goal to cause "chaos" and destabilize the island to a point where the tyrant can be toppled. The data in this study includes player behavior across the entirety of the game world.

Progress in the game can take place along a number of vectors, but primarily by advancing the main storyline or by causing "chaos". Chaos points are a progress measure earned in a variety of ways, primarily through missions of destroying government property, collecting items for the various rebel factions, etc. Chaos unlocks new agency missions and stronghold takeover missions. Building chaos and performing stronghold takeover missions increases the influence of the rebel factions, which unlocks further challenges and missions. Missions are generally classified in two types: 1) Agency missions and 2) Faction missions. Agency missions advance the main story line whereas the faction missions build chaos. The player can also collect various items that unlock or upgrade weapons and vehicles. Subsequently, the number of missions that players complete are indicative of their progress. As the game progresses, missions lead the player to explore different parts of the map with various means of transportation including a variety of cars, boats, planes, and via parachute. One of the special features of this game is the ability to allow a player to use a grappling hook to pull themselves quickly towards an object, used with the parachute to create a kite-like flight, and also to tie things together. Further, the player can call for a helicopter to be transported to another selected point to save the time travelling across the island. Both the options of transportation as well as the paths that players choose are incorporated into our analysis below. Finally, early in the game the player encounters a supplier from the black market to order progressively advanced equipment and vehicles. The use of these drops are included to identify players capable of deploying advanced game features to accomplish objectives. A summary of game features integrated into the analysis is provided in Table 1.

5 DATA AND PRE-PROCESSING

The dataset used in this work was provided by Square Enix, the publisher of JC2, via a research collaboration. The data is made up of a direct extract of random players from the game's servers in Fall 2010 from the EU version (data from the pirated version of the game removed, all data anonymized), across all three platforms of the game (PC, PS3, Xbox360). Due to confidentiality the raw data is not publicly available but supplementary materials include information and results across the analyses presented here. The data contains records of 5331 randomly sampled players across a 7 month period (from a population at the time of over 1.5 million). A total of 10,794,666 timestamped actions were included covering a large number of behavioral features. The data is formatted as an output to a SQL database query using Python. 61 players were removed due to a tracking error which meant that they were registered in terms of playtime, but no behaviors were logged. This left 5,271 players for analysis. This includes in-game geographical coordinates for player actions including time stamps. Also included is metadata such as the system used and system settings.

The players of the dataset exhibit a wide range of behaviors as is common in digital games [9, 14, 15, 24, 25], most features show a highly skewed distribution. Several metrics were defined and

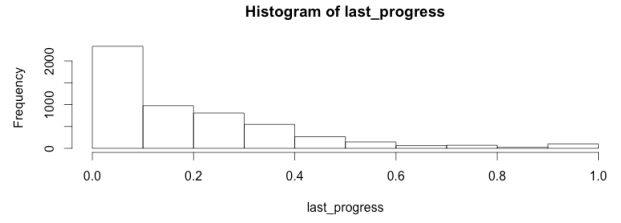


Figure 1: Progress of players in JC2. Note the marked drop-off after 10% game progress, which led to the players in the dataset being divided up into two groups: those with less than or equal to 10% progress, and those more than 10% progress.

calculated to be used in the analyses (Table 1). The Just Cause 2 dataset provides a comprehensive spatio-temporal coverage of the play histories. Initially, six different causes of death were included as features. However, correlations show that the sources of player death are highly correlated. The correlation matrix provides an overall r value of 0.94 or greater, which indicated that players who died a lot are likely to die from a variety of sources. Thus these are collapsed into an aggregate metric.

Early dropouts vs. committed players: One of the aims of the analysis was to evaluate behavioral differences across players who left the game early ("Early dropouts", $n=2306$), and those who keep playing for a longer period ("Committed players", $n=2965$). This allows profiles to be developed based on pre-defined parameters and provides a more granular view on the players. Based on the initial exploratory data analysis, a marked difference in player behavior was observed before and after they reached 10% game progress (for comparison, completing all main storyline missions comprise roughly 35% progress, the remaining progress is faction missions, ancillary achievements, collectable items, etc.); as well as a lot of players leaving the game around this progress point (Figure 1). Here progression is defined as a function of the entire base of missions and discoverable locations, unlike Pirker et al. [25] who only used the main storyline as the basis for evaluating progression. Each approach is useful but provides different viewpoints on progression analysis. Here the focus was on covering as much of the player behavior as possible.

The behavioral variance is possibly due to players initially needing to develop a familiarity with the game and its controls, as they progress through the first few on-boarding missions (10% progress is roughly the time when players have been through the introductory missions). Additionally, the players leaving the game early might constitute a different mix of skillsets and motivations. Changes in player behavior as a function of progress were also reported by Sifa et al. [29]. To accommodate these behavioral differences, the dataset was split into two groups based on their game progress values. This provides the ability to compare behaviors of those players operating early in the game, and those in later phases. In principle, this division can be made for any measure of progress, and multiple bins can be used similar to Sifa et al. [29]. This division will be used in the following analyses, with the following notation

Metric	Description/ statistics	Comment/Assumptions
K/D ratio	Ratio of enemy kills per player death event.	K/D ratio is used as a proxy measure of the players' expertise in JC2. This metric known from shooter games and esports[10] is a measure of how many enemies a player killed as a function of how many deaths the player suffered ($\mu = 8.2; \theta = 14.2$).
%Hardcore	Percentage of time played in "hardcore" mode.	Players can choose from four levels of difficulty: casual, normal, experienced and hardcore. Hardcore is the most difficult mode and requires higher skill and interest in this level of challenge (only 260 players used the "hardcore" mode).
Elite enemies	Number of enemy elites killed by the player.	Elites are a particular type of enemy encountered notably in the later part of the game. They are more sophisticated and better equipped, requiring some level of skill to defeat ($\mu = 133.8; \theta = 297$).
Weapon upgrades	Number of weapon upgrades.	Players can upgrade their weapons and vehicles through specific items found, or purchased via the black market. Using upgrades is an advanced skill feature ($\mu = 13.6; \theta = 31.5$).
Extractions	Number of times extraction has been used.	Extractions are a form of transportation in JC2, allowing instant transportation to a previously visited safe house location. The player is dropped from a helicopter, deploying a parachute ($\mu = 26.2; \theta = 50.3$).
Item pickups	Number of resource items played has collected.	Players can collect resource items during the game, including vehicle parts, armor parts, weapon parts and cash stashes. These are often found in out-of-the-way locations and represent a proxy measure of exploration behavior ($\mu = 119.8; \theta = 259$).
Playtime	Total playtime in seconds.	Total time spent for the specific playthrough of the game ($\mu = 262,000; \theta = 514,000$).
Sabotage points	Total amount of sabotage points earned.	Certain objects can be destroyed for Chaos points (i.e. progress). The activity of doing so could be called Sabotage, which generates sabotage points ($\mu = 143.3; \theta = 299$).
Game progress	Total progress in percentage.	Progress is a proxy measure for overall progress in the game, defined as a combination of missions completed, faction missions completed, area of the game that has been explored and other factors. ($\mu = 0.19; \theta = 0.2$).
Deaths	Number of times a player died.	Deaths from all potential sources (bullet, drowning, explosion, fire, impact, melee) ($\mu = 39.2; \theta = 51.7$).
Heavy drops	Number of heavy drops requested.	Players can use the black market to request a delivery of different vehicles to their location. Represents an advanced game feature ($\mu = 28.2; \theta = 87.8$).
Difficulty	Difficulty of game on 0-3 scale.	For each session played, a difficulty level is chosen by the player from casual, normal, experienced and hardcore. The game difficulty level for a player is calculated by averaging all difficulty levels ($\mu = 0.7; \theta = 0.6$).

Table 1: Behavioral metrics from the Just Cause 2 dataset

to separate between the two player groups: (1) Players whose maximum progress is 10%: Early Dropouts (43.7%); (2) Players whose maximum progress is greater than 10%: Committed Players (52.3%).

6 ANALYSES

In this section we describe three different analyses performed on the JC2 dataset: 1) DEDICOM is employed to build pattern groups for the spatio-temporal trails of the JC2 players. 2) Spatio-temporal profiles are constructed for the spatio-temporal trails and combined with behavioral features across the whole play history of the player;

3) Using the method from 2), a specific mission in Just Cause 2 is analyzed to evaluate the usefulness of the proposed approach at granular levels.

6.1 DEDICOM-based map partitioning and trajectory-based profiling

The focus here is on the time-stamped location trails of the players (in three dimensions: x, y and z). Player location is recorded in the JC2 telemetry engine every time any of a number of actions

happen, e.g. an enemy eliminated, the player dies, enters a vehicle, uses a parachute, etc. Combined, these form trails that are less accurate than frequency-based trails (e.g. recording location every second), but are more efficient in terms of bandwidth while preserving key behavioral information. The goal of the analysis is to segment the Panau map, group players based on their spatio-temporal behavior, and visualize the result. In order to accomplish this the DEDICOM algorithm and process are utilized. DEDICOM was adapted for use in 3D game environments by Bauckhage et al. [3] (the semi-nonnegative DEDICOM version from [31] is used here) who compared the approach with three other models, including k-means and spectral clustering, noting the superior ability of DEDICOM to identify structures in waypoint graphs that reflect the topology of game environments. We therefore do not compare multiple models here, but rather focus on investigating if DEDICOM can be utilized in OWG contexts, with the overall purpose of evaluating the usefulness of integrating spatio-temporal features with behavioral profiling for these games, described in Section 6.2.

DEDICOM is a low rank approximation of asymmetric similarity matrices which is attributable to Harshman [17]. The method finds its applications in areas such as analysis of asymmetrical directional relationships between objects or persons [17], natural language processing [6] and churn analysis [31]. Given an asymmetric square similarity matrix S , the objective of DEDICOM is to factorize S into lower-rank matrices A and R that respectively represent the matrix of loadings (or basis) and the matrix containing directional affinities [31]. All DEDICOM results in this paper were obtained from asymmetric matrices whose entries indicate spatio-temporal waypoint similarities that are generated by the procedure described in Bauckhage et al. [3]. That is, using k-means clustering to cluster locational information of players we extracted central waypoints and later created a waypoint graph encoding movement information between the automatically extracted waypoints. Note that the extracted waypoint graphs (especially for OWGs) are usually time or location-asymmetric. Upon convergence of this algorithm, we apply k-means clustering to the n rows of A . Although DEDICOM produces similar clusters as spectral clustering, it also characterizes affinities among the resulting clusters. For affinity matrix R , R_{ij} shows the affinity between cluster i and cluster j . The larger the number is, the more similar these clusters are.

Based on the setting of the game, we expect that players who progress more and stay longer in the game, Committed Players, have more accessibility to the map, thus their trajectories should be similar to each other. Early Dropouts will only move within certain areas as bounded by the mission, thus have high self-affinity for some trajectories. Therefore, data for these two groups of players were processed and evaluated separately. For each group, we first identify n prototypical waypoints from player's sequential 3D positions by applying k-means clustering. Having those n prototypical waypoints, the waypoint transition graph for each player is built. We then use the shortest Euclidean distance as a weighting scheme to obtain spatial similarity between any two waypoints (the number of transitions of a player between two waypoints is weighted by the corresponding shortest Euclidean distance). For each player p , we calculate one similarity matrix S_p from its waypoint graph.

Adding up all similarity matrices to the sum matrix S and performing the DEDICOM algorithm on matrix S (decomposing it



Figure 2: Example DEDICOM-generated waypoint graph map regions for Committed Players, showing the behavior-based partitioning, i.e. trajectory groups. The waypoint graph-based regions reflect overall trends in gameplay. Players are starting in the main continent (yellow segment). Following an agency mission placed in the magenta segment, each of the other segments form the basis for agency missions.

to ARAT), we obtain matrix A for all trajectories across players. The Panau map is segmented into different regions by performing another k-means clustering on rows of matrix A (n waypoints). For this analysis, DEDICOM provides 5 regions to demonstrate how to segment the game map based on player spatio-temporal behaviors. For the Early Dropouts sample of players, we observe a large segment on the coastal and edge of the map and small dense segments on the main continent and islands. This reflects the early game since players are starting with the first few missions and have relatively limited access to the map. All trajectories outside the missions appear to be random walks. In contrast with Early Dropouts, the Committed Players' trajectories have been separated evenly into 5 segments, each located in a major mission zone (Figure 2). These broadly reflect the location of the major agency missions, and one segment representing trajectories in side missions.

After getting a fixed A , we derive R_p (p refers to a particular player) so that AR_pAT approximates S_p . Applying k-means one last time on the flattened affinity matrix R_p , we can now cluster the JC2 players by their movements between the 5 trajectory groups, and compare the resulting player clusters. Matrices R_p of the cluster centers are compared to better understand the difference between players. The analysis distributes Early Dropouts into 6 trajectory-based clusters (Figure 3). We observe that all five trajectory groups for Cluster 2 have high self-affinity and relatively low affinities



Figure 3: Cluster centroids of matrix R for the *Early Dropouts* players. For each cluster the number corresponds to the description in the text. The number of players in each cluster is included below the cluster caption. Each affinity matrix is represented by segments forming trajectory groups as defined by DEDICOM

with the others. Cluster 3 is the smallest with very high self-affinity and inter-affinity between trajectory groups 1 and 2. Cluster 1 and Cluster 4 players have a similar pattern of affinity in the diagonal entries. Cluster 4 shows higher self-affinity in trajectory group 2 than 1. For the Committed Players the trajectories are much more differentiated, possibly as a direct result of these players spending more time in the game and thus having more time to explore the environments. Based on players' movements between the 5 trajectory groups, and the k-estimation techniques described above, an 8 cluster solution was developed for the Committed Players (Figure 4). We observe high self-affinity and low affinity between trajectory groups, and the difference between self-affinity and inter-affinity is substantially higher than Early Dropouts. This indicates that players generally move in a certain pattern during one mission, greatly influenced by the actual game environment and setting for a certain mission. If comparing pairs of affinities R_{ij} and R_{ji} , we also have more information about the player's movements between two trajectory groups, which shows that players jump from one trajectory group to the other group more frequently than the other direction. Visualizations of the trails representative of each cluster are not included here due to space limitations, but examples of the type of visualizations that can be created are included in the player mission trails in Figure 5 and 6.

6.2 Profiling spatio-temporal behaviors

In this analysis, the R matrix results were considered with the non-spatio-temporal metrics describing player behavior in JC2. Clustering is based on each entry of the R -matrix, but profiling the cluster tied back to the non-spatio temporal statistics of each cluster, resulting in profiles that integrate both sources of data that provides a more comprehensive view of player behavior.

For visualization purposes, for each player cluster (cluster of matrix R), we identify a typical player and visualize the player's activity trajectory. The typical player is the one whose R matrix is closest to the cluster's centroid. The closeness is measured by the Euclidean distance between the centroid R matrix and the player's R matrix. Following identification of the typical player, visualizations were created that show their geographical trails across the

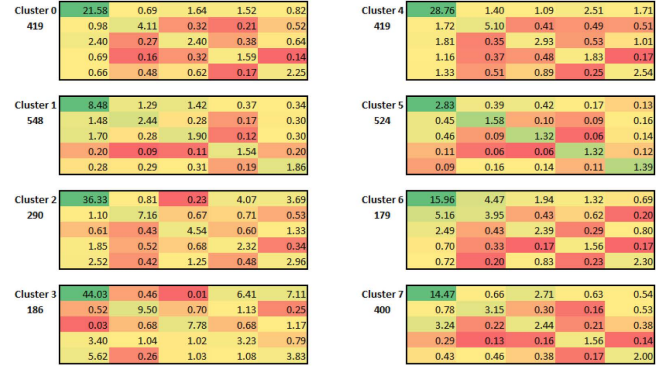


Figure 4: Cluster centroids of matrix R for the *Committed Players*. For each cluster the number corresponds to the description in the text. The number of players in each cluster is included below the cluster caption. Each affinity matrix is represented by segments forming trajectory groups as defined by DEDICOM

Panau map combined with behaviors (Figure 5 and 6). The player clustering analysis uses players' affinity R matrices as features, and groups the players with the same pattern of spatial temporal movements. For each player cluster, we then examined nine key non-spatio temporal statistics (Table 1).

Metrics represent a variety of aspects of the player behavior and include key behavior indicators such as the Kill-Death (K/D) ratio as an indicator of overall skill [7]. Though this particular metric is more prominent in online and competitive games, it is plausible that the K/D ratio is still informative as a comparative measure in single player games. While the ratio might not provide extensive information on a player who records a high K/D ratio, the ratios still provide a scale to compare players of different skills, particularly those that are struggling due to factors mentioned in this paper such as excessive difficulty. It would be reasonable to select an alternate set of metrics depending on the specific purpose of the analysis (see e.g. El-Nasr et al. [14] and Drachen et al. [10]). Due to the skewness of the non-spatio-temporal metrics, the data was log-transformed and the mean of the transformed data used for the purpose of comparing statistics across the clusters.

A combination of cluster balancing, elbow methods, silhouette coefficients, and interpretability [1, 27, 34] was used to determine the size of k , for all values less than 12. The elbow method (gap statistic is the statistical version) [34] uses the output of clustering algorithms such as hierarchical clustering, and compares the change in within-cluster dispersion with that expected under a null distribution. Silhouette coefficients represent clusters using silhouettes, based on the comparison of its tightness and separation. The silhouette shows which objects lie well within clusters vs. between clusters, allowing evaluation of the quality of the clusters [1, 27].

For the Early Dropouts, a 6-cluster solution emerged based on the estimation techniques described above. Six profiles are described in Table 2, with a summary of the behavioral statistics in Figure 7. For the Committed Players, an 8-cluster solution was reached, described in Table 3, with a summary of the behavioral statistics in Figure 8. In an applied context, two pairs of cluster for the Committed Players

Cluster	Description
Cluster 0: Elite players (16.8%)	These players are characterized by having the highest K/D ratios, spend at least some time on hardcore mode, spend a lot of time in the game and progress well into the game. They perform well across all metrics except for using few extractions. They prefer to use vehicles to travel, and spend most of their time near the West side of the game.
Cluster 1: Hardcore dropouts (60.9%)	These players dominantly start on hardcore mode but quickly leave the game. They do not use upgrades and extractions. They are only active in the first few regions of the game. They constitute over 60% of the players and indicate a potential problem with how new players understand the difficulty setting, leading them to choose hardcore mode but giving up due to the challenge level. Testing of this hypothesis should be done via user testing. Spatially they are confined to a small area of Western Panau and mainly navigate on foot while killing.
Cluster 2: Easy Moders (1.9%)	Of the Early Dropouts, these players leave the game last, use resource items, sabotages, kill many elites, etc. but play on the easiest game mode and have a relatively low K/D ratio. They are active in a broad range of the map and use different ways of getting around. All three platforms are used by these players.
Cluster 3: Explorers (6.5%)	These players are similar to Cluster 2, however, they more commonly use hardcore mode. Furthermore, their spatial behavior is also different, navigating the map broadly, primarily using vehicles to get around.
Cluster 4: Motoring Tourists (8.3%)	These players are overall similar in their behavior to the hardcore players in Cluster 0, although exhibiting slightly lower ranks across the performance-related features. However, while they access the same areas of the map and have a similar mission profile, they do not use the parachute but navigate only with vehicles and by walking. They also access similar core areas of the map as Cluster 1 players, but the latter are spatially less wide-roaming. Additionally, they chose different missions after the initial default missions. The Cluster 4 players also explore Panau in a more linear fashion than Cluster 2 and 3 players.
Cluster 5: Hardcore Players (5.7%)	Like Cluster 0, these players prefer the hardcore mode of the game. They generally have similar performance features as Cluster 0, and better performance than Cluster 4. However, their spatial behavior is markedly different. They have a much broader spatial range than Cluster 0, primarily using vehicles, although with an emphasis on parachutes in elevated topography and vehicles in low-lying areas.

Table 2: Behavioral Profiles - Early Dropouts. A summary of the behavioral statistics can be found in Figure 7

are so similar that an argument could be made for combining them, resulting in six more sharply distinguished profiles.

6.3 Mission-specific profiling

In this final analysis, the focus is on applying DEDICOM to investigate player behavior around a smaller segment of the game to evaluate the potential of the technique for more narrowly defined spaces than an entire playthrough.

Cluster	Description
Cluster 0: Flyers (14.1%)	These players use Easy or Normal difficulty. Most metrics are around the average for the clusters. The spatial behavior of the players presents a substantial amount of travel around Panau, using parachutes in mountain areas and flying vehicles to cross between islands. There is only a minimal use of ground-based vehicles and travel along roads. Players in this cluster rarely travel to the Selatan Archipelago, which may indicate a lack of engagement with the Roaches'-faction missions in this area.
Cluster 1: Parachute Jumpers (18.5%)	Clusters 0 and 1 are overall similar but cluster 1 players use the parachute more, also to cross between islands with long parachute flights. Their performance is similar to Cluster 0, albeit with less overall usage of the advanced game features. They also progress less overall.
Cluster 2: High Performers (9.8%)	Similarities with Clusters 0 and 1 in terms of performance, except they rank higher across most indicators. However, their spatial pattern differs, possibly due to having a higher game progress. Spatially, these players navigate across the entire game environment, using a variety of transportation modes, although showing a preference for using the parachute to cross between islands. They are ranked second in using the hardcore game mode.
Cluster 3: Hardcore players (6.3%)	Traditional hardcore profile, not only scoring high or the highest on the performance features, but also mainly playing using hardcore difficulty, and progressing the furthest in the game. Their spatial behavior is similar to Cluster 2, with an even wider dispersal around the islands, as would be expected given their increased progression. They use a diverse range of methods to cross larger distances.
Cluster 4: High Performers II (14.1%)	Similar overall to Cluster 2 but have progressed less in the game and their spatial behavior reflects this. There also appears to be some variation in the amount of travel done per playtime unit, although further analysis is necessary to verify this. In terms of hardware platform, this is the most equally distributed cluster. In a practical context it could be argued that the players of Cluster 2 and 4 could be combined.
Cluster 5: Low performers (17.7%)	The players in this cluster exhibit generally low performance, except for their K/D ratio. They do not take advantage of features such as extractions and upgrades, and progress less into the game compared to Clusters 0 and 2-4. There are almost twice as many Xbox users in this cluster as PS users. Spatially, they spend most of their travelling using ground-based vehicles, and parachutes to cross between islands. They have not progressed beyond the two largest islands.
Cluster 6 and 7: Stay-at Homers (19.5%)	The players of these clusters are similar in terms of their overall behavior, and it could be argued that they represent the same play pattern. It is not certain if it is enough to functionally separate them by mission completion or geographic access. The main difference appears to be in the K/D ratio, where Cluster 7 performs better. Performance is similar to Clusters 0 and 1 for several features, but they travel less.

Table 3: Behavioral Profiles - Committed Players. A summary of the behavioral statistics can be found in Figure 8



Figure 5: Example of a typical player trajectory from DEDICOM-generated cluster 4 ("Motoring Tourists") of the Early Dropouts. Blue point lines indicate players entering or exiting parachute mode. Yellow points and lines indicate players entering or exiting a vehicle. Red points and lines indicate enemy kills. Remaining player trajectory visualizations in the supplementary materials

For this case, the *Mountain Rescue* mission was selected. This is the 4th major mission in the game, and at this point, players can be expected to be familiar with the mechanics of JC2. It is also designed to be one of the most challenging missions in the game. Players are tasked with rescuing a hostage kept in a secluded military base in the mountains of Panau. Players must get into a helicopter, navigate to the base, land, and destroy four objectives, allowing the player to enter the base. Thereafter players will need to enter the building and eliminate all the armed guards and ninjas. Players must then hijack convoy cars when a hostage during the mission is found to have been relocated to the bottom of a nearby cliff. It is advantageous to use the helicopter to initially fly around the base and eliminate destructible objects, which makes later stages of the mission easier.

There are 159 players who started this late-game mission, with an average of 10-20 action types for each player in the dataset. Following the same procedure as in the previous analysis, three clusters emerge, with typical trajectories visualized in Figure 9 and described in Table 4.

In summary, we were able to identify different player action types and cluster them by behavior. Further, the method was also able to identify potential problems players may have experienced and consequent losses of interest (only cluster 3 players, 6%, complete the mission). The analysis thus provides insights useful for future user testing and behavioral analysis, e.g. investigating if these players' experiences exist in missions elsewhere with similar outcomes

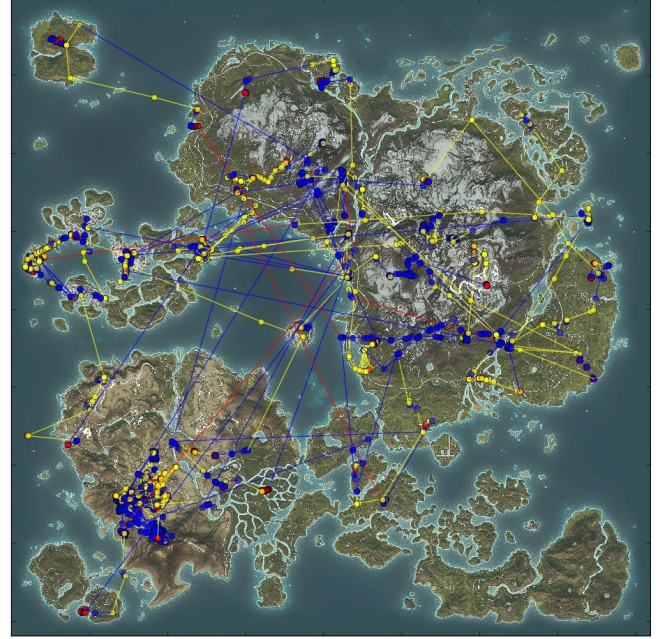


Figure 6: Example of a typical player trajectory from DEDICOM-generated cluster 0 ("Flying Players") of the Committed Players. Blue point lines indicate players entering or exiting parachute mode. Yellow points and lines indicate players entering or exiting a vehicle. Red points indicate enemy kills. Remaining trajectory visualizations available via the supplementary materials.

Cluster	Description
Cluster 1 (76.7%)	These are players that prefer using a parachute and mainly move vertically, focusing on destroying the four ventilation buildings. They do not reach later stages where the mission calls for navigation on foot inside buildings.
Cluster 2 (17%)	These players prefer moving on foot, and navigate vertically similar to Cluster 1, but horizontally in a much smaller area. They showcase a variety of action types such as killing of guards and ninjas that hide in the buildings as they navigate the base. They rarely use the parachute after leaving the helicopter. Their trajectories stay on the top left corner because they generally do not reach the final component of the mission which takes place down the shallows and lake (low point Fig. 5).
Cluster 3 (6%)	These players cover the most distance and use a variety of actions, abilities, and have the highest overall mission completion rates. They use the parachute to move up and down the mountain, eliminate guards and ninja on foot, and follow the escaping convoy in the last phase of the mission.

Table 4: Behavioral Profiles for the *Mountain Rescue* mission in *Just Cause 2*

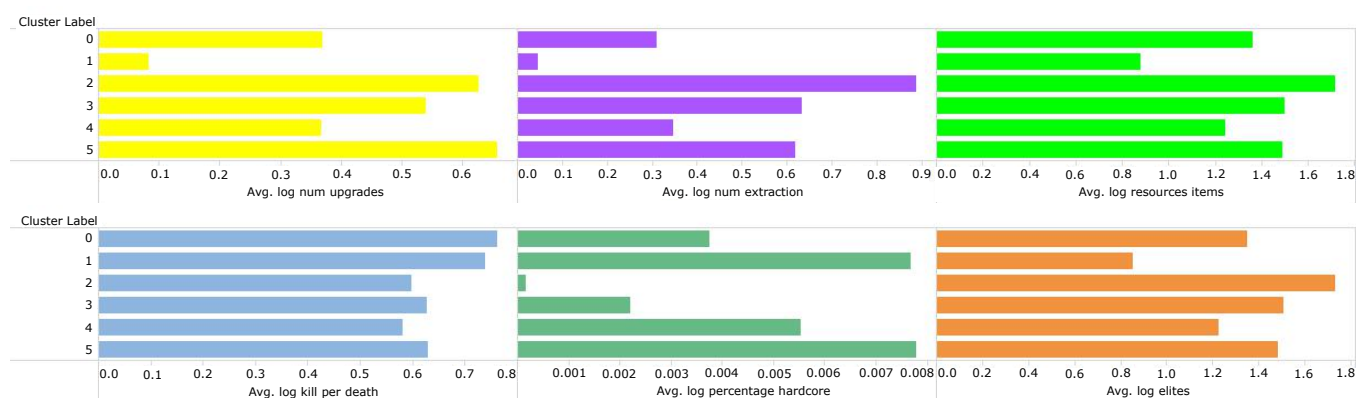


Figure 7: Summary statistics for the Early Dropouts in *Just Cause 2*. Note the log-scale of the X-axis. Metrics explained in Table 1, profile descriptions in Table 2

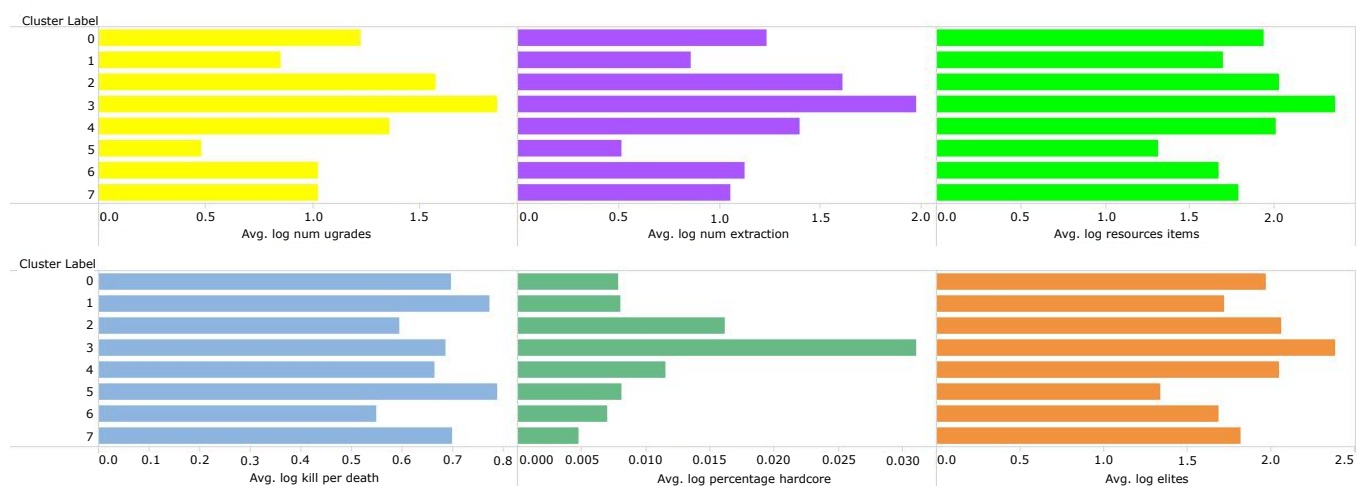


Figure 8: Summary statistics for the Committed Players in *Just Cause 2*. Note the log-scale of the X-axis. Metrics explained in Table 1, profile descriptions in Table 3

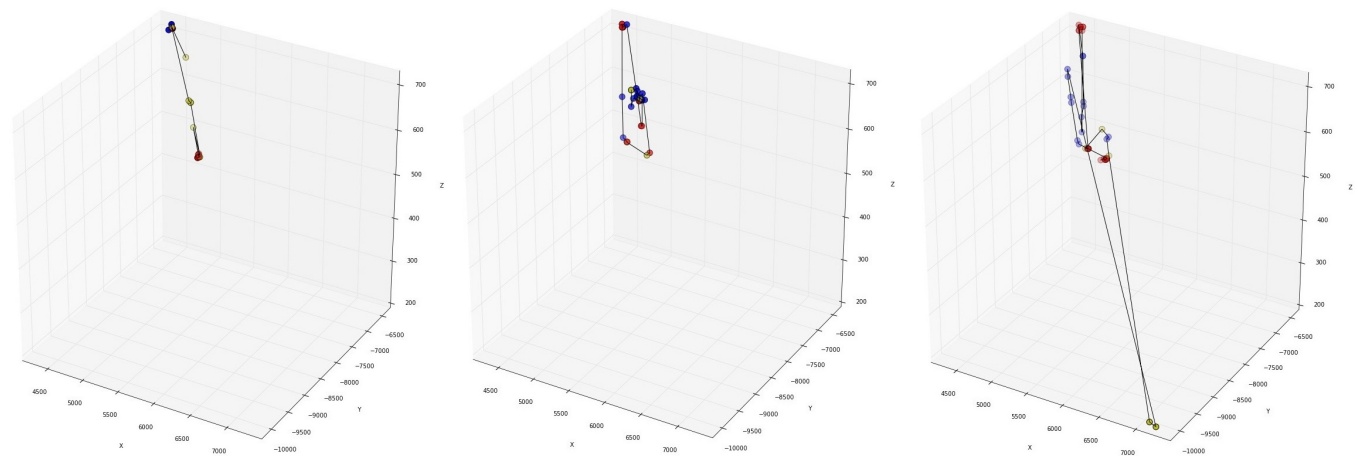


Figure 9: Typical DEDICOM-based player trajectories for cluster 1 (left), 2 (middle) and 3 (right) for the *Mountain Rescue* mission in *Just Cause 2*.

of abandonment. While based only on a single case example, the result indicates that spatio-temporal profiling using DEDICOM is applicable for detailed granular level of analysis where players are provided with an open range of opportunities to exhibit highly varied behaviors. With even more detailed information than available in the current dataset, corresponding more detailed trajectories would be possible.

7 DISCUSSION AND CONCLUSION

In this paper, a technique has been presented for condensing varied, voluminous behavioral telemetry data from OWGs into distinct profiles, that describe patterns in the behavior of the players of these types of games, and which notably takes into account the spatio-temporal dimensions of the playing activity. Based on the DEDICOM framework by Bauckhage et al. [3], we detailed behavioral data for over 5,000 players of *Just Cause 2*, translating them into a handful of distinct profiles. DEDICOM is applied to complete play histories of players, in addition to a smaller segment of play. In both cases the technique provides separated profiles that can be interpreted based on the constituent behavior and indicated potential reasons for early game abandonment. The approach advances the state-of-the-art in player profiling by incorporating the spatio-temporal behavior of players directly [5, 12, 13, 21, 23, 37]. One core objective of this paper was to investigate behavioral differences between different player types defined by early abandonment and commitment. We found that committed players were more able to harness the complete range of mechanics available to them over the course of play, as well as progress through a single mission further. In contrast, it can be observed that a significant majority of the Early Dropouts selected a difficulty higher than most other players, and subsequently stopped playing very early into the game. The ability to detect potential misunderstandings of players for the game's difficulty, especially within specific demographics could be a valuable application for game analytics. However, this would also require further unit testing and research to better understand how unsupervised cluster results may translate directly to consumers. Had we carried out our analysis without grouping players based on commitment, we likely would have missed the impact of difficulty on player commitment entirely. Additionally, we acknowledge potential concerns regarding the validity of our results. It is important to note that the success of our algorithm is not necessarily dependent on how we defined specific clusters or parameters, but rather how normally unused spatio-temporal dimensions can now be integrated into new analytical models. For example, our specific definition of what constitutes 'elite players' may possibly be invalid, but the same method we present here could be used for another definition using different parameters. Another goal of this study was to present the utility of spatio-temporal data to behavioral analytics, and we hope to have accomplished this by showing significantly varied movement between smaller subsamples of players and even within specific sections of play, such as for the *Mountain Rescue* mission. This ultimately expands the value of data already available to game developers and studios. Indeed, the best way to validate our results would be to develop alternative methods incorporating spatio-temporal user data, which prior to this study has been unexplored with real player telemetry.

The results of the work presented here may be specific to one game, but they provide a means to build profiles based on player activity in unrestricted game environments across different scales of granularity. While other OWGs may differ in their gameplay variables and user behaviors, the challenges that such data present are consistent. For any problem that contains a large range of spatial, temporal and varied data, the combinative methods used here are suitable. This paves the way for player profiling to be adapted to a variety of situations, with examples including detailed analysis of the performance and positional tactics of teams in esports [12, 13, 26], map balancing in first-person shooters, and composite navigational analyses in MMOGs [15, 19, 21, 37]. Finally, the approach could be applied to similar problems outside digital games, e.g. geographic consumer behavior analytics [36], esports analytics or any environment that contains valuable temporal-spatial data.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Square Enix for access to the *Just Cause 2* data. This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/L015846/1] and the Digital Creativity Labs (digitalcreativity.ac.uk), jointly funded by EPSRC/AHRC/Innovate UK under grant no. EP/M023265/1.

REFERENCES

- [1] S Aranganayagi and K Thangavel. 2007. Clustering categorical data using silhouette coefficient as a relocating measure. In *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, Vol. 2. IEEE, 13–17.
- [2] Christian Bauckhage, Anders Drachen, and Rafet Sifa. 2015. Clustering game behavior data. *IEEE Transactions on Computational Intelligence and AI in Games* 7, 3 (2015), 266–278.
- [3] Christian Bauckhage, Rafet Sifa, Anders Drachen, Christian Thureau, and Fabian Hadji. 2014. Beyond heatmaps: Spatio-temporal clustering using behavior-based partitioning of game levels. In *Computational Intelligence and Games (CIG), 2014 IEEE conference on*. IEEE, 1–8.
- [4] John Bohannon. 2010. Game-miners grapple with massive data.
- [5] Alessandro Canossa, Anders Drachen, and Janus Rau Møller Sørensen. 2011. Arrrrghh!!!: blending quantitative and qualitative methods to detect player frustration. In *Proceedings of the 6th international conference on foundations of digital games*. ACM, 61–68.
- [6] Peter A Chew, Brett W Bader, and Alla Rozovskaya. 2009. Using DEDICOM for completely unsupervised part-of-speech tagging. In *Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*. Association for Computational Linguistics, 54–62.
- [7] Olivier Delalleau, Emile Contal, Eric Thibodeau-Laufer, Raul Chandias Ferrari, Yoshua Bengio, and Frank Zhang. 2012. Beyond skill rating: Advanced matchmaking in ghost recon online. *IEEE Transactions on Computational Intelligence and AI in Games* 4, 3 (2012), 167–177.
- [8] Anders Drachen, Alessandro Canossa, and Georgios N Yannakakis. 2009. Player modeling using self-organization in Tomb Raider: Underworld. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. IEEE, 1–8.
- [9] Anders Drachen and Matthias Schubert. 2013. Spatial game analytics. In *Game Analytics*. Springer, 365–402.
- [10] Anders Drachen, Rafet Sifa, Christian Bauckhage, and Christian Thureau. 2012. Guns, swords and data: Clustering of player behavior in computer games in the wild. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*. IEEE, 163–170.
- [11] Anders Drachen, Christian Thureau, Rafet Sifa, and Christian Bauckhage. 2014. A comparison of methods for player clustering via behavioral telemetry. *arXiv preprint arXiv:1407.3950* (2014).
- [12] Anders Drachen, Matthew Yancey, John Maguire, Derrek Chu, Iris Yuhui Wang, Tobias Mahlmann, Matthias Schubert, and Diego Klabajan. 2014. Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2). In *2014 IEEE Games Media Entertainment*. IEEE, 1–8.
- [13] Christoph Eggert, Marc Herrlich, Jan Smeddinck, and Rainer Malaka. 2015. Classification of player roles in the team-based multi-player game dota 2. In *International Conference on Entertainment Computing*. Springer, 112–125.

- [14] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2016. *Game analytics*. Springer.
- [15] Wu-chang Feng, David Brandt, and Debanjan Saha. 2007. A long-term study of a popular MMORPG. In *Proceedings of the 6th ACM SIGCOMM Workshop on Network and System Support for Games*. ACM, 19–24.
- [16] Tim Fields and Brandon Cotton. 2011. *Social game design: Monetization methods and mechanics*. CRC Press.
- [17] Richard A Harshman. 1978. Models for analysis of asymmetrical relationships among N objects or stimuli. In *First Joint Meeting of the Psychometric Society and the Society of Mathematical Psychology*. Hamilton, Ontario, 1978.
- [18] Humphreys G. Hoobler, N. and M. Agrawala. 2004. Visualizing competitive behaviors in multi-user virtual environments. In *Proceedings of IEEE Visualization*. 163–170.
- [19] Jun H Kim, Daniel V Gunn, Eric Schuh, Bruce Phillips, Randy J Pagulayan, and Dennis Wixon. 2008. Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 443–452.
- [20] Larry Mellon. 2009. Applying metrics driven development to MMO costs and risks. *Versant Corporation* (2009).
- [21] John L Miller and Jon Crowcroft. 2010. Group movement in world of warcraft battlegrounds. *International Journal of Advanced Media and Communication* 4, 4 (2010), 387–404.
- [22] Pejman Mirza-Babaei, Günter Wallner, Graham McAllister, and Lennart E Nacke. 2014. Unified visualization of quantitative and qualitative playtesting data. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1363–1368.
- [23] Truong-Huy D Nguyen, Magy Seif El-Nasr, and Alessandro Canossa. 2015. Glyph: Visualization Tool for Understanding Problem Solving Strategies in Puzzle Games.. In *FDG*.
- [24] Dmitry Nozhnin. 2012. Predicting churn: When do veterans quit. *Gamasutra*, August 30th (2012).
- [25] Johanna Pirker, Simone Griesmayr, Anders Drachen, and Rafet Sifa. 2016. How playstyles evolve: progression analysis and profiling in just cause 2. In *International Conference on Entertainment Computing*. Springer, 90–101.
- [26] François Rioult, Jean-Philippe Métivier, Boris Helleu, Nicolas Scelles, and Christophe Durand. 2014. Mining tracks of competitive video games. *AASRI Procedia* 8 (2014), 82–87.
- [27] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.
- [28] Rafet Sifa, Anders Drachen, and Christian Bauckhage. 2018. Profiling in Games: Understanding Behavior from Telemetry. *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective* (2018).
- [29] Rafet Sifa, Anders Drachen, Christian Bauckhage, Christian Thureau, and Alessandro Canossa. 2013. Behavior evolution in tomb raider underworld. In *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. IEEE, 1–8.
- [30] Rafet Sifa, Anders Drachen, and Christian Bauckhage. 2018. Profiling in Games: Understanding Behavior from Telemetry. In *Social Interaction in Virtual Worlds*. 337–374.
- [31] Rafet Sifa, César Ojeda, and Christian Bauckhage. 2015. User churn migration analysis with DEDICOM. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 321–324.
- [32] Rafet Sifa, Sridev Srikanth, Anders Drachen, Cesar Ojeda, and Christian Bauckhage. 2016. Predicting retention in sandbox games with tensor factorization-based representation learning. In *Computational Intelligence and Games (CIG), 2016 IEEE Conference on*. IEEE, 1–8.
- [33] Christian Thureau, Kristian Kersting, and Christian Bauckhage. 2009. Convex non-negative matrix factorization in the wild. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*. IEEE, 523–532.
- [34] Robert Tibshirani, Guenther Walther, and Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 2 (2001), 411–423.
- [35] Giel Van Lankveld, Pieter Spronck, Jaap Van den Herik, and Arnoud Arntz. 2011. Games as personality profiling tools. In *Computational Intelligence and Games (CIG), 2011 IEEE Conference on*. IEEE, 197–202.
- [36] Marin Vukovic, Ignac Lovrek, and Hrvoje Kraljevic. 2012. Discovering shoppers' journey in retail environment by using RFID. In *KES*. 857–866.
- [37] Günter Wallner and Simone Kriglstein. 2014. PLATO: A visual analytics system for gameplay data. *Computers & Graphics* 38 (2014), 341–356.
- [38] Georgios N Yannakakis. 2012. Game AI revisited. In *Proceedings of the 9th conference on Computing Frontiers*. ACM, 285–292.
- [39] Georgios N Yannakakis and Julian Togelius. 2015. A panorama of artificial and computational intelligence in games. *IEEE Transactions on Computational Intelligence and AI in Games* 7, 4 (2015), 317–335.