Do they Play as Intended? – Comparing Aggregated and Temporal Behavioral Analysis in a Persuasive Gamified System

Enrica Loria
Fondazione Bruno Kessler
Trento, Italy
eloria@fbk.eu

Jessica Rivera-Villicana
Deakin University
Victoria, Australia
jessicarivil@gmail.com

Annapaola Marconi
Fondazione Bruno Kessler
Trento, Italy
marconi@fbk.eu

Abstract

Gamified systems nurture an ulterior goal set by their designers (e.g., a positive behavioral change). Behavioral profiling allows understanding whether users play as intended and reach such a goal. Analyzing in-game behaviors can also highlight unexpected interaction patterns or unengaged users. Current logging systems can track and store any in-game action. However, such high-dimensional data should be carefully processed to retain relevant knowledge while filtering unnecessary noise. Analysts can either aggregate data into a single data point per player or maintain temporal information. This study compares aggregated and temporal behavioral analysis conducted on a gamified system, promoting sustainable mobility (Play&Go). Results show how, in Play&Go, aggregated analysis conveys information on long-term winning strategies, whereas temporal analysis describes short-term strategies. Additionally, studying the temporal evolution of players' behaviors emphasizes a sharp division among engaged and unengaged users. We show how aggregated and temporal analysis hold a complementary view of players' experiences.

1. Introduction

Gamified systems, in contrast to traditional games, are characterized by an ulterior motive such as educating users, performing crowdsourcing activities, or promoting a positive behavioral change [1, 2]. Despite the wide consensus gamification found and existent successful examples, gamification can also have a neutral, or even negative, effect on players [3]. This variability can be explained by practitioners lacking specific guidelines to achieve their goals. Instead, they rely on theoretical and empirical knowledge. As a result, designing gameful systems is far from being a linear process, as external factors such as unexpected player behavior may affect the system's success. Hence, the design process should be iterative and allow the system's supervision and modification as it progresses [4, 5]. This raises the need to obtain knowledge regarding players' behaviors. Monitoring users interacting with the system can help (1) detect and fix design faults (e.g., is the gamification ulterior motive being pursued?); (2) moderate abnormal behaviors (e.g., cheaters); and (3) characterize the population to personalize experiences.

Players' datalogs can be analyzed towards this end, since they are more easily accessible by game masters (e.g., designers or domain experts) than direct interactions with players such as interviews, focus groups, or extensive observational studies. Datalogs, however, are often noisy and not wholly reliable to make inferences regarding player behavior [6]. Success depends on using the appropriate methods [7, 8]. Game analytics and telemetry tracking represent a considerable component of the process of analyzing player behavior. The outcomes of behavioral analysis can be exploited throughout the game's life cycle to evaluate players' experiences, ensure fairness, and perform predictive analysis [9]. However, when the goal is to produce actionable design insights, an interpretable visualization of the analysis outcomes is needed [9].

In many works on behavioral analysis [8, 5], players' profiles were extracted from aggregated game datalogs. Nevertheless, the dynamic nature of gameplay must be acknowledged as well [10]. Players tend to change their in-game behavior during their gameplay [9], irrespectively of the game platform or genre. Therefore, such behaviors need to be continuously monitored as the gameplay advances [10]. On the other hand, high-dimensional temporal data increases the likelihood of having redundant, misleading, and noisy data, which can risk obscuring the more relevant information [6]. It follows that, although temporal information does improve the power of predictive models [11, 12, 13] and players do vary their behavior throughout their gameplay [9, 14], it is still obscure whether designers can benefit from the analysis of such temporal data to understand how players engage with the game.

Motivated by this discussion, in this work, we

investigate the value of performing temporal analysis, which can be costly to deploy and maintain in the development cycle, against aggregated data analysis to produce actionable outcomes for designers. In this context, we define temporal analysis as using many data points for each player, collected over time at regular intervals. In contrast, in the aggregated analysis, we use one single data point per player, averaging all the activity recorded within the observation period. Despite the literature reporting applications of both approaches, there is yet, to the best of our knowledge, no direct comparison between them in the same gamified system. We performed Archetypal Analysis, known to produce more interpretable results than other clustering methods [10], on both aggregated and temporal telemetry data to answer the following question:

How does using aggregated or temporal analysis of telemetry data affect the understanding of players' in-game behaviors?

By answering this research question, we verify whether each type of analysis can convey different information on players' gaming experience. This work's contribution is twofold. First, the systematical comparison of the two approaches to data analysis shows how short-term and long-term winning strategies can differ. Both strategies, however, provide information on whether players' behavior is in alignment with the gamification goal. Second, we show that temporal data analysis emphasizes behavioral shifts and may provide additional knowledge on peculiar and unengaged players.

2. Background and Related Works

In the past decades, the use of gamification has gained popularity in different domains to achieve non-game-related goals via game mechanics [3]. However, gamification is not always successful, as its effects highly depend on its context and implementation [15]. Hence, research has focused on improving the design of such systems to maximize their success rate [16]. A gamified system's success is dictated by whether its goal is being pursued by players while identifying and minimizing the occurrence of negative effects such as churn [3].

Obtaining direct feedback from players to assess a gameful system's success is not always possible. However, modeling their behavior using data collected from their interaction with the system is a way for designers to infer whether players are pursuing the intended goal. Each player modeling approach involves different challenges to obtain meaningful information

about player behavior [6, 5].

The most popular technique used to profile players is clustering, which consists of finding groups of samples with similar characteristics in a dataset [17, 18, 19]. Since clusters can be inspected and interpreted to produce a human-readable description, the information retrieved can be exploited to optimize the game's design [10]. In recent Games User Research works [20, 21, 22], researchers have extensively used Archetypal Analysis, a soft-clustering method that, instead of assigning each data point to a specific cluster, calculates a likelihood or belongingness score of that element to be in each cluster. This notion supports previous findings stating that dominant types, represented by clusters, cannot fully describe an individual. Instead, players are a result of a combination of traits or factors [23].

While several studies have been conducted on aggregated telemetry data (e.g., averaging data for the whole duration of an experiment), a temporal view is crucial to understand dynamic interactions better [24]. Previous work has found that players can belong to different clusters in different stages of a game [14, 25], adding further support to the belief that player profiles need to be iteratively updated for them to remain useful [5]. Understanding the temporal aspects of gameplay is also essential in analyzing how engagement and frustration may vary over time through player behavior [26]. However, in most cases, temporal data is used to enhance the accuracy of prediction of specific events or churn prediction [27, 28]. Despite those models often being very accurate compared to aggregated models, they are not human readable. Thus, it is impossible for designers to be meaningfully informed by their outcomes.

In summary, while recent research suggests that temporal telemetry analysis can provide better results than aggregated data, those results generally refer to prediction models. However, there is no evidence that temporal analysis holds different insights than aggregated analysis for a gamified system's general design.

3. Use Case Scenario

In this section, we present the gamified system used to conduct the analyses: Play&Go. The choice of using Play&Go data was motivated by the following: Play&Go meets the description of a gamification system, as an ulterior goal is fostered through symbolic rewards while aiming to make behavioral change fun and entertaining. The gamified application is in continuous development, so insights from player data analysis can improve its future versions. Besides,

the app is available on Mobile App Stores free to download. Thus, there are no restrictions or selection criteria for participants, except for them playing within a specific geographic area. Moreover, participants gave their consent for using their data towards analysis and improvement. Finally, some of the authors participated in the design of P&G. Therefore, our domain knowledge of the platform and the environment, besides the opportunity to analyze anonymized data collected in the past are an advantage compared to using other games as a use case scenario.

While the summary provided is rather brief and outlines only information relevant to this paper, the gamified system is documented in detail in existing publications [29, 30].

3.1. Play&Go

Play&Go [29, 30] is a gameful system active in the Smart City of Trento (Italy) to elicit a positive behavioral change in terms of sustainable mobility. The gamified application is embedded in the city's system, sponsored by the municipality. Thus, the game is accessible to every citizen in possession of a smartphone. At the time of data collection for this study, players were well distributed over their age range: 22% between 16 and 20 years old, 27% between 20 and 35 years old, 31% between 35 and 50 years old, 18% between 50 and 70 years old, and 2% over 70 years old. Almost half of the players (51%) identified as females, and the remaining (49%) as males, none expressed other preferences. Citizens can participate in 6-month campaigns, recurring every year since 2016.

Players can track their movement within the city by specifying the transportation means they used: bike, bus, train, or walking. The validity of each trip is determined through automatic recognition algorithms. Each valid trip awards players 'Green Leaf' points, which vary according to the route's length and the environmental sustainability of the transportation means used. The gameplay is structured in weeks: the players' week performance is used to build the weekly leaderboard. Players in the highest position are awarded physical prizes. Alongside the weekly ranking, there is also a global ranking, used at the end of the campaign to determine the winners of special prizes. The concept of weeks is recurring, in that players engage in weekly challenges, which can either be single-player or multiplayer (2 players). Single-player challenges can be of three levels of difficulty (easy, medium, and hard), computed upon players' history and performances. A single-player challenge is characterized by a target (e.g., 5 Km) and the transportation means to be used

(e.g., Walk). General challenges also exist, where the objective is expressed in the number of Green Leaf points that have to be gained. If the player wins the challenge before the end of the week, additional points are awarded.

Play&Go also foresees unlockable features and mechanics, as players level up. For instance, while single-player challenges are automatically assigned to novices, as they reach level 2, players can select their challenge from a pool of 2 options. The options vary in difficulty or type - i.e., transportation means. Upon reaching level 3, the size of the pool of choices increases to 3. Multiplayer challenges can be of three kinds: competitive with a fixed target, where the winner is whoever reaches the goal first; competitive on players performance, where the winner is the player who achieved more by the end of the week; cooperative, where both players contribute towards reaching a common goal. Multiplayer challenges, which involve two players, are made available from level 4 to level 6. Upon reaching every milestone, players can decide which type of multiplayer challenge to unlock first. Players can engage in multiplayer challenges through an invitation system, in which they select the type of multiplayer challenge, the type of transportation means, and the opponent. The system automatically computes the target and the prize according to the players' performance and the estimated difficulty. It follows that the prize may vary between the two players, since the challenge may be easier for one of them. When an invite of a challenge is sent, the receiver can either accept or refuse it. In case of acceptance, the invitation is successful, and the challenge is assigned to both players. Each week players have two challenges assigned. The first is a single-player challenge assigned to novices and can be chosen by more experienced players, as mentioned above. The second challenge is either another single-player challenge or a multiplayer challenge, in case the players unlocked the multiplayer mode and were involved in a successful invitation. To prevent users from inviting them to a challenge, players can add them to their blacklist, and, eventually, decide to remove them in the future. Players need to complete the choice of the two challenges, called the programming phase, by each Friday for challenges to be active on the next week (Saturday to Friday). When players can choose between challenges but fail to do so, default challenges are assigned.

3.2. Dataset

We analyzed data from the gameplay of Play&Go [29, 30] between November the 2nd,
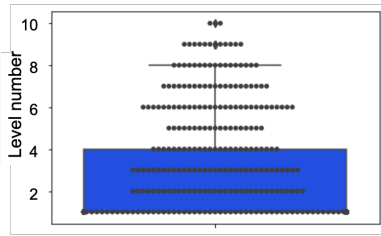
Figure 1: Box plot representing the distribution of the final level players achieved. The plot shows how players achieving a level higher than Level 8 are outliers of the population.

2019 and February the 28th, 2020, for a total of 17 weeks. We counted 425 players who performed at least one game action (other than the registration). The number of players retained in the gamified application for longer than one week was 248. The number drops to 194 when considering players that reached at least level 2 of the game. For this study, we examined a subset of those players whose maximum level was between 3 and 8. This sample resulted in 145 players. The rationale is that we wanted to include players that had the multiplayer mode available, which is unlocked at level 3. We also wanted to prevent the results from being influenced by outliers, which, as shown in Figure 1, were players having a final level higher than 8.

The selected 145 players, on average, participated in 11 game weeks (std = 5.2), over the 17 analyzed (Figure 2). The distribution of the level of activity (Figure 2), measured in the number of actions performed, is skewed towards zero with a long right tail (min = 6, mean = 166.5, std = 159, max = 844). The distribution of the Green Leaf points (Figure 2) achieved per week follows a similar trend, but the skewness and the tail are much less pronounced (min = 139, mean = 561, std = 271, max = 1.3k). Nevertheless, the distribution of level presents a low variability (mean = 5, std = 1.7). Ranking position is well distributed among players (min = 5, 25% = 43, 50% = 70, 75% = 95, max = 122). The features used to identify the archetypes are the following: points obtained, number of trips and kilometers per transportation modality, the total number of game actions performed, number of blacklist actions, the average difficulty of the individual challenges, number of single-player and multiplayer challenges, information on the types and amount of invites sent and received, percentage of accepted invites sent and received, percentage of individual and multiplayer challenges won, and percentage of challenges programmed (in contrast to being automatically assigned).

4. Modeling Player Behaviors

The following section presents the analyses performed on gameplay data to characterize players, both used as aggregated and temporal data.

We employ Archetypal Analysis (AA) (See [31] for more details) to produce clusters. AA is a soft clustering method, meaning that it produces a description of each data point as a combination of archetypes rather than as part of a single cluster. AA finds the extreme points representing the 'archetypes' in the dataset. As a result, each data point - i.e., each player - is represented through a series of belongingness scores, one for each archetype, showing how representative each archetype is of the player's behavior. Thus, instead of profiling each player by assigning them a single cluster, with AA players are modeled through a combination of archetypes, moving towards a trait-oriented approach.

Formally, AA is a matrix factorization technique that produces compact and interpretable data representations via the derived archetypes. Data points are described in a coefficient matrix, indicating the data points' belongingness ratios to each archetype.

Drachen et al. discuss how AA is more adequate to produce interpretable representations of player types than other widely used clustering algorithms, such as k-means clustering or PCA [10].

In the aggregated dataset, each entry represents a player, while the temporal dataset is divided into weeks - a concept highly tied to the game itself. Each row represents a player behavior within each week of their gameplay. The dataset contains players that have a level between 3 and 8. The features used are the same in both cases, as described in the previous sections.

Following the methodology of related research [20, 21, 22], we present an interpretation of the clusters (or archetypes) identified by the algorithms. The gamified system designers have also been consulted to validate the cluster interpretation exploiting their domain knowledge.

4.1. Aggregated Analysis

As typically done in clustering analysis, we selected the number of clusters using the 'elbow' method. It consists of running the algorithm with different values for k and plotting each of these values against a performance metric to identify an 'elbow' in the curve visually. The metric's value does not change drastically beyond this point. Figure 3 displays two elbows: for 2 and 6 archetypes. Since the use of 2 archetypes results in groups of active and non-active players, we selected 6 archetypes. We refer to this set as $\mathcal{A}$. A description of
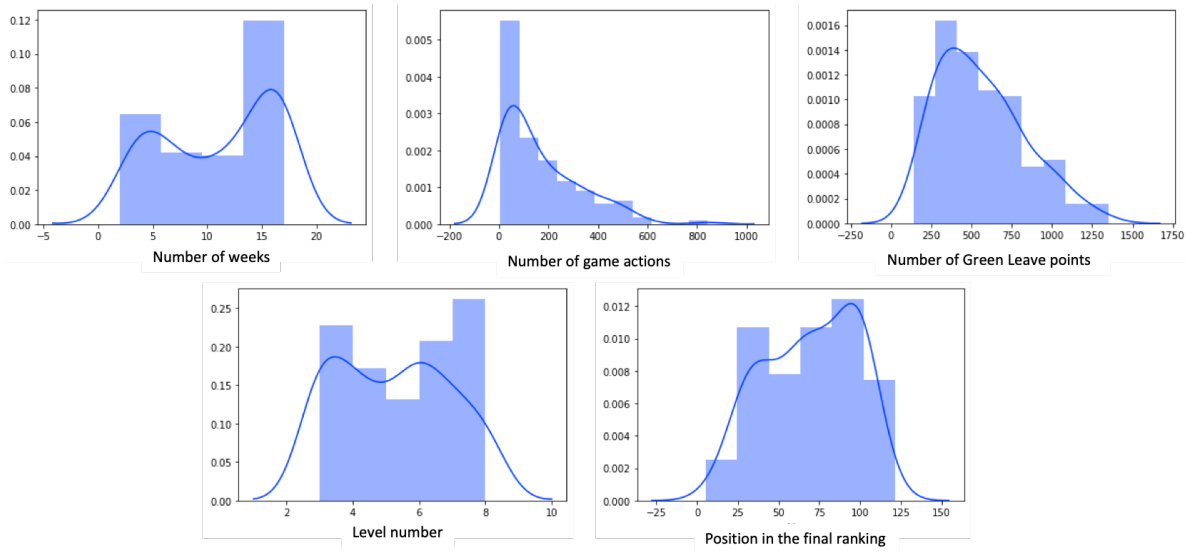
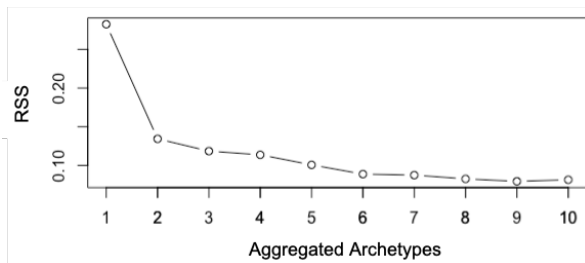Figure 2: Activity distribution in the selected sample of users .

Figure 3: RSS using a different number of archetypes for the aggregated analysis.

the archetypes follows.

A1: *Ghost*. Their level of activity is almost nonexistent.

A2: *Hostile*. They are multimodal travelers, with no evident preference. They tend to program a few challenges, with a medium difficulty, which are rarely won. They show an anomalously high number of actions on the blacklist. They also received a few cooperative invites, a low percentage of those were accepted. The multiplayer challenges were never won.

A3: *Loser*. They are quite active in terms of the number of actions performed. They are multimodal, with a high preference towards the bus and walking, and have tried both the challenge programming mechanism and the invitation systems on a small scale. Their single-player challenges are tendentially easy but are never won.

A4: *Average Full-feature*. They are quite active, and they intensively use all the available transport modes. They sometimes program their challenges, which are between a medium and a hard difficulty level. They do not always win them. They are also active in the invitation systems, despite being receivers rather than senders of challenge invites. They prefer cooperation to competition, but the percentage of won challenges is low.

A5: *Sporadic User*. They show a low level of activity and little to none interactions with other users (challenge invites). They mostly walk or use the bus. They programmed a few single-player challenges, with medium difficulty, and won them almost half of the time.

A6: *Green Socializer*. They are quite active in the game, with a strong preference for green transportation means walking and biking. They tend to program single-player challenges, usually between a medium and hard difficulty level and won most of them. They tend to send invites rather than receiving them. They prefer cooperation to competition.

The top of Figure 4 shows the distribution of the belongingness scores for each aggregated archetype over the 145 players. We can observe a predominance of the *Ghost* archetype, followed by the *Sporadic User* archetype. We can also observe that the *Hostile* archetype, in comparison to the others, is scarcely populated. Finally, Figure 4 shows how the more active archetypes present similar distributions.

Aggregated Archetypes ($\mathcal{A}$)
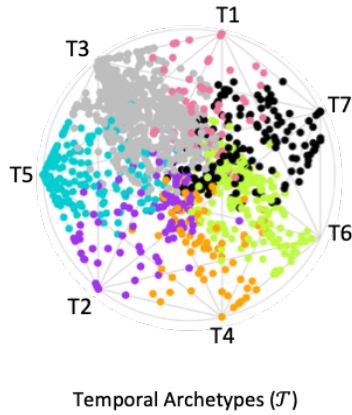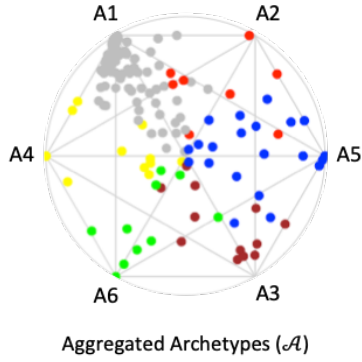

Temporal Archetypes ($\mathcal{T}$)

Figure 4: Scatter plot of the belongingness scores of the aggregated and temporal archetypes. The vertices represent the archetypes, while the dots represent the players. The closer the dot is to a vertex the higher the value for that archetype. The color of the dot is determined by the dominant archetype value.

4.2. Temporal Analysis

Temporal archetypes describe in-game behaviors occurring within a week. They were built using a dataset where each entry represents the players' weekly activity. To avoid introducing data dependency resulting from having several weeks for each player, we initially used a subset of one week of data with one entry for each player to identify the number of archetypes. In particular, we considered the week with the highest number of active players (to have similar environmental conditions). We chose week 5 with 104 players. Then, we computed the belongingness scores for every excluded entry according to the definition of the archetypes found.

As with the aggregated analysis (Figure 5), we found two elbows, corresponding to 2 and 7 archetypes. We choose 7 archetypes for the same reason. We refer to
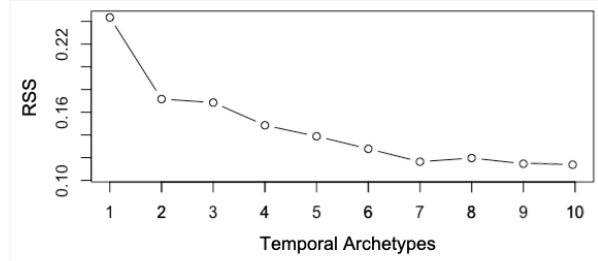

Figure 5: RSS using a different number of archetypes for the temporal analysis.

this set as $\mathcal{T}$. A description of the archetypes follows.

T1: *Wannabe Competitive Achiever.* They are quite active multimodal players, with a preference for green transportation (bike and walk). They always program hard challenges. Challenges are won half of the time. They also engage in multiplayer challenges. Although they send competitive invites, they receive only cooperative invites, which are always accepted.

T2: *Sporadic Social User.* Their activity level is shallow. They sometimes program challenges, but they usually lose them. They receive few invites (both cooperative and competitive), which are rarely accepted. Their win rate for multiplayer challenges is low.

T3: *Ghost.* Their level of activity is almost nonexistent.

T4: *Average User.* They have a moderate level of activity. They are multimodal, with a preference for walking and bus. They program a few easy challenges, which are seldom won. They also receive few invites (both cooperative and competitive), rarely accepted. Their win rate of multiplayer challenges is meager.

T5: *Social Initiator.* They are moderately active multimodal players, with a preference for walking. They often program the challenges with medium-to-hard difficulty. They have a good win ratio. They are almost always involved in multiplayer challenges, where they are the initiators. They have a very high acceptance rate to multiplayer challenges, although the win ratio is medium to low.

T6: *Green Loner.* Those players are moderately active. They have a strong tendency towards green transportation means. They tend to program medium to hard challenges, which are always won. They have no interaction with the invitation system.
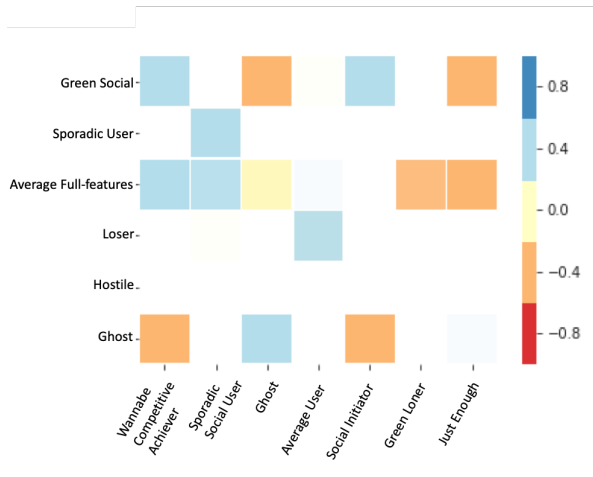
Figure 6: Heatmap of the correlation among the two sets of archetypes. The image overlays all the images of the 17 analyses. White cells represent values non statistically significant (p-value>.05). The p-values are corrected through the Benjamini-Hochberg method.

T7: *Just Enough*. Those players have a low activity level. They mostly walk or use the bus. They seldom program challenges, which are usually easy. However, their win rate is perfect. They have no interaction with the invitation system.

We also computed the belongingness score for the entries excluded (all the entries available per player). This led to a sample of 1590 entries for 145 players.

The bottom of Figure 4 shows the belongingness scores distribution for each temporal archetype over the 145 players and the 1590 entries. As in the aggregated archetypes, we can observe a predominance of the *Ghost* archetype. However, in contrast to the aggregated set, the more active archetypes are visibly more populated. We can also clearly see that the least dense archetype is the *Wannabe Competitive Achiever*, which represents a small slice of our population. Finally, Figure 4 shows that most of the data points are located in the middle of the figure, which means that a single archetype hardly characterizes most of the players.

4.3. Comparing Aggregated and Temporal Analysis

In this section we present a comparison of the sets $\mathcal{A} = \{A_1, A_2, A_3, A_5, A_6\}$, obtained with the aggregated analysis, and $\mathcal{T} = \{T_1, T_2, T_3, T_5, T_6, T_7\}$ with temporal analysis, to better understand the divergence of information they retrieve, if any. We did so by researching the correlations among the two groups

of archetypes. Since the scores that the users obtained were not normally distributed, we used Kendall's τ, being the best estimate for non-parametric data [32]. We used the Gilpin's correspondence table [33] to interpret Kendall's τ effect size according to the approximate Pearson's r equivalent. A value of $\tau = 0.20$ is considered a small effect, $\tau = 0.34$ is considered a medium effect, while $\tau = 0.50$ is considered a large effect.

We looked for correlations to understand whether and to what extent $\mathcal{A}$ and $\mathcal{T}$ were a different representation of the same concepts. To avoid dependency on the data, resulting from having multiple entries for the same player, we retrieved data from 1 week for the temporal data. We repeated the correlation analysis iterating on each week to rule out the possibility of the results being due to casualty. Then, we compared the outcome of the 17 - i.e., number of gameplay weeks - analyses. To treat multiple correlations, we used the Benjamini-Hochberg correction [34].

Figure 6[1] shows the resulting correlation heatmap. We observed how the low-activity archetypes in the two representations correlated (*Ghost* (A1, T3) and *Sporadic User*(A5, T2)). We also found how some specific archetypes existed in one representation and were completely absent in the other. For instance, the *Hostile* (A2) aggregated archetype was uncorrelated to the temporal archetypes, and the *Green Loner* (T6) and *Just Enough* (T7) temporal archetypes showed no positive correlations to the aggregated archetypes. Finally, some aggregated archetypes appeared to be, to some extent, the combination of two temporal archetypes: *Green Socializer* (A6) positively correlates to *Wannabe Competitive Achiever* (T1) and *Social Initiator* (T5). In contrast, *Average Full-feature* (A4) positively correlates to *Wannabe Competitive Achiever* (T1) and *Sporadic Social User* (T2). Therefore, although both sets of archetypes are related, they also provide a specific view of some aspects of the gameplay, which we will further investigate in the following sections.

5. Winning (or Losing) Behaviors

This section identifies emergent winning strategies by analyzing the best performing players' archetypes in contrast to the worst-performing players. The purpose of this is twofold. Firstly, we want to compare the outcomes of aggregated and temporal analysis. Secondly, we want to discover whether evident unbalances in the game mechanics exist by studying how winners behaved. Figure 7 shows how The belongingness scores of player archetypes correlate to

---

[1]The colors of the heatmap scale have been chosen to be color-blind friendly for accessibility reasons.
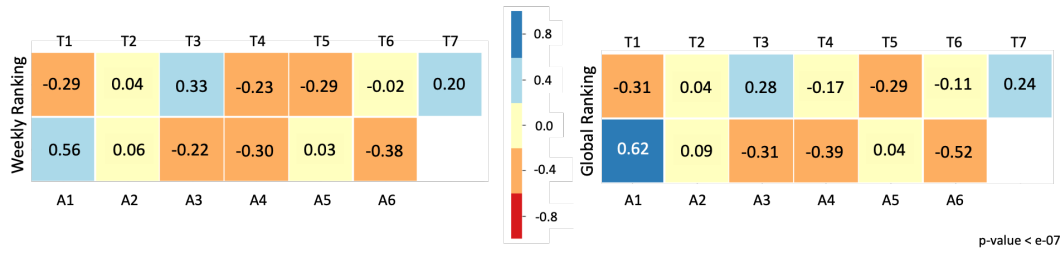
Figure 7: Heatmap of the correlation among the set of archetypes and the position in the weekly and global rankings. White cells represent values non statistically significant (p-value > .05).
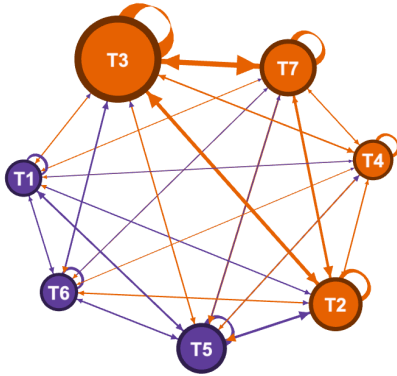
Figure 8: Graph of transitions among archetypes.

of recurrent transitions among archetypes. In other words, we studied whether some behavior transitions are more frequent than others to unravel more insights on how players interact with the gamified systems.

To model the transitions, we used a directed weighted graph (Figure 8), where the nodes are the temporal archetypes, and the edges represent a transition from an archetype X to an archetype Y. The weight of each edge represents how many times that transition occurred among players. In the graph presented in Figure 8, the size of the nodes is proportional to their weighted in-degree value, and the thickness of the edges increases with its weight. The nodes' colors show how the graph is divided into two communities, computed according to the strength of the ties they are involved in (modularity). From the division in communities, we observed how the archetypes were divided into two communities: one composed by T1 (*Wannabe Competitive Achiever*), T6 (*Green Loner*), and T5 (*Social Initiator*), and the other by T3 (*Ghost*), T7 (*Just Enough*), T4 (*Average User*), and T2 (*Sporadic Social User*) in the other. This suggests that, in most cases, players tend to maintain their level of engagement throughout the game. In other words, the game succeeds in keeping engaged the portion of interested users, while it lacks mechanisms to increase the participation of users not particularly invested in it.

7. Results and Discussion

By comparing the archetype representations of aggregated ($\mathcal{A}$) and temporal ($\mathcal{T}$) player data through correlation analysis, we found an overlap for what concerns low-activity archetypes. Despite this overlap, aggregated and temporal analysis also provided specific views on the most active users. For instance, temporal archetype analysis produced the *Wannabe Competitive Achiever* (T1), showing how some players try to engage in multiplayer challenges but have difficulty finding a match. Whereas, from the aggregated analysis, it

their position in both the weekly (on the left) and the global (on the right) leaderboards.

In the aggregated analysis, we found that the *Average full-feature* (A4) and the *Green Socializer* (A6) archetypes were higher in the weekly rankings (with a correlation of -0.3 and -0.38 respectively). However, when looking at the global ranking, it is evident how players with a higher belongingness score for *Green Socializer* (A6) dominated the leaderboard (-0.52).

In the temporal analysis, we found that the *Social Initiator* (T5) and the *Green Loner* (T1) archetypes were higher in the weekly ranking (both with a correlation of -0.29). The correlations remain similar when analyzing the position in the global ranking.

In summary, aggregated archetypes gave a clearer view of the successful behaviors in the long term, which brought players to the top of the global leaderboard.

6. Transitions Between Archetypes

In this section, we investigate how players' behavior varies throughout their gameplay. For simplicity, we consider dominant archetypes for each player each week, as we are interested in researching the existence

appears that most of the population dislikes competitive challenges.

Then, to compare the strategies that are rewarding in the short term (within each week) and the strategies useful in the long run (Figure 7), we used both the weekly and the global leaderboards. We found that, in both sets of archetypes ($\mathcal{A}$ and $\mathcal{T}$), the position on the leaderboard was related to the users' belongingness score in the low and high activity archetypes. Players with higher belongingness scores in high-activity archetypes assumed a higher position in the rankings and vice-versa. In the aggregated archetypes, we also observed how A6 showed a very high positive correlation (.6) in the global leaderboard. None of the archetypes in the temporal analysis showed a similarly strong correlation in the weekly or the global leaderboard. This highlights how (1) the archetype A6 (*Green Socializer*) represents a good strategy in the long term, and (2) aggregated analysis provides a better view on which are the behaviors that bring the players to succeed in the long run. In this specific context, the outcome suggests how players focused on green transportation means and involved in both single and multiplayer challenges were rewarded more than others, which is desirable for a system promoting sustainable transportation behaviors.

We also investigated the transitions among archetypes ($\mathcal{T}$) to understand how players oscillated between them. This analysis outlined a sharp division between archetypes with low and high activity by applying community detection algorithms. This highlights a fault in the game, as it is unlikely for users to shift from low activity to a high activity level. Instead, a more consistent shift among archetypes indicating different activity levels would be desirable, as it would be acceptable for players to oscillate among less active periods. On the contrary, our analysis outlines a group of users who never fully engage in the game, maybe due to specific preferences not met in its design.

7.1. Limitations

The main limitation of this study lies in its sample size. While retrieving data from a real gamified system is a benefit, in contrast to a controlled study, the number of users eligible for the study - i.e., that played for long enough - is limited. In addition, although our observations are context-independent, the analyses are conducted on data obtained from a single gamified app. Hence, the generalizability of the findings to other application domains should be further investigated. Lastly, despite using a method that produces more interpretable results (AA) than

mainstream approaches (e.g., k-means), this process still needs a human-in-the-loop to analyze and interpret the results. Nevertheless, this step is necessary for interpretability's sake of the final results and potential feedback to improve its design.

8. Conclusions and Future Works

In this work, we compared two types of behavioral analysis using aggregated and temporal data in a gamified environment. Although previous studies found that using temporal data is advantageous in predictive analysis, descriptive and exploratory behavioral analysis still relies only on aggregated data. To the best of our knowledge, an actual comparison between the two approaches on the same data to obtain human-readable and actionable information has not been made.

Our findings show that both aggregated and temporal analyses are needed to understand the dynamics occurring within the system. The aggregated analysis provides a global overview and the strategies that give the highest rewards in the long term, bringing the players at the top of the ranking. Temporal analysis is useful for detecting short-term rewarding behaviors. The temporal analysis allows identifying abnormal behaviors representing players' specific needs as well. Therefore, employing one method rather than the other depends on whether the designer needs low-lever or high-level insights. The aggregated analysis may be enough to understand whether the ulterior motive is ultimately pursued. However, the temporal analysis provides low-level information and can highlight design faults that might be overlooked otherwise.

In future works, we will perform more thorough analyses on the transitions between archetypes, considering all belongingness scores instead of the dominant archetype only. Moreover, we will move towards a players-centric approach, going in-depth on individual gameplay to study and visualize each player's gameplay's evolution to assist designers further while detecting general design issues.

References

[1] K. Huotari and J. Hamari, "A definition for gamification: anchoring gamification in the service marketing literature," *Electronic Markets*, vol. 27, no. 1, pp. 21–31, 2017.

[2] H. Warmelink, J. Koivisto, I. Mayer, J. Hamari, and M. Vesa, "Gamification of the work floor: A literature review of gamifying production and logistics operations," in *Proceedings of the 51st Annual Hawaii International Conference on System Sciences*, 2018.

[3] J. Koivisto and J. Hamari, "The rise of motivational information systems: A review of gamification research,"

International Journal of Information Management, vol. 45, pp. 191–210, 2019.

[4] A. Alshamrani and A. Bahattab, "A comparison between three sdlc models waterfall model, spiral model, and incremental/iterative model," *International Journal of Computer Science Issues*, vol. 12, no. 1, p. 106, 2015.

[5] R. Sifa, A. Drachen, and C. Bauckhage, "Profiling in games: Understanding behavior from telemetry," *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective*, 2018.

[6] D. Hooshyar, M. Yousefi, and H. Lim, "Data-driven approaches to game player modeling: a systematic literature review," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–19, 2018.

[7] 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*, pp. 163–170, IEEE, 9 2012.

[8] 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 Computatonal Intelligence and Games*, vol. 0, IEEE, 7 2016.

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

[10] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, "A Comparison of Methods for Player Clustering via Behavioral Telemetry," 7 2014.

[11] B. G. Weber, M. Mateas, and A. Jhala, "Using Data Mining to Model Player Experience," in *FDG Workshop on Evaluating Player Experience in Games*, 2011.

[12] A. Saas, A. Guitart, and A. Perianez, "Discovering playing patterns: Time series clustering of free-to-play game data," in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 0, IEEE, 7 2016.

[13] Z. Yu, J. Lian, A. Mahmoody, G. Liu, and X. Xie, "Adaptive user modeling with long and short-term preferences for personalized recommendation," in *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2019-August, pp. 4213–4219, International Joint Conferences on Artificial Intelligence, 2019.

[14] J. Pirker, S. Griesmayr, A. Drachen, and R. Sifa, "How playstyles evolve: Progression analysis and profiling in Just Cause 2," in *Lecture Notes in Computer Science*, vol. 9926 LNCS, pp. 90–101, Springer Verlag, 2016.

[15] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work? – A literature review of empirical studies on gamification," in *2014 47th Hawaii International Conference on System Sciences*, pp. 3025–3034, 2014.

[16] E. Marcucci, V. Gatta, and M. Le Pira, "Gamification design to foster stakeholder engagement and behavior change: An application to urban freight transport," *Transportation Research Part A: Policy and Practice*, vol. 118, pp. 119–132, 2018.

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

[18] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering Game Behavior Data," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, pp. 266–278, 9 2015.

[19] L. A. L. Rodrigues and J. D. Brancher, "Improving Players' Profiles Clustering from Game Data Through Feature Extraction," in *2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, vol. 2018-Novem, pp. 177–17709, IEEE, 10 2018.

[20] E. S. Siqueira, C. D. Castanho, G. N. Rodrigues, and R. P. Jacobi, "A data analysis of player in world of warcraft using game data mining," in *Brazilian Symposium on Games and Digital Entertainment, SBGAMES*, vol. 2017-Novem, pp. 1–9, IEEE, 6 2018.

[21] M. H. Schiller, G. Wallner, C. Schinnerl, A. Monte Calvo, J. Pirker, R. Sifa, and A. Drachen, "Inside the Group: Investigating Social Structures in Player Groups and Their Influence on Activity," *IEEE Transactions on Games*, pp. 1–1, 7 2018.

[22] A. Rattinger, G. Wallner, A. Drachen, J. Pirker, and R. Sifa, "Integrating and inspecting combined behavioral profiling and social network models in Destiny," in *Lecture Notes in Computer Science*, vol. 9926 LNCS, pp. 77–89, Springer Verlag, 2016.

[23] E. Loria and A. Marconi, "Player types and player behaviors: Analyzing Correlations in an On-the-field Gamified System," in *CHI PLAY 2018*, (New York, New York, USA), pp. 531–538, ACM, 10 2018.

[24] R. Bakeman and J. M. Gottman, *Observing interaction: An introduction to sequential analysis*. Cambridge university press, 1997.

[25] J. Kang, M. Liu, and W. Qu, "Using gameplay data to examine learning behavior patterns in a serious game," *Computers in Human Behavior*, vol. 72, pp. 757–770, 7 2017.

[26] M. Soppitt and G. McAllister, "Understanding player experience using sequential analysis.," in *DiGRA Conference*, 2011.

[27] A. Saas, A. Guitart, and A. Periánez, "Discovering playing patterns: Time series clustering of free-to-play game data," in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8, 2016.

[28] J. L. Hsieh and C. T. Sun, "Building a player strategy model by analyzing replays of real-time strategy games," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 3106–3111, 2008.

[29] M. Ferron, E. Loria, A. Marconi, and P. Massa, "Play & go, an urban game promoting behaviour change for sustainable mobility," *Interaction Design and Architecture(s)*, no. 40, pp. 24 – 45, 2019.

[30] A. Marconi and E. Loria, "Sustainable Mobility in Smart Cities: The Key Role of Gamified Motivational Systems for Citizens' Engagement and Behavior Change," in *Implications of Mobility as a Service (MaaS) in Urban and Rural Environments*, pp. 211–246, IGI Global, 2020.

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

[32] D. C. Howell, *Statistical methods for psychology*. Cengage Learning, 2012.

[33] A. R. Gilpin, "Table for Conversion of Kendall'S Tau to Spearman'S Rho Within the Context of Measures of Magnitude of Effect for Meta-Analysis," *Educational and Psychological Measurement*, vol. 53, pp. 87–92, 3 1993.

[34] S.-Y. Chen, Z. Feng, and X. Yi, "A general introduction to adjustment for multiple comparisons," *Journal of thoracic disease*, vol. 9, no. 6, p. 1725, 2017.