UNIVERSITY OF TRENTO

DOCTORAL THESIS

Alone with Company: Studying Individual and Social Players' In-game Behaviors in Adaptive Gamification

Author:
Enrica LORIA

Supervisor:
Dr. Annapaola MARCONI

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in

Information and Communication Technology

Information Engineering and Computer Science Department
Head: Prof. Paolo Giorgini

April 2021

"*You gain strength and courage, and confidence by each experience in which you really stop to look fear in the face. You must do that which you think you cannot.*"

Eleanor Roosevelt

To Michel, who always believed in me,
even more so when I did not...

UNIVERSITY OF TRENTO

Abstract

Information Engineering and Computer Science Department

Doctor of Philosophy

Alone with Company: Studying Individual and Social Players' In-game Behaviors in Adaptive Gamification

by Enrica LORIA

Humans procrastinate and avoid performing activities that they deem dull, repetitive, and out of their comfort zone. Gamification was conceived to reverse the situation by turning those activities into fun and entertaining actions exploiting game-like elements. In practice, however, many challenges arise. Gameful environments cannot satisfy every player's preference and motivational need with a one-fits-all strategy. However, meeting players' motivational affordances can provide intrinsic rewards rather than extrinsic (e.g., points and badges). Producing intrinsic rewards is desirable as they are more likely to foster long-term retention than the extrinsic counterpart. Therefore, gamified systems should be designed to learn and understand players' preferences and motivational drivers to generate specific adaptation strategies for each player. Those adaptation strategies govern the procedural generation of personalized game elements - examples are task difficulty, social-play versus solo-play, or aesthetic tools. However, an appropriate personalization requires intelligent and effective player profiling mechanisms. Player profiles can be retrieved through the analysis of telemetry data, and thus in-game behaviors. In this project, we studied players' individual and social behaviors to understand their personalities and identities within the game. Specifically, we analyzed data from an open-world, persuasive, gamified system: Play&Go. Play&Go implements game-like mechanics to instill more ecological transportation habits among its users. The gamified app offers various ways for players to interact with the game and among one another. Despite Play&Go being one of the few examples of gamification implementing more diverse game mechanics than solely points and leaderboards, it still does not reach the complexity of AAA entertainment games. Thus, it limits the applicability of an in-depth analysis of players' behaviors, constrained by the type of available features. Yet, we argue that gameful systems still provide enough information to allow content adaptation. In this work, we study players' in-game activity from different perspectives to explore gamification's potential. Towards this, we analyzed telemetry data to (1) learn from players' activity, (2) extract their profiles, and (3) understand social dynamics in force within the game. Our results show how players' experience in gamified systems is closer to games than expected, especially in social environments. Hence, telemetry data is a precious source of knowledge also in gamification and can help retain information on players' churn, preferences, and social influence. Finally, we propose a modular theoretical framework for adaptive gamification to generate personalized content designed to learn players' preferences iteratively.

Contents

List of Figures

List of Tables

List of Abbreviations

A **A**chiever
AA **A**rchetypal **A**nalysis
BC **B**etweenness **C**entrality
ANN **A**rtificial **N**eural textbfNetwork
CHI **C**omputer **H**uman **I**nteraction
CC **C**loseness **C**entrality
CCA **C**anonical **C**orrelation **A**nalysis
CM **C**entrality **M**etric
D **D**isruptor
DC **D**egree **C**entrality
EFA **E**xploratory **F**actor **A**nalisis
EC **E**igenvector **C**entrality
FS **F**ree **S**pirit
F2P **F**ree to **P**lay
GUR **G**ames **U**ser **R**esearch
IoU **I**ntersection **o**ver **U**nion
PvP **P**layer versus **P**layer
PSN **P**layer **S**ocial **N**etwork
P **P**hilanthropist
PB **P**layer **B**ehavior
PR **P**age **R**ank
PT **P**layer **T**ype
R **P**layer
RF **R**andom **F**orest
RL **R**einforcement **L**earning
RS **R**ecommender **S**ystem
S **S**ocializer
SNA **S**ocial **N**etwork **A**nalysis
UI **U**ser **I**nterface

Chapter 1

Introduction

Adaptive Gamification embraces the need for tailored content in gameful environments [26, 119]. While gamification is the usage of game-like elements in contexts other than games [53], the term *Adaptive* introduces an additional layer in which players' personalities are studied and accommodated. The goal is to engage and retain users for longer to increase the likelihood of achieving the gamification goal, which is challenging to pursue [84]. In other words, Adaptive Gamification reverses the trend of a "one-fits-all" gameful design [84, 198] and moves towards personalized game content [173, 79]. However, producing adaptive content requires player modeling methods, describing users' intent, and providing actionable information on the content to produce.

Player modeling or profiling can be achieved either through theoretical [18, 66, 36] or data-driven approaches [92, 202]. The former exploits psychological theories and empirical findings usually operationalized in self-reports to extract user types [219], personality traits [103], or motivational affordances [237]. The latter relies on gameplay telemetry data and numerical features to model player behaviors. Both approaches come with benefits and weaknesses. Self-reports may suffer from perception biases from players, regardless of them being malicious or unintentional [224, 194]. Telemetry data can also be biased, especially when many redundant features exist [92]. As a result, analysts should interrogate a domain expert and model the system with care. Yet, data-driven approaches hold a significant advantage: they are versatile and easily adaptable [62, 92]. This ductility suits the mutable nature of players' behaviors within the game. Humans are complex and (seemingly) unpredictable beings. Thus, before defining a method for adaptive gamification design and maintenance, a deeper understanding of players should be achieved.

Games User Research (GUR) is rich of studies investigating telemetry data and behaviors within entertainment games (e.g., [202, 181, 55]). Those studies show how to analyze gameplay datalog to convey information on players and their gaming experiences [158]. Games, however, hold the advantage of massive multi-dimensional data, lacking in common gamified systems. Despite gamification data usually being more modest, meaningful insights can also be produced from the interaction with gameful apps. Game research can be used as a guideline and source of inspiration, providing methods and approaches adapted in gamification research. This observation is very broad per se, hence leaves much room for investigation. For instance, it is still unknown how much knowledge the analysis of players' behaviors within the gamified environment convey in terms of their overall experiences and preferences. Hence, the need for more research on the potential of gamification telemetry data and the extent to which techniques used in games can be employed.

Players' behaviors not only reflect their personalities and activities but also show how the community of players connects. Besides the environment, also, and especially, other individuals have a high effect on people's actions. Even in the real-world, peer pressure and social relationships influence our thoughts and actions [5]. Consequently, a person's life and essence are the results of cultural, environmental, and *social* factors. This also occurs in virtual worlds. Recent findings show how players build their own social network [226], interact with one another, and form active communities and groups [195]. In those communities, players can be *influenced* by other individuals [37, 195], favoring their permanence in the virtual world. These pieces of evidence lay the premise for the existence of social contagion also in gamification. Sociality has a huge psychological impact on players. Besides fostering retention and increasing engagement [181], interacting with others has positive effects also on well-being and mental health [105]. Despite few studies proving the benefits of social gamification [96, 231], the dynamics and implications of social interactions are widely under-exploited in gamification. Hence, the need for more in-depth research on this topic.

In summary, tailored content is needed as gamification can benefit from the deployment of ad-hoc game elements instead of treating players as a uniform unit [172, 103, 79]. However, to generate personalized content self–adapting to players' experience, a deeper understanding of users' interaction patterns should be achieved. Those interaction patterns become even more complex and meaningful when multiplayer game elements are offered, and thus, social play exists [121]. Data-driven approaches are widely used on games' data for many analyses, spanning from prediction to procedural content generation [55, 158]. Yet, little is known of how telemetry data can be exploited in gamification and how descriptive it is on players' experience and personality.

In this thesis, we dissected gameplay and analyzed it from two perspectives: individual preferences and social interactions. First, we researched whether the in-game activity can be used to predict churn and players' preferences. Then, we investigated social phenomena occurring within games and gamified systems aimed at finding influencers. Specifically, we studied whether influence can be used to assist the pursuit of the gamification goal. Our case study is a persuasive gamified system (Play&Go), active in the Italian region of Trentino to pursue greener transportation habits among its citizens. Play&Go implements a variety of game elements and allows several interaction patterns among its players, including the possibility of playing with others. Part of our analysis on in-game social dynamic is preliminarily conducted on games' data and then replicated and extended in the analyzed gamified system. The motivation to include these analyses is twofold. First, the preliminary evaluation of new algorithmic methods requires a social network of a larger scale of dimension than the network existent in Play&Go. Games offer such a vast population of players. Second, this dual analysis allowed us to compare social phenomena emergent in games and gamification. Although researchers argue that the social, motivational drivers leading users to play are in force in both games and gameful environments [96], systematic comparisons are still lacking. Therefore, our work also contributes to the field by building up a new piece in the bridge between games and gamification.

1.1 Research Questions

Existent research claims for more adaptive content in gamification [121, 26, 80]. Towards this, a better understanding on players' experience and in-game behaviors is needed to (a) **assess players' experiences**, (b) **learn their preferences**, and (c) **study social influence**. Having a complete overview of players' gameplay can inform an iterative **adaptive gamification framework** that continuously generate tailored content and, potentially, can deduce preferences' shifts.

Assessing Player Experience

Player experience can be (partially) evaluated by analyzing interaction patterns in the game. Game research includes a plethora of studies on entertainment games that exploit player behaviors to predict future actions [131, 116], profile players [181, 181], or inform marketing strategies [233]. These analyses are enabled by the multifaceted nature of game data and the substantial number of users involved. Gamification, on the other hand, often offers a restricted number of game elements and mechanics. Many gamified systems rely on basic elements, like points and badges [52, 121]. Such a simplicity discourages complex analyses, requiring many diverse features and a consistent population of users. Although researchers call for more variegated and diverse gamified applications [184], the complexity of games is unlikely to be achieved. Gamified systems and games have different scopes and belong to different domains, despite sharing similar motivational affordances [96]. In the spirit of those common points, however, gamification research can be inspired by game research and adjust methods and analysis already popular in Games Analytics.

Players rarely play as originally intended by the designers. While unexpected behaviors may be a marginal issue in games, as long as users keep playing, the problem can be more pressing in gamification. Unforeseen behavioral patterns may hinder the achievement of the gamification goal. In-game behaviors can be used towards this. Specifically, we researched how to treat telemetry data to extract players' winning strategies and anomalous patterns [145]. Therefore, we asked (*RQ_I_Monitor*): "**How can we process telemetry data to understand players' behaviors and strategies in a gamified system?**"

Nevertheless, providing a global overview of how the gameplay is progressing is not sufficient to grasp the multiple shades of players' experiences. For instance, this kind of analysis cannot provide information on a very relevant class of users: uninterested players at risk of abandoning the game—i.e., churn. Games researchers showed a great interest in churn prediction analysis [179, 131, 132]. Predicting players' abandonment is essential in entertainment games, as churners' timely detection may inform contingency strategies to avoid such a churn. In games, churn prediction has economic value, as retaining customers is more remunerative than attracting new ones. Although marketing reasons are generally not the primary focus in gamified environments, churn prediction is also important in gamification. Long-term retention is beneficial in the pursuit of the gamification goal. Considering limited gamification data, we researched whether information on players' activity and participation in the game is sufficient to predict their abandonment [143, 139]. In the form of a research question, we investigated (*RQ_I_Churn*): "**How can we exploit players' in-game participation behaviors to predict churn in a gamified system?**"

Learning Player Preference

Assessing and monitoring players' experience allows designers and developers to identify unengaged players and behaviors misaligned with the gamification goals. The information retrieved should be matched with appropriate changes in the gameplay.

However, to tailor the gameplay, high-level knowledge of players' experiences is not enough. Methods are needed to model or profile players' preferences. Player models can then be used to personalize the game content.

Player modeling or profiling associates to players, or a subset of players, the specific game elements or mechanics that they prefer, and thus, that motivate and engage them. In the gamification literature, much research has been conducted on theoretical taxonomies to build player types or traits [219, 216]. Those taxonomies are often grounded in psychological studies and are concretized in questionnaires and self-reports. Despite theoretically grounded approaches hold the advantage of being justifiable and explainable, self-reports are liable to memory loss and biases [224]. Data-driven approaches, on the other hand, rely on objective telemetry data describing players' actions and behaviors in the game [92].

Having investigated how players' idealized version of their preferences and their tastes contextualize in the specific gamified application, we studied whether telemetry data can also be used to learn player preferences. First, we analyzed preferences explicitly expressed by users through choices performed in the game. Hence, we proposed a method to translate their interactions into numerical values, easier to process and elaborate [138]. Then, connecting to the discourse on limited gamification data and the possibility of exploiting it to predict player churn, we researched if participation data can also be used to learn implicit preferences. Specifically, we exploited players' participation and activity data as feedback for a Reinforcement Learning-like algorithm. Then, we used users' implicit preferences to evaluate the preferences learned [141]. This study answers the question (*RQ_I_Learn*): "**Can we learn players' preferences from their implicit in-game participation behaviors following the Reinforcement Learning paradigm?**"

Given the existence of two different modeling approaches, data-driven and theoretical, we investigated the relationship between players' self-report and in-game behaviors. Among several theory-driven taxonomies designed for games, the Head User Type is the first model expressly conceived for gamification [140]. Hence, we investigated whether theory-driven and data-driven approaches lead to different adaptation strategies for the same user, where the term adaptation strategy refers to the type of content to generate for the user [136]. In this analysis, we defined the concepts of idealized and contextualized preferences. Idealized preferences are obtained from theoretical models and describe players' perceptions of the self. Contextualized preferences are extracted from telemetry data and reflect players' behaviors within the game. Hence, we asked (*RQ_I_Pref*): "**How do players' idealized preferences, expressed through the Hexad User Types Questionnaire, reflect their contextualized preferences, extracted from in-game interactions?**"

Studying Social Influence

Players' behaviors in gameful systems are also affected by their direct or indirect interactions with others [82], besides being conditioned by users' individual tastes. While player modeling can hint at users' preferences towards multiplayer game content, it is agnostic to social dynamics and influence. Yet, the social environment and the role the users assume within it provide important information on players' personalities. Even though players' may appear disinterested in social game elements in favor of solo play, they may still be affected by social comparison and influence.

Social influence describes the phenomenon by which some individuals (influencers) affect others' behaviors and choices [3]. Originally found in social media, influencers have also been detected in games [37]. Early works studying social networks in games analyzed communities built around game titles but existent on third-party platforms - e.g., Twitter [225], Twitch [230], or matchmaking websites [195]. Those studies already identified more cohesive and small communities, often representing guilds and groups, where the whole group activity is influenced by the presence of specific individuals, such as moderators and leaders. Like social media, Twitter and Twitch feature very visible users that exert influence over the gamer network, for instance, by conditioning the popularity of specific games. However, social influence was never analyzed within the gameplay, using telemetry data, and extracting in-game behaviors. A notable exception is Cannossa et al.'s work [37], where influencers are described as well-connected and well-positioned users in the player social networks. Hence, those influencers were characterized by having many connections and by assuming strategical roles within the network. Although influencers were, then, found to impact players' long-term retention in the game analyzed (Tom Clancy's The Division), they did not possess particular properties—i.e., they were discerned from power users or elite players.

Recalling the definition of social influence as individuals conditioning others' behaviors, we further investigated games retention influencers. Guided by the Social Network Analysis (SNA) literature, we adopted a different approach to identify them. Instead of researching central and popular users, as in the previously mentioned study [37], we detected influential users impacting others' behaviors. Therefore, having defined the behavior of interest (in the form of a series of game features), given two players X and Y connected at a time t, X influences Y if Y's behavior becomes more similar to X's after time t. We defined a semantic algorithm to detect influencers and researched them in the online multiplayer game Destiny [144]. Specifically, we asked (RQ_S_Inf): "**How does social influence manifest in online multiplayer games, and how retention influencers differ from central players in the network?**"

Having found that semantic influencers had a higher impact on long-term retention than central users, in Destiny, we studied whether the phenomenon also occurs at a higher level. In other words, we investigated whether the status of influencers persists across game titles. To this end, we collected data from a popular game provider, Steam, and computed the semantic influence algorithms [137]. By answering the question (RQ_S_Steam) "**How does social influence in the Steam players' community differ from social influence within games, and which properties do influencers possess?**" we increased the current knowledge on influencers. For instance, we learned whether a retention influencer in a specific game is more likely to also be an influencer in a different title or if they encourage others to play new games.

Following the works on social influence in Destiny [144] and Steam [137], we analyzed social relationships in our gameful application (Play&Go). As we discussed, retention is also vital in gamification. Hence, successful gameful systems can highly benefit from the presence of individuals who can help pursue a longer permanence in the game. Towards this, we analyzed the two-player challenges in our persuasive gamified system (Play&Go) and derived a players' social network. Besides seeking players' retention, gamification designers have to ensure the gamification goal's pursuit (e.g., crowdsourcing, education, behavioral change). Considering the strength of

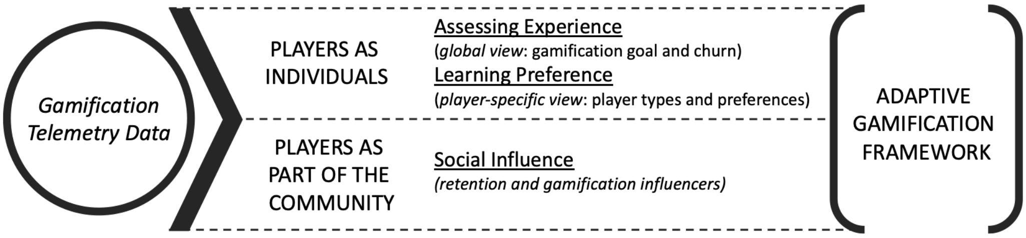


FIGURE 1.1: **Thesis roadmap**, representing the flow used in the manuscript connecting our analyses and finding. The specific investigations on players' experience and preferences (players as individuals), connected to the study on social influencers (players as part of the community), contribute to the definition of a conceptual adaptive gamification framework. Those three topics (players as individuals, players as part of the community, and the adaptive frameworks) are each presented in a dedicated chapter (Ch. 4–6).

the semantic influence algorithm that allows specifying the behaviors we are interested in influencing, we also investigated the presence of other types of influence aligned with the gamified app's underlying purpose. In Play&Go, the gamification goal is to promote green mobility habits. Hence, we researched the presence of influencers impacting others' transportation behaviors [142]. Therefore, we asked (*RQ_S_Gamif*): "**What are the different types of influence(rs) existing in gameful systems?**", aimed at researching the equivalent of games' retention influencers but also the existence of gamification influencers helping to achieve the gamification goal.

Adaptive Gamification Framework

Understanding and modeling players' preferences and social interactions are only the prefaces of the adaptation process. The information obtained needs to be operationalized into concrete tailoring strategies ad-hoc designed for each player. The mainstream approach for adaptation is relying on player profiling [160] or motivational [160] theoretical models. However, the results can be conflicting [78], especially when players are reduced to a single dominant type [78]. Sometimes those theoretical models are combined with recommender systems [218], which work well when small, focused tailoring is required [112], but cannot scale with the complexity of the task of producing a complete model.

Besides the task being challenging, gamification design adds a layer of complexity, as players' enjoyment and needs to connect with the underlying goal [26]. Consequently, the success of the gamification application is highly dependent on players'—hopefully long-term—engagement and the pursuit of the goal [89]. Moreover, people's behaviors are mutable, and preferences may shift over time [62]. Adaptation strategies should also account for this mutability, hence the need for automatic, dynamic, adaptive methods for tailoring gameful systems [119]. Those automatic systems should consider players' specific preferences and in-game social rules while keeping users' behaviors aligned with the gameful app's purpose.

Following our analyses on players' behaviors [145, 139], preferences [141, 136], and social connections [142], we combined our findings to study the feasibility of an adaptive gamification framework. In the form of a research question, we asked (*RQ_Fram*): "**How can we model player in-game behaviors and social interactions to inform the generation of adaptive, dynamic game content?**"

TABLE 1.1: Mapping of the research questions to the hypotheses, chapters and publications.

ID	RQ	H	Chapter	Publication(s)
RQ_I_Monitor	How can we process telemetry data to understand players' behaviors and strategies in a gamified system?	H1a&b	Ch. 4.1	[67, 151, 145]
RQ_I_Churn	How can we exploit players' in-game participation behaviors to predict churn in a gamified system?	H1c	Ch. 4.1	[143, 139]
RQ_I_Learn	Can we learn players' preferences from their implicit in-game participation behaviors following the Reinforcement Learning paradigm?	H2a	Ch. 4.2	[138, 141]
RQ_I_Pref	How do players' idealized preferences, expressed thought the Hexad User Types Questionnaire, reflect their contextualized preferences, extracted from in-game interactions?	H2b	Ch. 4.2	[140, 136]
RQ_S_Inf	How does social influence manifest in online multiplayer games, and how retention influencers differ from central players in the network?	H3a	Ch. 5.2	[144]
RQ_S_Steam	How does social influence in the Steam players' community differ from social influence within games, and which properties do influencers possess?	H3b	Ch. 5.2	[137]
RQ_S_Gamif	What are the different types of influence(rs) existing in gameful systems?	H3c	Ch. 5.3	[142]
RQ_Fram	How can we model player in-game behaviors and social interactions to inform the generation of adaptive, dynamic game content?	H4a&b	Ch. 6	[135, 34]

1.2 Hypothesis and Contributions

In summary, in this thesis, we present a multi-perspective analysis of players' in-game behaviors to inform a theoretical adaptation framework (Figure 1.1). Given prior findings in Gamification and Games User Research (GUR), we hypothesized that:

(1) *In-game behaviors describe players' experience in gamification.* Gamification telemetry data can be used to (H1a) identify players' strategies, (H1b) monitor behavioral evolution, and (H1c) predict churn.

(2) *Player preferences can be extracted from their in-game behaviors.* Players' level of activity varies with the type of game element they interact with. Hence players' activity can be used to learn their preferences (H2a). Besides, in-game behaviors describe players' preferences contextualized in the specific virtual environment and do not necessarily match general theoretical preferences (H2b).

(3) *Online gameful environment can foster social influence.* Influential individuals affecting others' retention do exist in games (H3a), across different games (H3b), and in gamification (H3c).

(4) *Telemetry data can be used to inform an adaptive gamification framework.* Players' datalogs can be processed and organized to produce automatic content generation (H4a), dynamically adjusted at runtime as the game progresses (H4b).

Testing these hypotheses results in the following contributions to the GUR and gamification research. First, we connect gamification research to game research by showing how telemetry data can also be used in gameful systems to extract important information on players' experience. Second, we show how in-game behaviors can be used to predict players' churn and thus estimate the goodness of their experience. Third, we provide a data-driven approach to learn players' preferences in gameful environments. Forth, we provide a systematic and visual comparison of players' self-reported preferences and preferences extracted from their behaviors. Fifth, we suggest a method for identifying influencers in games and gameful systems and show how those influencers impact others' retention and can help to achieve the gamification goal. Finally, we present a modular, theoretical adaptation framework embedding all our findings.

Table 1.1 provides an overview of the research questions and how they relate to the hypotheses, the chapters, and the publications.

1.3 Thesis Structure

The manuscript is organized as follows. Chapter 2 introduces the necessary background on game research, player modeling, and social network, as well as the related works on adaptive gamification and social play. Chapter 3 describes the data and the case studies analyzed: the persuasive gamified system (Play&Go), the game Destiny, and the game provider Steam. Chapter 4 focuses on the individual play experience of users and presents the analyses on churn prediction and player preferences. Chapter 5 describes the study of social play and influence in games and gamification. Chapter 6 presents the adaptive gamification framework, built upon the studies discussed in Chapter 4 and Chapter 5. Chapter 7 presents a general discussion on the thesis and the limitations of this research. Finally, Chapter 8 concludes this thesis with a summary and suggestions for future work.

Chapter 2

Related Works

In the following chapter, we will review relevant works from the literature. First, we will introduce gameful systems and research the motivations for adaptive gameplay in gamification. We will then show how players' telemetry data is a precious source of information to assess players' experience and preferences. Having discussed works analyzing players' individual gameplay, we tackle games' social aspect. We provide evidence on games' being a way to connect individuals and form communities and incubators of peer pressure and social contagion.

2.1 Gameful Environments

In its simplest, most common definition, *gamification* is described as the *the use of game design elements in non-game contexts* [53]. It has been used for about ten years in educational settings to increase learner performance, motivation, or engagement [11, 48], and to make users intrinsically motivated in performing a certain task [163]. In the last decade, gamification became a widely employed term, with its own field in Computer Human Interaction (CHI) [198, 166]. Previous studies have shown that gamification can turn unpleasant tasks into fun ones and enhance the user experience in numerous contexts and domains [84, 198]. In addition to academia, gamification has leaped the industry and has become an established practice in user experience design [166, 198], while also gaining popularity in different domains [121]. However, gamification is not always successful, as its effects highly depend on its context and implementation [84]. Hence, research has focused on improving the design of such systems to maximize their success rate [152], dictated by the pursuit of the gamification goal while minimizing negative effects, such as churn [121]. Towards this, gamification analytics describes the toolset employed to measure the impact and improve gameful apps [89]. Although researchers found that gamification generally produces positive effects [231], implementing gameful mechanics does not implicate automatic increases in users' activity [81]. The difficulty of gamification lies in the meaningful connection of players' needs to the gamification layer [26]. Conversely to the action, generally, gameful apps promote, the choice of playing should be a voluntary behavior [191], intrinsically motivated. A behavior is intrinsically motivated if it fulfills the needs for competence, autonomy, and relatedness [49, 190, 191]. All of these motivational needs are commonly satisfied by playing games [191]. Nevertheless, the satisfaction of such motivational affordances is highly personal [49, 190]. It derives that treating players as a homogeneous unit is an unsuccessful strategy [172]. Thus, the gameful affordances implemented in gamification research should be diversified to allow the creation of more inclusive gameful experiences [121].

2.2 Adaptive Gamification

Gamification is mostly used in a "one-size-fits-all" manner, built of a static set of gamification elements (such as points, badges, or larboards) [84, 198]. At first, gamification research primarily focused on investigating its applicability in different domains. Whereas more recent research aimed at understanding why gamification works and which factors play a role in the success of gamified interventions [166]. Studying the effects of gamification showed that gamification should be tailored to the users to be effective, with pieces of evidence coming from the domain of health [173], sport [134] and learning [113, 48, 128]. The results revealed that interpersonal differences affect the perception of gamification elements (i.e., not all users are motivated by the same elements). When preferences are not met, researchers obtained inconclusive or even negative outcomes deriving from a static gamification approach [6, 198, 84]. Thus, the question of how to tailor gamified systems towards the user arose and became an important research direction.

Previous research has revealed that demographic factors such as age [23], gender [171], or personality traits [103] explain user preferences for gamification elements to a certain extent. The need for diversified content led to the development of many theoretical models aimed at classifying players according to their personality [18], motivations [237], and play styles [36]. Additionally, self-assessments are also employed to measure the level of enjoyment and the level of flow of players [98]. However, there was no model or framework, which specifically targets personalization in gamified systems until Marczewski [153] proposed the Hexad User Types model. The model consists of six player types, which differ in the degree to which they are driven by autonomy, relatedness, competence, and purpose (which are core pillars of Self-Determination Theory [190]). Tondello et al. [219] developed and validated a questionnaire to extract the Hexad User Types [216]. Based on this, the Hexad user types model could be used to investigate user preferences in gamified systems across various domains, reaching from physical activity [7], health [174] or energy conservation [122] to the education domain [162]. These studies supported the Hexad model's usefulness to explain user preferences in gamified systems by consistently connecting the perception of gamification elements and Hexad user types across domains. This is additionally supported by research conducted by Hallifax et al. [79]. They compared the Hexad model against other factors and player typologies and concluded that the Hexad model is advantageous in explaining user preferences for gamification elements. Hence, the young field of tailored, or adaptive, gamification is mostly focused on user modeling for a future personalization, adaptation, or recommendation of game elements.

Despite those promising outcomes, adaptive gamification is still an under-investigated topic, and further research is needed to understand better how fully improve player experience. Such emergent field benefits of a theoretical base build upon several studies conducted in real-world learning settings. In static adaptation approaches, profiles are identified, often relying on theoretical models (e.g., the Bartle Player types [18], the Brainhex player satisfaction model [165], the Hexad player types [219], or the categories of players described by Ferro et al. [66]). Users are sorted into different categories based on these profiles; different game elements are given to each of the different categories of players [80]. In dynamic adaptation, the game elements' deployment is informed by players' activity in the game. A user profile is built by monitoring her in-game behaviors. In the work of Paiva et al. [175], users' actions are divided into either collaborative, gamification, individual, or social interactions. The adapted element is the player's goal, which is assigned according to the type of actions performed. Another example is Jagust et al.'s [100] study, in which two dynamic adaptation scenarios are presented. In the first situation, the players' task is timed. When the task is completed, players are given less time for the following one. In the second situation, the difficulty is defined by the completion time instead of the target score. The score increases as they carry out the task correctly. In another work, the customization is in the types of badges presented to and in the feedback given [113]. Other studies use a hybrid approach in which they define static adaptation rules for given player profiles. When the profile of a user changes, different game elements are given to her [160]. The Brainhex model inspires the profiles. In the systems suggested by Knutas et al. [120], an algorithm uses players' profiles and interactions. In this case, the authors use the Hexad player profile and users' skills. They recorded students during their project meetings to classify their interactions. Then, they proposed the game elements according to those profiles and interactions. However, in this work, a method to detect these actions in real-time lacks [120].

Researchers claim the need for more dynamic tailoring, also relying on other features than static player profiles [119]. Therefore, the usage of implicit data, as telemetry logs, is encouraged. Within games research, player profiling and content generation already heavily rely upon data-driven approaches (e.g., [202, 148]), confirming how content adaptation can lead to higher motivation and longer-term engagement [180]. This trend is also picking up in gamification research. For instance, a recent study on gamified educational systems promotes an automatic identification of student flow experience to measure the systems' influence on the learning outcome. Their findings show how adding dynamic difficulty adjustment and other flow dimensions can improve the gamification experience [177]. Other works also show the benefits of using data-driven approaches in the education domain [214].

All in all, the field of adaptive gamification is still young and growing, especially for what concerns dynamic adaptation [80]. As such, gamification research still misses coherence in research models, as well as theoretical foundations [121]. Moreover, the motivational affordances implemented in the majority of the studies remain 'goals and objectives', 'multimedia feedback', and 'metaphorical/fictional representations' [228], neglecting others such as the expression of self-identity [240]. Since different things motivate different people, it can be expected that personalizing the incentives and the way the rewards are presented to the individual would be beneficial [222]. In particular, tailoring at the level of social influence strategies may increase the effects of the persuasive technologies [107]. Yet, the conversation is still focused on what can and needs to be personalized, rather than on how to tailor experiences [119]. Thus, researchers call for more studies on automatic adaptation and dynamic modeling [119, 121], with associated impact evaluation [119]. Scholars and practitioners should exploit implicit behavioral data [51, 89] to update and improve the gamification model in a cyclic manner [121, 119], benefiting by periodically producing novel content [121]. Gamification studies would benefit from broader use of theories to account for the complexity of human behavior and a more thorough exploration of the many opportunities coming from the world of games [184]. Hence, game and gamification analytics theories and empirical findings must be studied to derive truthful and reliable models of players' experiences and personalities.

2.3 Modeling Player Experience

Through game [62] and gamification [89], a better understanding of players can be achieved, which can be beneficial for both customers and suppliers.

Game metrics can measure quantitative and qualitative attributes [62]. It is a common practice to employ self-assessments to model player experience (e.g., PENS [191], GEQ [98] and IEQ [102]). This type of data source is known as subjective Player Experience Modeling (PEM) [150]. Despite the many techniques that exist to address the tool's subjectivity, especially on large-scale quantitative studies, some flaws persist. Answers might be biased either because players have a wrong perception of themselves, want to please their interlocutor, or have a different scale of measure than the other participants [62]. An additional issue may also be memory faults caused by answering questions after long gameplays.

The usage of psycho-physiological measurement during gameplay [149], called objective PEM [236], is another option. For instance, physiological signals are an alternative form of input, proved to be an effective way to evaluate the experience [149, 150]. Nevertheless, such external tools cannot always be easily integrated into the system, either for economic or logistic reasons. 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 [91]. Each player modeling approach involves different challenges to obtain meaningful information about player behavior [92, 202].

Data mining on user behaviors is a practice aimed at supporting the design and development of systems. When the data analyzed comes from games and gameful environments, the term used is Game Analytics. Researchers and practitioners are interested in detecting meaningful information from players' interactions [133]. This information can also inform the customization of users' experiences. On the one hand, the designers' and developers' interests lie in retaining and attracting many users. On the other hand, users seek engaging experiences. To this end, a plethora of sources can be used to gain knowledge about players [99]. Gameplay data is also a valuable source of information. Gameplay-based PEM [236] is driven by the assumption that the quality of player experience is deductable from in-game interactions. Not only this approach allows having immediate feedback on the way users interact with the game, but it also eliminates the biases likely to occur in more subjective methods [194]. The objective nature of data-driven gameplay-based PEM led to the extensive use of those metrics to assist the design and development of games [115]. Gameplay data is very insightful when measuring the player experience is the purpose [62], while being easy to retrieve through an automatic logging system [115]. From the raw gameplay data, we can extract several metrics. Examples are metrics to measure basic learning indicators to adapt to the difficulty level and the evaluation of players' performance [114]. Nonetheless, such methods are not unflawed. Some issues are generalizability problems, algorithmic efficiency, and the ability to consider player variation - e.g., skills development - in time [92]. Thus, to avoid misleading interpretation, data should be meticulously analyzed [92].

2.3.1 Churn Prediction

Measuring and estimating players' experience is only the first stage. Knowledge of users' level of satisfaction can be exploited as feedback to improve the overall system.

Delivering a high-quality experience is the key to maintain the system alive. An example of the opportunities deriving from gameplay analysis is churn prediction studies [56, 210, 22].

The likelihood of a player abandoning the game can be measured by analyzing gameplay data and identifying the behaviors that can be symptoms of an abandonment [62]. Many studies addressed the problem of assessing whether churn is indeed predictable and which features should be studied to help prevent it (e.g., [179, 76, 158, 108, 147]), often using context-specific game metrics [147] or metrics solely applicable to full-featured games [76]. The vast majority of studies use aggregate data to predict churn [147, 76], instead of analyzing temporal data. Data dynamicity should also be considered since it is proved that players, rather than having a static in-game behavior, tend to change it throughout their gameplay [62]. An exception is the study of Hadijii et al. [76], in which they defined a generic model that considers temporal features.

Nevertheless, some researchers also investigated the power of more generalizable measures, such as users' activity, to predict churn and in-game purchases. Xie et al. [233] propose a prediction model using frequency-based data to classify disengagement and first purchase. Liu et al. [131, 132] adopted a different approach by analyzing churn at both a macro and a micro level using graph mining. They collected and explored a considerable amount of data from a Chinese game provider, showing that players' activity across different applications is a predictor of churn likelihood. Also, they present a model to rank applications according to the users' abandonment rate. While they provide findings that monitoring players' interaction with the application is a predictor of abandonment, they exploit cross-game information to make predictions available in a game provider but lack specific applications. Moreover, the focus is on free-to-plat games. Churn prediction in those games is very specific due to the nature of those games, in which players tend to churn very quickly [116]. Therefore, there is no evidence that their models, methodologies, and findings hold in such a different context as gamification.

Churn prediction is undoubtedly essential to timely identify players at risk of abandonment. However, identifying those players is truly meaningful if associated with strategies to avoid such a churn. A viable strategy might be to introduce or improve the generation of personalized content for those users at risk, as players are notably more likely to be retained when the system accommodates their tastes.

2.3.2 Assessing Player Preferences

The self-report method (i.e., asking users to self-report on their perceptions and preferences) is the most common mode of assessment [224]. In gamification, this research method is the most frequently used to assess preferences for gamification elements and tailoring gamified systems [119]. Results related to personalized gamification were established mainly through surveys, storyboards, and questionnaires [119].

For instance, Orji et al. [173] used storyboards to explain persuasive strategies and investigate how far personality traits moderate the self-reported preferences. Similarly, storyboards have been used to assess self-reported perceptions of gamification elements and whether they are correlated to Hexad user types by Altmeyer et al. [7, 8] and by Hallifax et al. [79]. Another way to illustrate gamification elements is using textual descriptions. Tondello et al. [219] studied correlations among Hexad user types and game elements in a general context. They textually described the gamification elements and asked participants to self-report whether and how they felt motivated. Kotsopoulos et al. [122] followed a similar method and used self-reported preferences to inform personalized gamification to reduce energy consumption. The self-reported perception was also used by Jia et al. [103] to assess the motivational impact of gamification elements, illustrated through videos.

Self-reports offer many advantages, such as easy interpretability, the richness of information, and sheer practicality. However, there is also substantial criticism of this method [224]. A major issue with self-reports is the credibility of answers (i.e., whether the respondents' answers are accurate). Past research has also demonstrated problems such as social desirability bias (participants answering in a socially desirable manner), acquiescent responding (participants tending to agree with statements), or constraints on self-knowledge. Especially the latter (i.e., the fact that participants might not be able to report their perceptions or preferences due to a lack of recall, experience, or knowledge), provides a threat to personalization research in gamification, mostly relying on self-reports.

Alongside self-reports, grounded in psychological theories, researchers also investigated the possibility of directly understanding players' experience by analyzing their interactions with the game. Those methods exploit telemetry data build from players' in-game interactions. While much research has been done on games, for conciseness's sake, we primarily reviewed gamification studies, as this is our main focus. Gamification research, inspired by current trends in games, is approaching the usage of machine learning to predict and model players' behaviors. Churn prediction, for instance, can be used to identify non-engaged users [17]. Recent studies also show how the relationship between video gaming preferences and gamification feature categories can differ from what theoretical frameworks state, especially regarding social-orientation [88]. Nevertheless, it is unknown whether such a dichotomy can be attributed to surveys' flaws, previously described, or other phenomena.

Users' actions are driven by motives, which can be implicit and explicit [154] and rarely correlate among one another [90]. Implicit motives describe impulsive behaviors [154], whereas explicit motives drive responses in structured environments resulting from elaborated choices [154, 155]. Therefore, explicit motives can be accounted for people's actions only in certain situations [33] and they do not necessarily correspond to what they actually enjoy [156]. Users' preferences can vary with the environment they are placed into [221]. Besides, they are also affected by the physical medium they are using [209]. On the other hand, implicit motives can explain long-term behaviors applicable in many more various domains, rather than in concrete goals within specific environments [156].

Those findings raise the question of whether static player types derived from self-reports are robust enough to embed and model what players actually enjoy, instead of only the superficial conscientious choices. This issue is extremely complex due to the many confounding factors. Hence, we believe that the first problem to tackle is complementing this research by analyzing the users' player types (self-report) and in-game behaviors (spontaneous actions) when interacting with a gameful system. Towards this, our scope is to trace a thread tying theoretical models and data-driven analysis to research whether those approaches are effectively pointing in different directions, or rather they converge into similar results.

2.3.3 Data-Driven Player Profiling

The most popular technique used to profile players is clustering, which consists of finding groups of samples with similar characteristics in a dataset [20, 203, 50]. 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 [57]. In recent Games User Research works [203, 195, 185], 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. 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 [13]. Previous work has found that players can belong to different clusters in different stages of a game [181, 106], adding further support to the belief that player profiles need to iteratively be updated for them to remain useful [202]. Understanding the temporal aspects of gameplay is also essential in analyzing how engagement and frustration may vary over time through player behavior. However, in most cases, temporal data is used to enhance the accuracy of prediction of specific events or churn prediction [192, 94]. Despite those models often being very accurate compared to aggregated models, they are not human readable. Thus, designers cannot be meaningfully informed by their outcomes.

When adaptation and personalization are introduced, most of the time, researchers and practitioners focus on how to profile players' individual preferences and adjust the game accordingly. Consequently, social game mechanics are usually reduced to one of the elements that players may or may not like. However, many studies provided evidence on how sociality is essential in games, which connect individuals through play. Sociality is multifaceted, and thus in-game relationships need to be closely studied, as well.

2.4 Social Play

The belongingness hypothesis states: "*human beings have a pervasive drive to form and maintain at least a minimum quantity of lasting, positive, and significant personal relationships*" [21]. Humans' need to feel close and connected to others is also considered a fundamental construct in self-determination theory [49], as part of the concept of *relatedness*, and further accepted in other psychological theories [63]. The individual's need to belong is, often, what influences their choices and decisions to begin an activity or behave a certain way [21]. People seek relatedness in any kind of activity [21], often in games [191, 187]. By their nature, games embed some degree of social connection, either directly or indirectly, differently from non-gameful actions [189]. Players are more likely to keep being engaged and keep playing when their need to belong is fulfilled [183], as the nature of in-game relationships can impact their behaviors and participation [99, 223]. Social play can produce a feeling of well-being and an increase in players' performance [182].

Researchers found how commitment to the game benefits from peer-to-peer communications [46], because feelings of relatedness increase enjoyment [46]. Not only explicit but also non-verbal communication can foster closeness and social play [127]. However, not every multiplayer game successfully produces belongingness and relatednesss [60, 86]. People feel connected when a sense of presence [47] and virtual community [24] is created, in which virtual social agents are perceived as real-world social individuals [127]. Well-designed social game mechanics can result in players' will to complete their tasks and be retained for longer: they are more motivated to have success [167]. Within virtual environments, players can form long-lasting friendships within games, which can continue not only in the real world but also in other games [44].

2.4.1 Social Play in Gamification

Gamification is strongly related to motivational design, specifically prevalent when trying to involve users in tedious or repetitive tasks and to trigger new behaviors. According to one of the most used definitions, gamification is the use of game-like elements in non-game contexts [53]. In many studies, this definition has been translated into the implementation of the "blueprint triad" [52, 121], meaning points, badges, and leaderboards (often called PBL). Gamification research still heavily relies on achievement-oriented affordances. [121], despite social-oriented affordances showing promising new perspectives [83, 48, 231].

Popular approaches toward gamification—to effectively achieve long-term players' retention—must move towards intrinsic rather than extrinsic motivation. Intrinsic motivation can be driven by the primal need for a feeling of belongingness. In a recent study on gamification, social game elements positively satisfied the autonomy, competence, and relatedness needs [231]. Hence, gameful systems can benefit from the inclusion of game mechanics related to cognitive and social motivations [1]. Social gamification helps achieve more successful gamification examples, as players' are more motivated and engaged than in reward-based gamification [213]. Besides, users are driven towards technology, in the long-term, when social influence is in place, especially when they can be positively recognized by peers and are offered the possibility to network [83]. A sense of social interaction conditions the motivation of using the system, increasing users' engagement [186].

Social influence benefits are found in different application domains. For instance, in gamification for health, players were found keener to engage in physical exercises when playing with friends [83]. Players' engagement and participation were proportional to the number of friends they had in the service. Moreover, having a partner to play with eases the process of weight loss [126]. Learning can also be improved when social gamification is used [123, 48]. Learners not only were retained for longer in the system, but they also showed better performance in the tests [123]. Social play not only can be exploited to positively instill individual behavioral change but also impacts community activities. For example, crowdsourcing can be made more engaging by using competition [2].

However, divergent opinions exist on how social dynamics should be implemented and whether they may negatively affect users. [2] argue about the benefits of competitive game elements. When they compared the competition to cooperation, they obtained detrimental results for the competition. Other studies found that competition negatively affects social behaviors [88], too. In detail, the authors hypothesized how those adverse effects may be explained by an overachiever's presence impacting players less skilled in the activity [88]. Researchers claim that a competitive environment inevitably fosters negative social interactions [16, 97], whereas cooperative gamification is likely to positively connect users [164]. In turn, users experiencing positive social feedback may be more prone to continue using the systems [212] and benefit from it [83].

Thus, more research should be conducted to identify the variables that affect social play, engagement, and participation within gamified systems.

2.4.2 Social Network Analysis in Games

Researchers and designers can rely on Social Network Analysis (SNA) to monitor the status of in-game social interactions. Player social relationships can be modeled through graphs, which successfully represent interaction patterns among a group of people [193]. Social networks manifest when direct or indirect social interactions are allowed. Social media and standard Online Social Networks (OSNs) explicitly define connections among the users, linked because they are related or share interests. On the other hand, despite being originated from indirect connections, implicit social networks are as rich as information. Hence, they may enforce similar social rules. For instance, online multiplayer games, being a social phenomenon, encourage social interactions. These interactions, or relationships, can also be interpreted as social networks, thus being modeled using traditional SNA techniques.

Online multiplayer games convey information on the social aspect of gaming [69] and help understanding social relationships in a highly digitalized world [59]. The study of how players socialize through games can lead to better social environments in games [59, 58]. For instance, studying the players' network in the form of a graph can highlight how players' activity is reflected in others' experiences [235] and when the permanence of certain players can condition others' behaviors [235]. Inspecting the player network hinted that pending more time in teams is not a synonym of being more social, as players' interests in social interactions can be merely functional to the game [95, 58]. Similarly, toxic interaction patterns may emerge from the analysis [104].

Although the usage of SNA in games is still young, researchers have already analyzed the social roles of players. Not only group formation represents a pillar of the player community [59] but, sometimes, loyalty to the guild led players to prioritize its growth rather than their own [9]. The team organization and connectedness also benefit the individual performance [117] and retention [182].

2.4.3 Social Influence (in Games)

Social networks are an essential tool to identify and understand influencers in online networks. In a social network, nodes often tend to resemble their neighbors. This similarity happens either because similar individuals tend to connect or because they mimic the behavior of some other individuals [43]. The first phenomenon is called homophily [157], or selection, while the second is named social influence [3].

Influence is a widely studied topic in social network analysis, and yet there is no agreement on the definition of an influential person [146]. From state of the art, two types of influencers can be distinguished: (1) individuals affecting the spread of information or behavior [200]; and (2) individuals manifesting a particular combination of desirable properties, which span between expertise and position in the network [70]. Many terms have been used to address those influential users. When they impact other behaviors, those individuals are referred to as opinion leaders [61], innovators [35], key-players [27] and spreaders [75]. When they are well position and connected in the whole network, they are usually called celebrities [168], evangelists [12] or experts [75].

Using centrality measures to identify influencers has been proven to be a relevant approach [74, 14]. More specifically, in- and out-degree, betweenness, eigenvector, and closeness are the more widely used metrics [197]. Although these measures are distinct, they are conceptually related [188]. While, due to their definition, those metrics seem to be very aligned to the influencers of the second type, there is no trivial evidence that they are sufficient to identity influencers of the first type - i.e., influencing people's behavior. Instead of being fundamental to keep the community connected, these specific kinds of influencers foster similarity among the nodes in that others tend to emulate them.

Many researchers have studied and modeled the concept of influence and its spreading throughout the network. Generally, in those works, influence is said to occur when B performs an action after A performed it. The probability of influence degree can be learned from a log of users' actions [73]. A similar interpretation of influence, which is more tied to the individual's identity than their actions, is the study of the conditional probability that similarity increases from t-1 to t between two nodes that become linked at time t [197]. Also, a combination of the two approaches is used, modeling both user attributes and actions over time. Tan et al. have used the latter methodology. [211] to compute the likelihood that the user also performs the action, which is increases when one's friends are performing said action. Various ways of measuring the influence of users have been analyzed [15, 32, 38]. Studying the increase of similarity among users over time also allows modeling the idea of reinforcing influence when the interaction perpetuates. It is shown that similarity steadily increases even after the first interaction, although at a decreasing rate [43]. Influence has also been used to differentiate between strong ties to weak ones [232]. But most of this work has focused on online social networks, while work on influencers in games is still rare.

Multiplayer, or social, games foster social relationships by nature, and thus, can also be modeled through a graph. Researchers studied groups and community [60], investigating the impact of social structures in gameplay [176, 182]. Properties of groups and guilds, for example, are indicators of players' in-game activity and retention [185, 195]. Although the player communities formed around games hold important information about the social dynamics occurring [226], the psychological drivers moving players' behaviors are still underexplored [62].

The possibility that any behavior can be influenced has also been explored in games. Alongside general studies on social structures and communities [195, 226], recent works investigated the existence of influencers in popular games. *Structural* influencers, or central users, have been analyzed in the game *Tom Clancy's The Division* [37]. The authors found how central players were different from both *power users* and *traditional players* and fostered long-term retention in others. While these studies highlight the importance of influencers in the players' community, they are still limited to similar domains. Researchers claim that the same motivational affordances move players in both games and gamified environments [96]. However, the findings on social influence cannot be assumed to generalize also in gamification.

Chapter 3

Case Studies

In the current chapter, we will present the use cases analyzed in the manuscript and an overview of the features and some statistics of the collected data.

First, we will introduce the core application domain: **Play&Go**, a persuasive gamified system fostering ecological transportation habits. Play&Go is active in a whole Italian region. Players participate in the gamification campaign on a voluntary basis, rather than being invited to a lab experiment. Since 2015, yearly gamification campaigns have occurred. Users also can provide feedback to the designers, either by directly approaching them via email and during the yearly events organized or by completing in-app surveys. Players' feedback is then analyzed and exploited to improve the next edition of the application in a cyclic co-design process. The most notable product of co-design has been the introduction of multiplayer mechanics in 2018. In this thesis, we analyzed three editions of the game, starting from 2018.

Play&Go implements a variety of game elements, both single and multiplayer. Such a variety is suited to enable our research on adaptive gamification, as it gives the freedom and space to analyze different behaviors and offer personalized content. Nevertheless, the size of the players' population is still massively reduced compared to traditional games, whereas the study of novel algorithmic approaches required a more consistent player base. To fulfill this need, we also analyzed game data from a Free-to-Play online game (**TagPro**), an online multiplayer FPS (**Destiny**), and a game provider (**Steam**). Including game data has an additional benefit other than testing new approaches: bridging gamification and game research, highlighting the differences but, especially, the similarities among these seemingly disjoint worlds. Games and gamification players are moved by similar motivational drivers, yet game research is and cannot be immediately generalized to gameful systems. Despite bridging gamification and games is not the main scope of this thesis, we still partially address this issue in the process of designing new analyses to study individual and social behaviors in gameful environments.

3.1 Gamification Data: Play&Go

Play&Go [67, 151, 109], as illustrated in Figure 3.1), is a gamified system, in the form of a mobile application, with the goal of producing a positive mobility behavioral change in its users towards sustainable transportation means. The app is sponsored by Trento and Rovereto's municipality, the two cities of the Italian province of Trentino. The sponsorship allows organizing yearly 6-month gamification campaigns

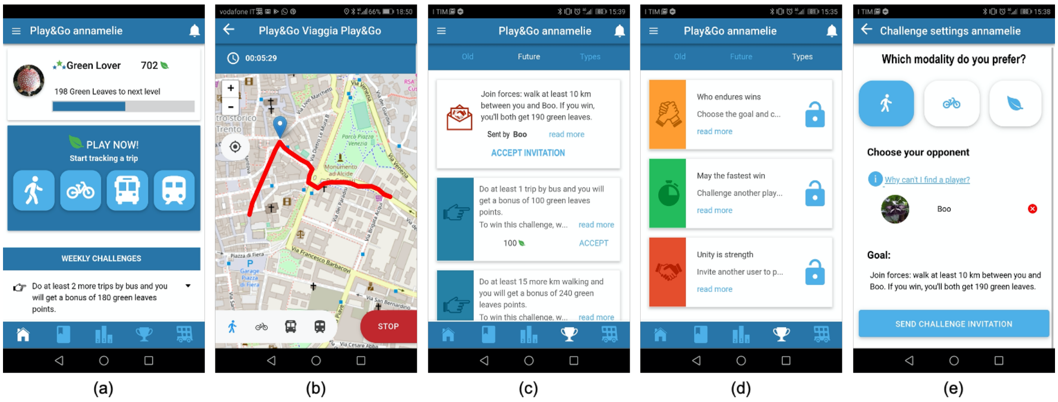


FIGURE 3.1: Screenshots of the Play&Go application representing: (a) home page; (b) the realtime tracking view; (c) the selection of personalized weekly challenge, where players can accept a received invite (if any), send a new invite (if unlocked) or choose a single-player challenge, (d) the unlockable multiplayer challenge modes; and (e) the creation of a new invite. Please note that the figures refer to the last edition of the game analyzed (2019).

in which citizens can compete for physical prizes offered by local partners (e.g., museum tickets). Prizes are awarded weekly to best-performing players within the game week and at the conclusion of the gamification campaign to the top players.

Players can earn *Green Leaves* points by tracking their trips around Trentino through the app, by specifying whether they are walking, riding a bike, or using a public transportation mean (bus or train). They can visualize their trip in real-time (Figure 3.1b) and switch transportation modality as needed. Once the journey ends, players can stop the tracking, and an automatic validation algorithm analyzes the trip. A trip is deemed valid if the declared means is reasonable, considering its time, location, and speed. For instance, bus trips are cross-validated with bus timetables, and acceptable walk trips are under an estimated velocity. Valid trips are translated into *Green Leaves* points, whose number is determined by the length and the eco-sustainability of the trip - i.e., bike trips are more ecological than bus trips.

Figure 3.1a shows the Play&Go home page where users can visualize their current status, start a new tracking, and access their profile. A sense of mastery is conferred to players by associating them with a level that describes their experience in being a green citizen (e.g., Green Soldier, Green Warrior, Green Guru). Green Leaves points also have the meaning of experience points and are used to reach a new level. Hence, on the home page, users can visualize their progress. Players can compare their level of expertise - i.e., points - with others through weekly and global leaderboards. Weekly leaderboards consider players' activities within the specific weeks, whereas global leaderboards report point gathered throughout the gameplay. The purpose of weekly rankings is to include fairness towards late-comers in the gamification campaign, which may be penalized in the global ranking. The leaderboards also have a strong extrinsic value as they are used to select the winners of the sponsored physical prizes.

From the leaderboard (or through the search features), players can access other users' profiles and visualize a high-level overview of their status - e.g., level, profile picture, and a cumulative value of trips and kilometers recorded by transportation means.

Play&Go also features weekly challenges, to maintain users in a state of flow. Thus, challenges are specifically deployed for each user using a Recommendation System

(RS), taking into account players' mobility habits and performance [112]. An example of a challenge is the following: "This week, do at least 5 km by bike to win a bonus of 120 Green Leaves points". Challenges have a weekly validity and, upon completion, award players additional points, defined in accordance with the challenge's difficulty for the specific player. Through the challenges, players are further motivated to improve or maintain their performance, if already optimal. Yet, only goals possible for the user to achieve are suggested to keep them interested without risking frustrating them.

Weekly challenges are not necessarily automatically assigned by the system. Players can perform a choice among 2 or 3 options generated by the RS. The options can vary in the target and difficulty requested, where the difficulty is a value computed upon the estimated effort the specific player is expected to do. The reward associated with the challenge is proportional to its difficulty. As the game is organized in weeks - i.e., weekly validity of the challenges and weekly leaderboard - there is a specific timeframe within which the choice must be performed. Challenges are activated each Saturday. This selection period of 3 days is called *programming phase*. If users lack in making a selection, the system enables the challenge best ranked by the RS.

Other than consulting the leaderboards, players can more actively engage with other users through two-players challenges. Generally, multiplayer challenges are initiated by one of the two players involved through the in-game invitation system (Figure 3.1d). The invite's sender chooses their opponent among a subset of the players' list, comprising users having at most two levels of distance from the player. This constraint avoids producing highly unbalanced challenges, trivial for one end and frustrating for the other. When creating an invite, players can also specify the parameter upon which the challenge will be evaluated, between walk, bike, or Green Leaves points in general. Having specified the opponent and the parameter (or challenge type), the RS will calculate the best target and reward according to both players' previous performance, with the primary goal of generating a fair and appealing challenge for both parties. Therefore, it may happen that the reward, in case of a win, is higher for one party than the other because the challenge is slightly more difficult for that user. The target, on the other hand, is always the same for the two players. Users are made aware of this possibility to allow them to make an informed choice. Finally, the sender needs to specify the multiplayer mode. Play&Go offers three types of social game elements (Figure 3.1c): cooperative challenges, time-based competitive challenges, and performance-based competitive challenges.

In the **cooperative challenges**, players have a common goal and, no matter the individual contribution, if they reach the goal before the challenge expires, they both win. Since the purpose is to make players work together, the target is common, and the prize is the same for each party - e.g., "Join forces: bike at least 10 km between you and *Player2*. If you win, you'll both get 120 green leaves.". In the **time-based competitive challenges**, the first player who reaches the goal wins. The target is calculated so as that it is challenging enough for both players but still achievable. However, since it is likely to happen that the challenge is easier for one of the players, the reward is proportional to the level of difficulty calculated for each party: - i.e., the strongest players would get a lower reward than the weakest. An example of a time-based challenge is the following: "Bike 4 km before your opponent Player2. If you win, you'll get 100 green leaves. If Player2 wins, she-he will get 150 green leaves". In the **performance-based competitive challenges**, the player with the highest performance in the week the challenge is active wins. Due to the definition of the challenge itself, there is no predefined target. The reward is a booster calculated on the weekly points obtained to foster the best performance possible - e.g., "Walk more km than your opponent Player2. The winner will get a 40% bonus of all the points got during the week.".

Once the player is satisfied with the setting, they can send the invite (Figure 3.1d). During the invitation phase, players can actuate a strategy. For instance, they may challenge a less experienced player in a competitive match to increase their chances of winning the reward. The invite is then sent to the other player, who can accept or reject it (Figure 3.1c). The sender can also decide to cancel the request and make another one. Players can make only one request at a time and, once a challenge has been accepted, the players involved cannot make other invitations or accept other challenges until the following week.

As players progress in the game and increase levels, they unlock new game features. Using unlockable elements has a dual purpose. First, players may be overwhelmed if presented with a series of unknown features. The confusion might be frustrating and lead them to disconnect. Second, some players are additionally motivated by the presence of unlockable content. Reaching a new level, up to Level 6, is a milestone for players and results in new mechanics or elements.

- Level 1 - players receive their first customized individual challenge.
- Level 2 - players can choose one among two customized individual challenges.
- Level 3 - players can choose one among three customized individual challenges.
- Level 4 - players can unlock one among the three modes of multiplayer challenges available - cooperative, time-based competitive, and performance-based competitive challenges. From this point onward, they can invite other players to a multiplayer challenge (Figure3.1d).
- Level 5 and 6 - players can unlock another multiplayer challenge mode.

Finally, players possess a personal blacklist, in which blocked users are stored. If a player blocks another user, then they will not receive invitations from them. We included this possibility to allow players to protect themselves from unwanted repetitive invites to multiplayer challenges, as this is the only in-game social interaction possible. Players can also decide to unblock a user at a later stage.

In this project, we analyzed data from three gamification campaigns:

a. The 3rd edition, from September 2017 to March 2018.

b. The 4th edition, from October 2018 to April 2019.

c. The 5th edition, from September 2019 to February 2020.

The core mechanics remain the same across all editions. However, we introduced multiplayer challenges, the possibility to send invites, and the leveling system in the 4th edition. Although players asked for competitive and cooperative game elements, during Edition 4, very few of them engaged in them, whereas the others manifested an issue of trust and confusing interactions. Up to this edition, players could have one active challenge each week, regardless of the type. A major criticality was making mutually exclusive single-player and multiplayer challenges. Therefore, in Edition 5, we allowed two active challenges per week, at most one of them being multiplayer. Besides, to overcome trust issues, we included an *injection* mechanism: players not having active multiplayer challenges were matched through a simple matchmaking algorithm. Hence, a system-generated multiplayer challenge was assigned to them whenever possible. The matchmaking algorithm is naive, as it matches players with similar expertise and performance. For the remaining players, for which the matching was not possible, the RS generated a second single-player challenge.

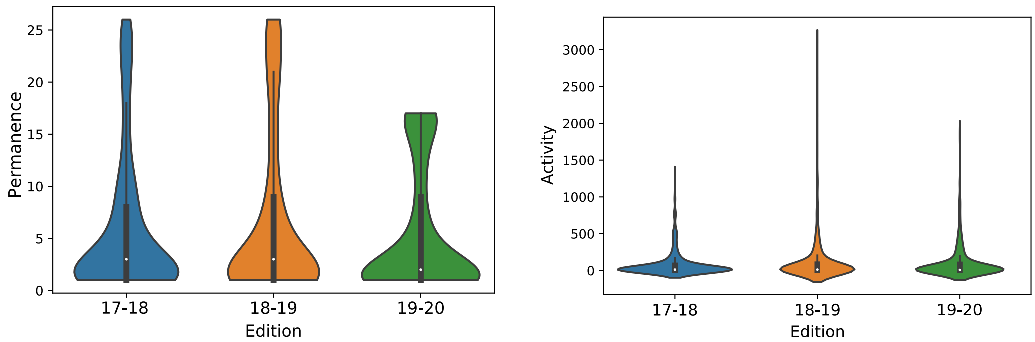
Dataset Description

Play&Go embeds a logging system. Thus, information on players' game actions is stored, as well as the time when the action occurred. Therefore, the raw logs contain detailed data on users' activities, except for the in-app interactions (e.g., visit a specific app page or time spent on a page). Logged actions may be categorized as:

- Information on trips tracked (e.g., transportation means, number of kilometers, and points awarded);
- Information of single-player challenges (e.g., target, reward, and whether it was completed);
- Information on multiplayer challenges (e.g., multiplayer mode, winner, and type);
- Information on invites (e.g., sender, receiver, and whether accepted);
- Information on levels - i.e., the time when a new level is reached;
- Information on unlocked items - i.e., time and type of item unlocked; and
- Information on the blacklist (e.g., blocked users and unblocked users).

We can process and analyze a series of low-level game actions and features through the game's logs. In the following, we will describe some general features used (or combined with others) in the studies conducted. Please note that later chapters may introduce new features specific to the analysis presented. The features can be organized in macro-categories. Players' **activity** is described through the number of trips and kilometers tracked (divided by each transportation means), the number of game actions performed, and the number of *Green Leaves* points collected. Players' activity is also evaluated using the number of active days, where a day is considered active if at least one game action is performed, and the `frequency of usage` value. The *frequency of usage* value is computed as the percentage of active days over the players' gameplay length, computed as the number of days elapsed between the registration day and the last game action. Players' **performance** is represented by the final game level reached; their win ratio in general and divided by challenge's mode (i.e., single-player, cooperative, and competitive); the average difficulty for single-player challenges; and the positions in the weekly and global leaderboards. Players' **social interactions** are modeled through the number of recommendations sent to friends and family; the number and type of invitations sent; the number and type of invitations accepted; and their activity on the blacklist. Players' **agency** is described by the percentage of single-player challenges programmed; and the percentage of multiplayer challenges for which they are initiators.

In the remainder of the manuscript, we will analyze both the feature values aggregated over the whole gameplay and aggregated over each game week. In the first case, referred to as *aggregated data*, for each player, a single data point exists, whose dimension N is the number of features used in the study. In the latter case, referred



(a) Permanence (b) Activity

FIGURE 3.2: Distribution of players' permanence in the gamification campaign (number of weeks spent playing) and their activity (counted in number of game actions performed). It should be noted that the last edition was shorter than the previous two due to the covid19 emergency.

to as *temporal data*, for each player i there are t_i data points of dimension N, where t_i is the number of weeks the player was in the game.

For conciseness's sake, we hereby report basic statistics on the players' population across the three editions of Play&Go analyzed. In the following chapter, relevant features will be described in more detail. Exception made for the 5th edition, interrupted due to the Covid19 emergency, each gamification campaign lasted 25/26 weeks. The players' population size slightly decreased over the years, from 642 registered users in the 3$^{\text{rd}}$ edition, to 587 and 425 in the 4$^{\text{th}}$ and 5$^{\text{th}}$ respectively. Despite this decrease, players' activity and permanence increased (Figure 3.2). With the passing years, users were more likely to perform more game actions and remain in the game for longer. Hence, the game found consensus in the majority of its users.

At the beginning and conclusion of each gamification campaign, a survey is submitted to users in the form of a special challenge. The *registration survey* is built of two parts: general information on users and the Hexad User Types Questionnaire [215]. General information comprehends basic demographics (age and gender) and estimates of the own mobility habits (e.g., preferred transportation means and average weekly kilometers traveled). The Hexad User Types Questionnaire is used to model each user according to the player types defined in [153] - i.e., Philanthropist, Socializer, Achiever, Player, Free Spirit, and Disruptor. The *final survey* collects impressions and feedback from users. For instance, players can evaluate game elements, declare whether the game impacted their transportation behaviors, and manifest their will to participate in the following campaigns. The survey also foresees a free-text box for unstructured comments.

In every edition, we found the players very well-distributed among males and females, both representing half of the population, with a variation of few percentile points. Players' age distribution was also consistent throughout the years. Two-thirds of the players were in the age range 20-35 and 35-50 years old, one-third for each category. Younger (16-20 years old) and older (50-70 years old) users were less present, amounting to one-fifth of the population each, on average. A very small minority of the users were older than 70 years.

The distribution of the player types scores also remained very consistent across the editions, as Figure 3.3 shows.

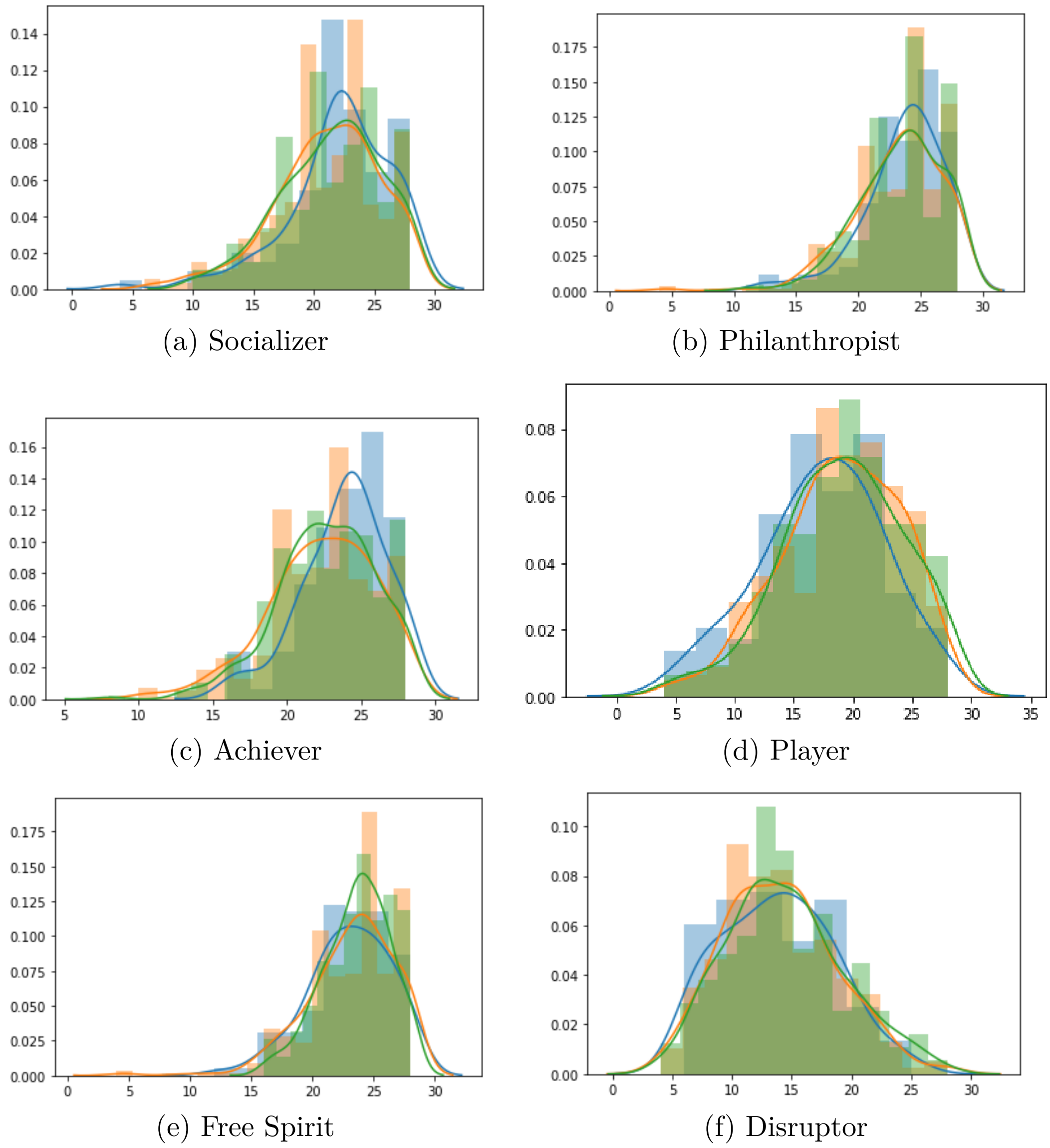


FIGURE 3.3: Distribution of the Hexad User Types scores across the three editions of Play&Go analyzed: 2017-2018 (blue), 2018-2019 (orange), and 2019-2020 (green).

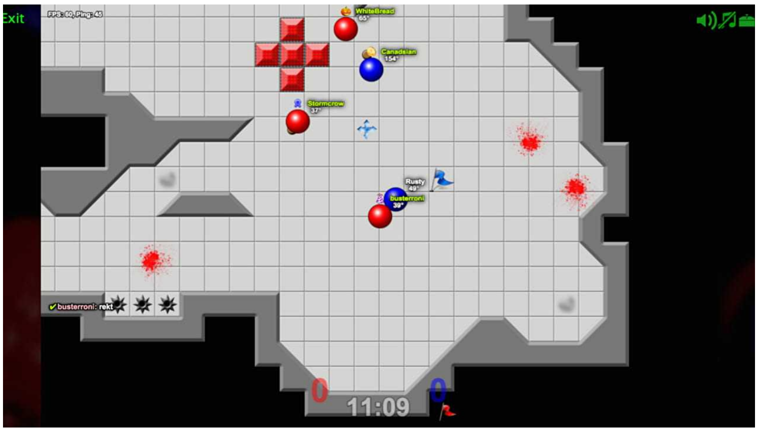


FIGURE 3.4: Gameplay screenshot of the game TagPro.

3.2 Games-related Data

Gamification is indeed the primary application domain for this project. However, during our analyzes, we also exploited data coming from games. This section describes the scope and dynamics of the system analyzed: two online games belonging to different genres and a game provider.

3.2.1 TagPro

TagPro[1] (see Figure 3.4) is an online, multiplayer, casual Free-to-Play (F2P) created by Nick Riggs. The game is a classic desktop game, available online, and was first released in February 2013.

Players are organized into two teams (red and blue) and, to win, have to capture the opponents' flag. Each team can include up to four players, and the game characters can either be controlled through the WASD or the arrow keys.

The game offers two modes, distinguished by the flag type. In the traditional mode, the two teams are spawned to the opposite sides of the map, and each has its own flag. Their goal is to steal the opponents' flag and bring it to their team general quarters (the spawn point). In the alternative mode, instead of the two flags, there is only one neutral flag. The teams compete to be the first that bring the flag to their home–base.

The simple core mechanic of *"capture the flag"* games is complicated by power-ups and other map elements. Game elements exist to either bring benefit to the own teams or hinder the progress of the enemy team.

For instance, the player carrying the flag can be "popped" by the enemies, which results in an invisible explosion. Following the explosion, the flag is brought back to its original location - e.g., the team home–base, in the traditional mode, or the neutral point, in the alternative mode. In the traditional mode, players also "pop" if they are both flag carriers and touch each other, unless they gas a power-up. Whereas in the alternative mode, when a player touches a carrier, the carrier pops, and the flag is transferred to the opponent.

[1] https://tagpro.koalabeast.com/

Table 3.1: Temporal statistics of the Destiny dataset

Property	Value
Observation Period	09 Sep 2014 – 11 Aug 2015
Players	10 037
Matches	26 120
#Snapshots (days)	336
#Snapshots (weeks)	48
#Snapshots (months)	$\sim$ 12 (4 weeks each)

The winner team is the one that first reaches a score of 3 points differential. A point is obtained when a team brings the opponents' flag to their general quarter. For the point to be awarded, the flag carrier of the enemy team should be popped before. In other terms, when the flag is brought to the home–base, the opponent team should not be in possession of their flag. This rule favors strategies and teammates' organization in specific roles. For example, teams often feature *offenders* that tackle the opponent flag carrier intending to pop them once their team stole the other flag.

Dataset Description

TagPro gameplay data–logs[2] store basic information on game matches, such as the composition of the teams, the final score, and the timestamp of the match. As users can also build a profile, in the log, there is also explicated whether the players were authenticated or not. In case of authentication, the player logged in to their profile before joining the match. Only logged players were included in the dataset, as we needed to retrace their activity throughout the observation period.

We retrieved data from September 2015 to April 2016 for a total of 6 months (25 weeks). Our population size was 75k players, participating in 3 game actions, on average (std 3.78).

3.2.2 Destiny

Developed by Bungie, *Destiny* is a popular online multiplayer first-person shooter (MMOFPS) video game. Although Destiny is frequently described as an FPS, it includes many role-playing elements (RPG) and massively multiplayer online (MMO) games. The game is located in a multiplayer *"shared-world"* environment, as in MMOs, and allows the development of a character throughout the gameplay, as in RPGs.

Players take the role of a Guardian and must protect the last safe city remained on Earth from dangerous alien species. Guardians travel across different planets to fulfill their mission, find and exterminate alien enemies for humanity's salvation. Guardians are also tasked with the research of a celestial white body, called the Traveler. The ultimate goal is to relive the Traveler, who allowed humans to navigate in the universe.

The gameplay is divided into player versus environment (PvE) and player versus player (PvP) activities. Regardless, competition is a core aspect of *Destiny*. PvP matches exist in *The Crucible*, a hub separated from the principal game world, where players fight in restricted areas. The Crucible features a playlist of PvP modes and

[2] Data is publicly available at `https://tagpro.eu/`

TABLE 3.2: Basic aggregated statistics for the 10k Destiny players across the 48-week observation period.

Variable	Min	25%	50%	75%	Max	Mean	Std
#matches	17	22	35	78	3.8k	83	146.5
score	2	44.45k	75.2k	164.4k	7.3kk	176.7k	326.3k
time played (s)	2.97k	12.54k	20.2k	4.44k	2.1kk	4.74k	82.44k
kill	2	245	398	877	33k	951	1.75k
death	27	245	393	889	46k	934	1.6k
assists	4	84	136	300	16.9k	316	559
average lifespan (s)	540.78	1.24k	2k	4.4k	185k	4.6k	8.2k
character level	10	29	33	34	40	31	3.8

allows a maximum of twelve players per match, depending on the specific mode, such as *Control*, *Clash*, and *Skirmish*. Players are organized into two teams where the goal is to either conquer a map area or kill all the opponents.

Dataset Description

We analyzed data from the PvP Crucible matches in *Destiny*, between September 2014 and August 2015. In our time window of 48 weeks, we retrieved a population of 3M of players. We reduced the population to a sample size of 10k players for computational constraints. The subset of players is constituted of users that were retained for at least five weeks in the game. This temporal constraint derives from the requirements of the study, whose details are outlined in Chapter 5.

The final dataset contained 26k matches PvP Crucible matches. For each match, the following information is available: the teams' composition, the timestamp of the start of the match, how long each player stayed in the match, and basic stats (e.g., kills, assists, and deaths). Table 3.2 shows basic statistics for the players' population, describing their activity within the observation period.

3.2.3 Steam

Steam[3] is an online game provider - i.e., an online library of games, which are usually uploaded by game studios, publishers, or indie developers. The games are showcased in store pages, presenting information on the game, such as plot, description, screenshots, reviews, ratings, tags, and genre.

Users need to register to the platform to access its services. They have their own profile page, public by default, where information on their activity is shown. For instance, games can be played through the platform, and players' activity is visible in their profiles, such as their time playing and unlocked achievements. Players can also review games. A game's rating is described on a scale from overwhelmingly negative (0%-19% of positive reviews) to overwhelmingly positive (95%-100% of positive reviews).

Besides building their game library, users can also connect to other players through friendship requests and join or create groups. Steam groups connect people with shared interests, enabling them, for instance, to discuss specific games or share information. Although the social component is not essential for the core purpose of

[3] https://steampowered.com

Steam - e.g., collect and play games - the community is very active, and thus, it is interesting to research to gain insight on players.

Dataset Description

We collected daily updated information about players' activity in the form of time spent playing on each game they owned. The observation period covered five weeks, from April 13th, 2020, to May 17th, 2020.

Given that we consider a user active if it has played at least one game during the observation period, we found that only $51k$ out of $191k$ players were active during the five weeks we crawled their activity. We removed these users from the friendship graph, along with the nodes that became disconnected. This resulted in a final sample of **39,354** users and **218,432** edges. Although we acknowledge that our dataset partially represents the Steam network, the size of the crawled data was limited by the APIs constraints and for computability reasons, especially related to the *semantic* influence algorithm (see Chapter 5). Furthermore, it is worth noting that we conducted our analyses with careful consideration of players' privacy. In Steam, user ids are public and can be used to retrieve user's profiles online by including the id to the website URL. Therefore, to protect users' privacy and eliminate potential risks, we mapped each user to a new unrelated number. Additionally, Steam protects players' activities and guarantees no disruption of the user experience caused by logging or crawling processes.

Regarding the players' activity, we obtained a final sample of 28 days after removing the outliers caused by an API server failure. Table 3.3 describes the users' level of activity during the observation period. Generally, the players tended to be very active; on average, they played for about 17 days during the observed 28 days ($std = 8.42$). The majority of the players were engaged in few games, and only a few outliers played more than ten games. Users have a mean of about 4 playing hours on each active day, and only a small part of the population played less than ~$2h$. From this, it can be deduced that our population is represented by predominantly regular players who play almost every day.

Games Metadata

During the 5-weeks observation period, our user population played about $17k$ different games. For each game, we scraped its metadata, including the game genre, steam-defined tags, user-defined tags, and reviews trend.

For the majority of the games, very few players in the sample played them. Specifically, the distribution of players per game has a long right tail, with a mean of 25.8 players and a very high standard deviation of 232.8 players ($min = 1, 25\% = 1, 50\% = 1, 50\% = 3, 75\% = 10, max = 26,590$). We observed that only a sixth of the games (~$3k$) had at least 10 players, about 500 games had more than 100 players, 35 games had more than $1,000$ players. Only one game had more than $10k$ players ($17,273$ players in the sample). The game is *Counter-Strike: Global Offensive*, a *Free-To-Play Action Game*, released on August 21st, 2012, with more than 4 million reviews and a general rating of *Very Positive*. Generally, the types of reviews they have received for the game tend to from *Mixed* to *Very Positive* review results, as defined by Steam. Thus, the games played by our population meet the consensus of the entire community. Less and almost irrelevant are the unpopular games.

TABLE 3.3: Statistics of players' activity.

	Min	**25%**	**50%**	**75%**	**Max**	**Mean**	**Std**
Minutes spent playing	0.67	111.29	185.71	282.43	997.50	209.65	132.01
#Games played	1	3	6	11	387	8.87	9.8
Active days	2	10	19	25	28	23	6.5
Active period	1	22	26	27	27	23	6.5

The types of games contained in the dataset are quite heterogeneous. While we can observe how both user-defined tags and Steam-defined tags show a predominance of indie games and single-player games over multiplayer titles, these are not necessarily the games with a higher number of players. On the contrary, we observed that players tend to play different single-players games while they were more homogeneous in the choice of collaborative and multiplayer games. This behavior can be explained by the social component being a core aspect of the multiplayer game at a micro-level and also reflected in the more general players' network at a macro level. In other words, in contrast to individual games, multiplayer games require a community to be fun. Hence, a type of consensus among users is needed for a multiplayer game to be played.

Chapter 4

Understanding Players as Individuals

When they conceive a game, designers have a precise idea of how such a game will be played and how the gameplay would look like. This idea is even more delineated in gameful applications and games with a purpose. Designers have a precise *ulterior goal* they aim at achieving by exploiting game elements and mechanics. However, humans are complex and unpredictable. Consequently, the expectations rarely match reality.

Games analytics can help unravel players' behaviors and interaction patterns to inform and adapt the game's design. Analyses on a global level, observing the population as a whole, can convey knowledge on design faults and unordinary usage. For instance, players' winning strategies may emerge, or cheating behaviors can be detected. Player-centric analyses, on the other hand, can produce a finer-grained understanding of users towards their profiling. Player profiles can be used for the generation of tailored game content aimed at matching players' skills and tastes. Producing content under players' specific preferences has the goal to deliver more enjoyable gaming experiences and retain them for longer, avoiding a premature churn.

Players cannot be treated as a uniform unit. The literature informs us on how humans express their individuality in every action they undertake. Games, which were born to allow people to experiment and learn, are the ultimate self-expression tools. Players choose games that accommodate their motivational drivers and needs and indulge their personality and preferences. As a result, to intrinsically motivate users to play, games need to implement methods aimed at understanding players' personalities, preferences, and behaviors. Recent studies confirm that also gamification benefits from tailored content, hence the popularity increase for Adaptive Gamification. In this chapter, we understand what information can be extracted by analyzing their in-game activity and their in-game behaviors to model players' experience.

4.1 Assessing Players' Experience

The analysis of players' in-game behaviors can produce valuable feedback on their gameful experiences. In this section, we first exploit telemetry data to describe and model the players' population from their behaviors. Those high-level player models allow the identification of game strategies and monitor the oscillation between active and inactive periods throughout the gameplay. After this global investigation on the behavioral trends among Play&Go users, we studied more in-depth players' experience to identify users at risk of churn. This second analysis shows how players' activity within the app can still provide meaningful insights, despite gamified applications being limited in the diversity of analyzable features.

4.1.1 Observing Gameplay

The initial game design and the actual gameplay can diverge, especially in gameful systems, where a precise underlying goal is pursued. For instance, our gameful application (Play&Go) is conceived to promote a behavioral shift towards ecological mobility habits (See Chapter 4 for more details on the app). Players are awarded points for their daily journeys once they track them using the app. Those points have both an extrinsic and intrinsic meaning. Higher scores increase the chances of obtaining real-world prizes but also allow them to gain experience in the game and achieve a sense of mastery. Hence, users are motivated to gain points and *win*. The game designers and promoters, however, want to prize sustainable transportation behaviors. The analysis of in-game behaviors to extract players' winning strategies can verify whether those strategies align with the gamification goal. In this study [145], we process Play&Go telemetry data to characterize the population of players, understand who the winners are, and model changes in their in-game interactions and behaviors.

Datalogs, however, may be unreliable in inferring players' behaviors due to noise [91], which can be contained if the appropriate methods are used [54]. The main issue is whether to process aggregated game data, as frequently done in behavioral analysis [54, 202], in contrast to temporal data, acknowledging the dynamicity of gameplay [57]. Aggregated data describe players using a single data point, averaging their activity within the observation period. Conversely, temporal data represents users through many data points, collected at regular intervals.

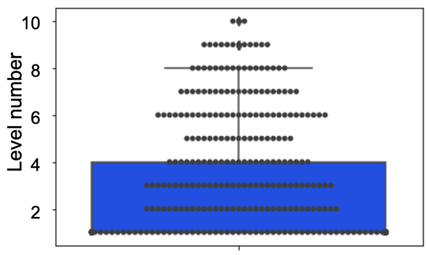


FIGURE 4.1: Boxplot of players' level distribution.

Dataset

For this study, we analyzed data from the last Play&Go gamification campaign, held between November the 2nd, 2019, and February the 28th, 2020.

In our 17-week-long observation period, 425 players registered to the game and tracked at least one trip. Among those players, 248 remained in the campaign for more than one week, but only 194 reached Level 2. We analyzed a subset of those, with a maximum level between 3 and 8, amounting to 145 users. The lower bound of 3 was set to exclude players not having access to multiplayer challenges, while the upper bound of 8 was included to remove outliers (show in Figure 4.1).

Our sample of players was fairly active (Figure 4.2). On average, they played 11 weeks (std = 5.2) over the 17 analyzed. Players performed 166.5 actions (std = 159), on average, with a medium global score of 561 points (std = 271).

To characterize and model the recurrent behaviors in the population of players, we used the following features: 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).

Data is organized into two datasets, both with the same features. The aggregated dataset contains one entry for each player that averages their activities throughout their gameplay. The temporal dataset, on the other hand, is built of multiple entries for each player, representing a snapshot of their activity within a specific interval. Specifically, we consider weekly snapshots, as the gamification campaign is organized in weeks (see Chapter 3).

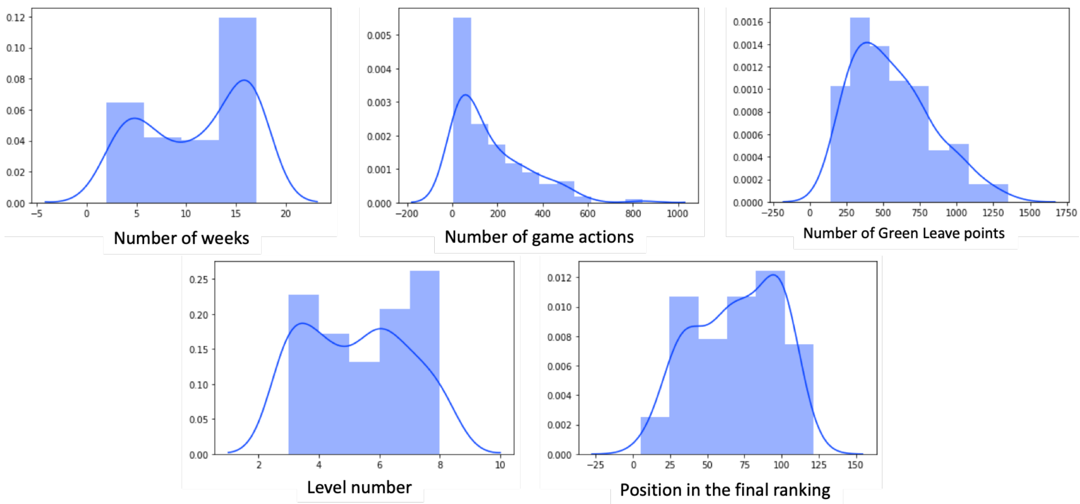


FIGURE 4.2: Activity distribution in the selected sample of users .

Modeling In-game Behaviors

Players' in-game actions (or behaviors) were modeled using a soft-clustering technique: Archetypal Analysis (AA). AA is a matrix factorization method that produces a description of each data point as a combination of archetypes, where the archetypes represent extreme points in the dataset. For a more detailed description of AA, please refer to Chapter 2.

Semantically, archetypes can be interpreted as traits or behaviors, and the data points are represented through a combination of those traits or behaviors. In the analysis of the aggregated dataset, a data point corresponds to a player, while in the analysis of the temporal dataset, a data point represents how a player behaved during a specific game week. Hence, a player is described by many data points.

In the following, we present the algorithm's archetypes, as classically done in related research [203, 195, 185]. The archetypes have also been inspected and validated by the game designers, exploiting their domain knowledge.

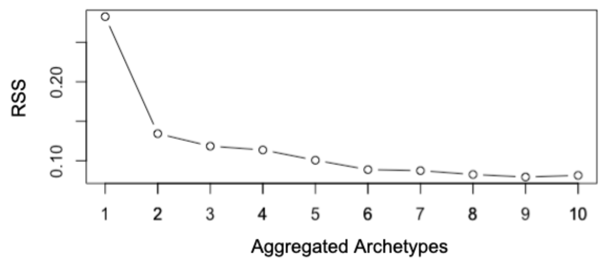


FIGURE 4.3: RSS using a different number of archetypes for the aggregated analysis.

Aggregated Analysis. We obtained the first set of archetypes by running AA on our aggregated dataset. AA, like many other clustering techniques, requires the number k of clusters - in this case, archetypes - to be set a priori. The approach typically used is the 'elbow' method, consisting of running the algorithm with different values for k. The results are then plotted against a performance metric, and the k value is selected in correspondence with an 'elbow', which can be visually identified in the plot. Figure 4.3 displays two elbows: for 2 and 6 archetypes. Using two archetypes produced a division into active and non-active players. Therefore, we retained six archetypes, later referred to as the set $\mathcal{A}$.

A1: *Ghost* are characterized by very few game actions. Their in-game activity is almost none.

A2: *Hostile* are characterized by an elevated number of actions on their blacklist. Their travels are multimodal, and their activity is below average. They also tried customizing single-player challenges, with a low winning rate despite the medium difficulty, and few multiplayer challenges, which were never won.

A3: *Loser* are characterized by their very low winning rate in single-player challenges, despite being easy. Yet, they are fairly active in terms of the number of game actions performed. They show a preference for the bus and walking transportation means but have also tracked trips with public transportation means. Those players have tried the challenge programming mechanism and the invitation systems on a small scale.

A4: *Average Full-feature* are characterized by being active in every available game element and mechanics. They intensively use all the available transport modes. They sometimes program their challenges, either of medium or hard difficulty. 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* are characterized by a tendentially low level of activity. They almost never interact with other users but programmed a few single-player challenges with medium difficulty and won them almost half of the time. They mostly walk or use the bus.

A6: *Green Socializer* are characterized by a strong preference and high activity in green transportation means (walking and biking). They have a good winning rate of single-player challenges, often programmed to be of either medium or hard difficulty. They tend to send invites rather than receiving them, with a preference for cooperation rather than competition.

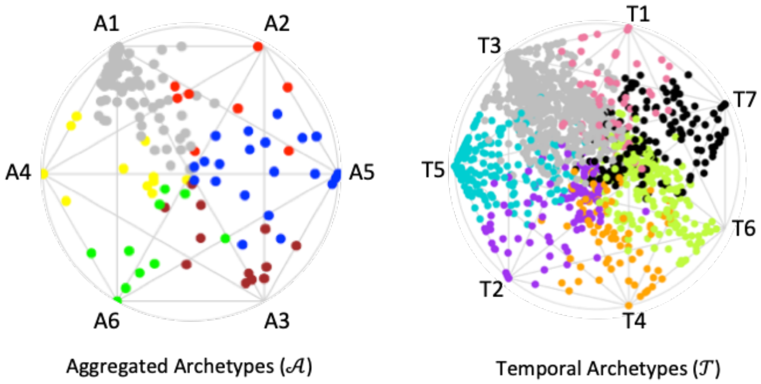


FIGURE 4.4: Scatter plot of the belongingness scores of the aggregated and temporal archetypes. The vertices represent the archetypes, while the dots represent the players. The closer the dot is to a vertex, the higher the value for that archetype. The color of the dot is determined by the dominant archetype value.

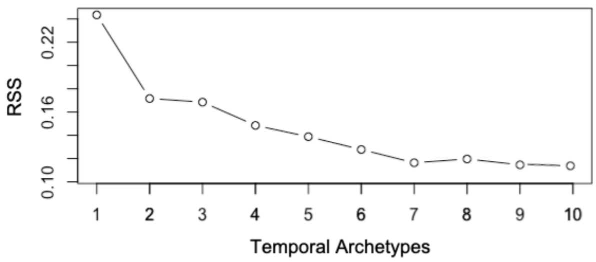


FIGURE 4.5: RSS using a different number of archetypes for the temporal analysis.

The leftmost plot in Figure 4.4 shows the belongingness scores distribution for each aggregated archetype (in $\mathcal{A}$) over the 145 players. The *Ghost* (A1) archetype characterized a big portion of the users, followed by the *Sporadic User* (A5) archetype. On the other hand, the *Hostile* (A2) archetype, in comparison to the others, is scarcely populated. The remaining more active archetypes - i.e., A3, A4, and A6 - show similar distributions.

Temporal Analysis. Aggregated analysis extracted and characterized players' behaviors throughout their gameplay. With temporal analysis, on the other hand, we characterize in-game activity on a weekly basis. Towards this, we used the temporal dataset, in which the entries represent players' weekly actions. This dataset contains several weeks for each user. To eliminate data dependency, we run AA on a subset of the dataset, considering activity in game-week 5. We choose this week because it featured the highest number of players (104). Then, we computed the belongingness scores for every excluded entry according to the definition of the archetypes found. Similarly to the aggregated analysis, we selected the number of temporal archetypes using the 'elbow' rule on the archetypes screeplot. We found two elbows, corresponding to 2 and 7 archetypes. We choose seven archetypes for the same reason. We refer to this set as $\mathcal{T}$. A description of the archetypes follows.

T1: *Wannabe Competitive Achiever* are characterized by programming hard single-player challenges, won half of the time, and by sending many competitive invites. Although their competitive invites are almost never accepted, they receive many cooperative invites, which they accept. Those players are active and multimodal, with a preference for green transportation means (bike and walk).

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

T3: *Ghost* are characterized by very few game actions. They in-game activity is almost none.

T4: *Average User* are characterized by a moderate level of activity. Their trips are multimodal, with a preference for walking and bus. They program a few easy single-player challenges, seldom won. They receive, but rarely accept, few invites (both cooperative and competitive). Their win rate of multiplayer challenges is also low.

T5: *Social Initiator* are characterized by their strong involvement in multiplayer challenges, where they are the initiators. Their invites acceptance rate is high, although the win ratio is medium to low. They are moderately active, and their trips are usually multimodal, despite a preference for walking. They often program their individual challenges with medium-to-hard difficulty and maintain a good win ratio.

T6: *Green Loner* are characterized by having no interaction with other players. They are moderately active and have a strong tendency towards green transportation means. They tend to program medium to hard challenges, which are always won.

T7: *Just Enough* are characterized by a low activity level, sufficient to win their challenges (easy and seldom programmed). Their trips are mostly walked or tracked by bus. They have no interaction with the invitation system.

Having defined the temporal archetypes $\mathcal{T}$, we computed the belongingness score for the excluded entries - i.e., the remaining weeks for all 145 players.

The rightmost plot in Figure 4.4 shows the distribution of the belongingness for each temporal archetype over the 145 players and the 1590 entries. The *Ghost* archetype is also one of the predominant archetypes, as in the $\mathcal{A}$ set. Differently to the aggregated set, however, the more active archetypes are visibly more populated. The least dense archetype is the *Wannabe Competitive Achiever*, representing a small portion of the population. Finally, Figure 4.4b shows how the majority of the data points are positioned at the center of the figure. Hence, a single archetype hardly characterizes most of the users.

Players' Winning Strategies

In the description of the archetypes, we provided a preliminary visual comparison of the two 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. In the following, we expand those comparisons by researching correlations among the belongingness scores players obtained in the two groups of archetypes. We used Kendall's τ, as the scores were not normally distributed [93], and the Gilpin's correspondence table [72] to interpret

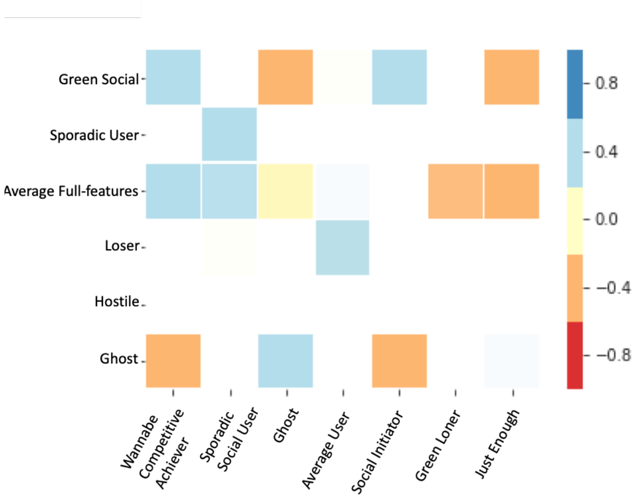


FIGURE 4.6: Heatmap of the correlation among the two sets of archetypes. The image overlays all the images of the 17 analyses. White cells represent values non statistically significant (p-value>.05). The p-values are corrected through the Benjamini-Hochberg method.

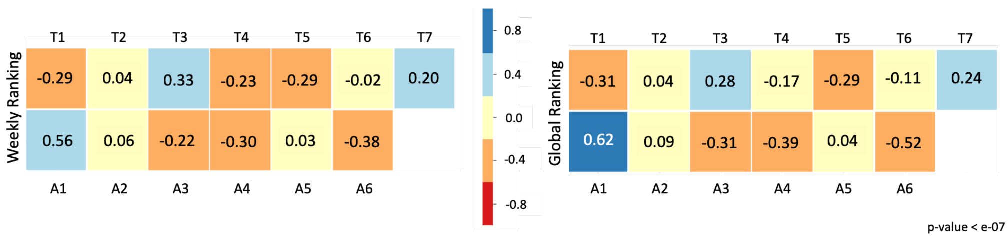


FIGURE 4.7: Heatmap of the correlation among the set of archetypes and the position in the weekly and global rankings. White cells represent values non statistically significant (p-value > .05).

the effect. 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.

To avoid data dependency, we analyzed one week of the temporal dataset at a time. 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., the number of gameplay weeks) analyses and treated multiple correlations through Benjamini-Hochberg correction [40]. Figure 4.6[1] shows the resulting correlation heatmap.

We found a strong correlation among the low-performing archetypes (*Ghost* (A1, T3) and *Sporadic User* (A5, T2)). While, some aggregated archetypes correlated with a couple of temporal archetypes, as they were a combination of those. For instance, *Green Socializer* (A6) positively correlates to *Wannabe Competitive Achiever* (T1) and *Social Initiator* (T5), and *Average Full-feature* (A4) positively correlates to *Wannabe Competitive Achiever* (T1) and *Sporadic Social User* (T2). We also found how some specific archetypes existed in one representation and were completely absent in the other. 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. Therefore, the two sets of archetypes are related, although providing additional details on some aspects of the gameplay.

We further investigated how different is the information conveyed by aggregated and temporal archetypes in the analysis of players' winning strategies. Towards this, we correlated the players' belongingness scores in each archetype to their position in the weekly and global rankings (Figure 4.7).

In the $\mathcal{A}$ set, we found that the *Average full-feature* (A4) and the *Green Socializer* (A6) archetypes assumed high positions in the weekly rankings (with a correlation of -0.3 and -0.38, respectively). In contrast, high positions in the global leaderboard were occupied by users with higher belongingness scores in the *Green Socializer* (A6) archetype (-0.52). When investigating the $\mathcal{T}$ set, the *Social Initiator* (T5) and the *Green Loner* (T1) archetypes were dominated 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.

Behaviors' Evolution

The archetypes featured different levels of activities and interaction patterns, both in the $\mathcal{A}$ and $\mathcal{T}$ sets. In this paragraph, we researched frequent behavioral changes by analyzing temporal data. To reduce complexity, we characterized players' weekly activity using their dominant archetype in that week. Then, we studied recurrent transitions among archetypes (or behaviors). In other terms, we collected more information on how players interact with the game by modeling the frequency of those behavioral transitions.

The transitions were modeled using a directed weighted graph (Figure 4.8), where the nodes are the temporal archetypes, and the edges $e = (u, v)$ represent a shift from

[1]The colors of the heatmap scale have been chosen to be color-blind friendly for accessibility reasons.

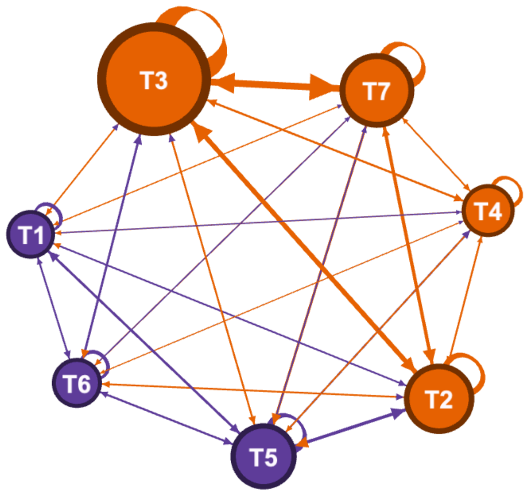


FIGURE 4.8: Graph of transitions among archetypes.

an archetype u to an archetype v. The weight of the edge shows how frequent the transition was among players. The size of the nodes in Figure 4.8 is proportional to their weighted in-degree value and the thickness of the edges to their weights. The graph representation shows how many connections are directed towards the *Ghost* (T3) archetypes. The active archetype more frequent, on the other hand, was *Social Initiator* (T5).

Then, we partitioned the graph using the modularity score, computed upon the strength of the ties. As a result, we obtained two communities identified in the graph with two different colors. The first community includes the most active archetypes: T1 (*Wannabe Competitive Achiever*), T6 (*Green Loner*), and T5 (*Social Initiator*). The other community is mostly composed by low-activity archetypes: T3 (*Ghost*), T7 (*Just Enough*), T4 (*Average User*), and T2 (*Sporadic Social User*). This suggests that, in most cases, players tend to maintain their level of engagement throughout the game. Therefore, 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.

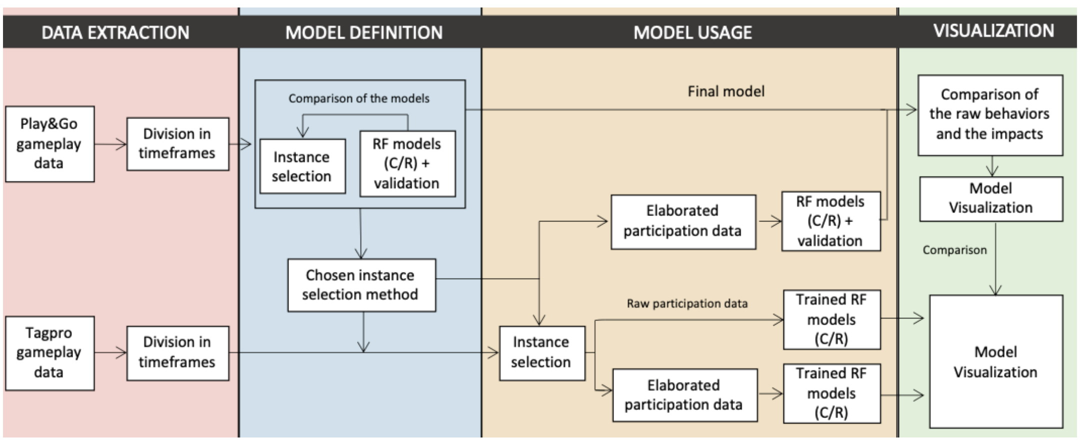


FIGURE 4.9: Visual representation of the analysis workflow for the churn prediction study.

4.1.2 Churn Prediction

In the previous study, we characterized the population of players by analyzing their interaction patterns conveyed information on their behaviors, winning strategies, and how they shifted among the behavioral archetypes during the gameplay. Those analyses highlighted the presence of unengaged users, oscillating among low-level activity archetypes. This finding raised a warning for designers and developers. Yet, no additional information is provided on how to anticipate players' disengagement and prevent them from abandoning the game. Timely identifying players at risk of churn - i.e., abandonment - allows introducing ad-hoc intervention strategies to foster players' retention.

Churn prediction has widely been investigated in entertainment games, where a rich set of features can be exploited to inform the prediction. Gameful systems, on the other hand, often rely on simple interactions, hence lack complex attributes or properties. In this study, we tackle the problem of predicting players' churn in our gamified app, Play&Go, using only participation data. We show how information on players' in-game activity is sufficient to predict their churn [143, 139], and investigate different methods for data preprocessing and the analysis setting [139]. Finally, we research the generalizability of the findings and the models by using as validation set data from a F2P game (TagPro) [139]. Figure 4.9 shows the analysis workflow.

Constructs and Metrics

In this study, we collected and analyzed telemetry data from the 3rd and the 4th editions of Play&Go to model players' level of participation. In practice, we measured the *intensity* and the *frequency* of players' activity in the game. The intensity of usage is evaluated using the `number of Actions` performed and the `Green Leaves points` obtained by the player. The `Active Behavior` evaluates the intensity of their participation. To this extent, we separately calculated the number of points gathered and the number of actions performed during each week. The `Frequency of Usage` indicator evaluates players' commitment (`Committed Behavior`) to the game, as the number of active days over the gameplay length.

Similar to our previous studies, we modeled the temporal aspect of the gameplay by fragmenting it in daily and weekly time windows to "acknowledge the dynamic, cyclical nature of gamification" [121]. Those time windows, referred to as **timeframes** in the remainder of the manuscript, allows us to evaluate players' progress punctually. We also measured players' time investment in the game as the total amount of time they spent in the game (`Gameplay time`). Finally, we computed our dependent variable in the churn prediction analysis (`Churn Time`) - i.e., when and if a player would abandon the game. The Churn Time value lies in the $[0; 1]$ interval and is calculated as the number of timeframes the player remained in the game over the total number of timeframes left before the game ended. If the player never abandoned the game, Churn Time was 1; in contrast, if the player left in the following timeframe, the churn time was 0.

The Play&Go dataset counted 13k players, for whom we had data for up to 25 weeks - i.e., the length of the campaigns. Table 4.1 shows the distribution of the values of the features, as an aggregate value across players' gameplay, and the dependent variable, Churn time. Besides gamification data, we also collected telemetry data from a F2P online game (TagPro), used as the validation set. TagPro players were 75k, analyzed in an observation period of five months. Table 4.2 shows the features distribution.

TABLE 4.1: Distribution of the values of the features in the Play&Go dataset as aggregated throughout their game experience.

Features	min	25%	50%	75%	max	mean	std
Frequency of Usage	0.04	0.04	0.06	0.1	0.34	0.07	0.04
Number of Points	3	64	100	117	522	96	50
Number of Actions	1	1.3	2	4	18	3	1.6
Frequency of Usage (Impact)	-0.9	-0.08	0.16	0.49	1	0.2	0.38
Number of Points (Impact)	-0.5	-0.43	-0.4	-0.36	-0.03	-0.39	0.06
Number of Actions (Impact)	0.01	0.3	0.2	0.86	0.99	0.38	0.39
Gameplay Time	0.01	0.02	0.1	0.73	0.99	0.33	0.38

TABLE 4.2: Distribution of the values of the features in the casual game TagPro dataset as aggregated throughout their game experience.

Features	min	25%	50%	75%	max	mean	std
Frequency of Usage	0.04	0.04	0.04	0.08	0.97	0.06	0.03
Number of Points	-306	28	58	117	8916	99.34	150
Number of Actions	1	1	2	4	291	3.1	3.78
Frequency of Usage (Impact)	-0.64	-0.04	-0.04	-0.38	0.95	-0.4	0.1
Number of Points (Impact)	-0.79	0.48	0.7	0.86	1	0.65	0.27
Number of Actions (Impact)	-0.59	-0.3	-0.25	0	0.95	-0.2	0.3
Gameplay Time	0.01	0.01	0.1	0.3	0.9	0.2	0.18

Measuring the evolution of players' behaviors

Besides researching whether raw participation data is sufficient to predict players' churn, we also investigated the elaboration of those values embedding information on players' improvement (or worsening). Towards this, we define a generalizable function, called **impact function**, as it measures the *impact* the game has on players concerning a specific behavior or feature. The values produced by the impact function are called **impact values**.

Each participation feature, as previously described, is separately processed by the function as a separated *indicator* of participation. The outcome is negative when the participation is reduced, positive otherwise. Specifically, the `FoU` indicator describes the `Committed Behaviors`, while the `GA` and the `GP` indicators describe the `Active Behavior`.

We define the *impact function* as:

$$f : [0,1] \times [0,1] \rightarrow [-1,1]$$

The function takes as input the couple $(baseline, current)$. The baseline is the value of the indicator in the previous time window, whereas the current is its present value. The domain of the function lies in the interval $[0,1] \times [0,1]$, as the input values can be normalized without loss of generality. The goal is for players to have impact values close to 1.

Our function is initially defined by five *tuning points*. As commonly done in works presenting a utility function (e.g., [220]), we assume this function is specified by the

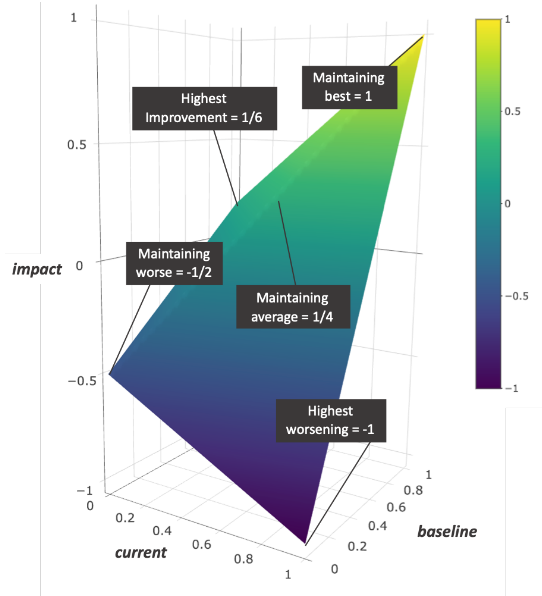


FIGURE 4.10: Our instance of the impact function.

user using their domain knowledge. Those points can be used to guide the behavior of the function by specifying the desired values for key situations: Highest improvement, Highest worsening, Maintaining best, Maintaining average, and Maintaining worse. As a result, the designer or analyst can manage the importance of key events. For example, higher importance could be assigned to maintaining the improvement instead of showing sudden peaks in performance. The function was then obtained by interpolating the tuning points.

The five points are defined by the pairs $(baseline, current)$:

- **Highest improvement** when the player, from doing nothing ($baseline = 0$), reaches the highest possible value ($current = 1$).
- **Highest worsening** when the player, from achieving the best performance ($baseline = 1$), does nothing ($current = 0$).
- **Maintaining best** when the player keeps the best performance ($baseline = current = 1$).
- **Maintaining average** when the player performance are average ($baseline = current = 1/2$), but neither an improvement or a worsening occurs.
- **Maintaining worse** when the player does nothing and keeps in doing so ($baseline = current = 0$).

In the current study, the values were defined as follows. **Highest worsening**, the lowest value, was set to -1, while **maintaining best** was set to 1. Players were penalized for being inactive by setting the **maintaining worse** value to $-1/2$. We also slightly penalized players active every other timeframe by setting the **highest improvement** value to $1/6$. Finally, the **maintaining average** value was set to $1/4$ as a mild reward for players maintaining an average performance but lacking in ambition. A visual representation is shown in Figure 4.10.

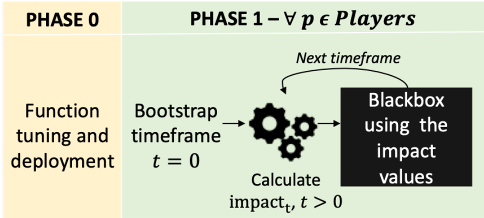


FIGURE 4.11: Macro-phases of the function employment.

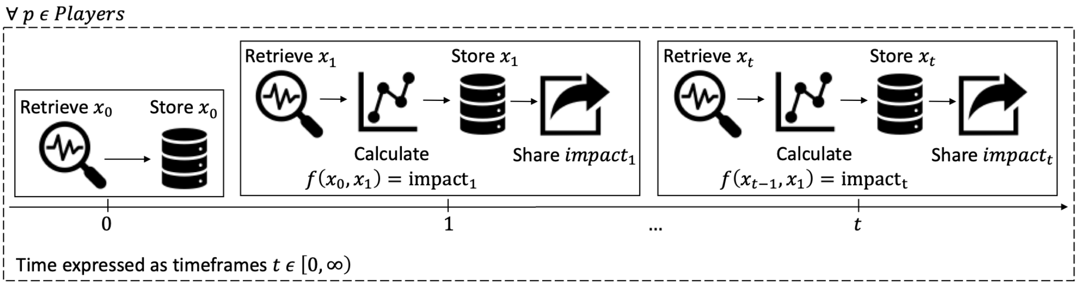


FIGURE 4.12: Usage of the function over time, organized in timeframes t.

The definition of the function tuning points and deployment is the first step of the function employment (Phase 0 in Figure 4.11). The second phase concerns its usage at runtime. The impact values are iteratively measured every timeframe (Phase 1 in Figure4.11). The behavior of the function in this phase over time is described in Figure 4.12. The function values obtained at timeframe 0 are used as a baseline. In the subsequent timeframes, the indicator values are retrieved and the impact calculated. The impact values are then stored and used in the new iteration.

The impact function comes with the following benefits. First, the values are adjusted to players' specific performance and behaviors, as each provides their personal baseline for the computation. As a result, the problem of contaminating the outcomes with outlier values is reduced. Second, the metric is *dynamic* metric and continuously measures players' behaviors. Hence, it is sensitive to changes. Finally, the function can be adapted to the domain and the system's goal by modifying (a) the indicator used and the tuning points defined.

Instance selection method

In this study, we analyzed players' in-game activity over time. Hence, the Play&Go and TagPro datasets comprised multiple entries for each player, describing their behaviors within a specific time window. A dataset record, or data point, is an instance. The instance takes the form of an array, of size n, with n being the number of features. In other words, players are described by multiple n-sized vectors in the dataset. Consequently, the datasets suffer from data dependency, which can affect the accuracy of the predictions due to introduced bias. Instance selection, or sampling, can help to minimize data dependency [130]. The prediction model can be more accurate and reliable once noise and redundancy are removed [10]. In the literature, widely used instance selection methods are simple random sampling, stratified random sampling, and adaptive sampling [130]. We used and compared one stratified random sampling and two types of adaptive samplings. In the stratified random sampling, for each user, an entry is randomly selected among the records available. Whereas

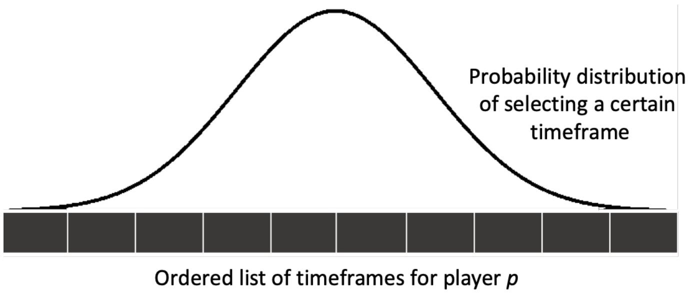


FIGURE 4.13: Representation of how the probabilistic selection of timeframes works.

the adaptive sampling we defined are: (a) the *Gaussian Selection*, and (b) *Uniform Selection*. In the Gaussian Selection sampling method, an entry in the player history was retrieved following a probability distribution described by a Gaussian curve (mean = 0, std = 1), shown in Figure 4.13. This sampling had the purpose of giving a higher priority to records in the middle of players' gameplay and limit the chances of selecting timeframes at the extremes. Penalizing entries referring to the first and the last weeks of gameplay are less relevant in a prediction analysis than records regarding the middle of the experience. In addition, limiting data from early weeks helps to ensure the heterogeneity of the sampled entries, which are more frequent in the dataset. Finally, in the Uniform Selection approach, we retrieved the same week for each player. This sampling is aimed at limiting biases introduced by external co-funding factors (e.g., environmental) and ensure an analogous global game status.

Predicting Churn

Previous analysis on players' behaviors for churn prediction [143], using an ANN model, showed how in-game activity could be a predictor of players' abandonment. In this analysis [143], we accommodated the game (Play&Go) structure and defined the timeframe length to one week. Then, we researched how the number of weeks considered in the prediction impacted the accuracy of the results by analyzing 2, 3, and 4 weeks memory.

In the present study, the memory used in the analysis was drastically reduced to one-timeframe memory—meaning using data only from the immediately preceding timeframe. We applied another change. Instead of using 1-week time windows, we used 1-day timeframes, as weeks may be relevant in Play&Go, but have no meaning in other domains. The final modification concerned the prediction model used. From an ANN approached, we moved towards a tree-based approach, which is preferred in churn prediction studies in games [179, 76]. The advantage is in the production of smaller and more informative models [147, 76]. Among the existing tree-based models, Random Forest (RF) proved to be particularly effective [42] and fast to train [204], similar to ANN performance [4].

We conducted both regression and classification studies. Specifically, in the regression analysis, we predicted `Churn Time`, a value in the interval $[0; 1]$, describing when and if a player will leave the game. We decided to perform a regression rather than a survival analysis because there is no censored data [65]. We analyzed data from the complete gamification campaign. Therefore there are no users "dying" after the end of the observation period. In the classification analysis, we predicted whether players churned in the form of a Boolean value: true or false. Churners (churn = True) were users whose churn time value was higher than the classification threshold.

Although the threshold is highly context-dependent[242], we used the common value of0.5. Studying the impact of different thresholds is out of the scope. Yet, it should be addressed in future works. We also acknowledge that this threshold is not ideal in an imbalanced dataset like ours. However, we resolved the issue by using an oversampling technique, discussed in the remainder of the section.

The prediction analyses were performed using the Python sklearn implementation, with 10-Folds cross-validation. The forest size for the RF algorithm was 100 trees, empirically defined [196]. As we anticipated, the dataset presented a modest imbalance in the values of the dependent variable. Hence, we improved the model's performance [101, 45], using an oversampling algorithm [39]: SMOTE. SMOTE performs very well, also for very imbalanced datasets [19] and when outliers are present [77]. In addition, SMOTE variants for regression analysis exist [220, 30], although being a complex and young area of research [125]. SMOTE for regression, or SMOGN [29], allows making the classification and the regression results more comparable, which guided our choice. In both classification and regression, we performed oversampling only on the training set, with a sampling rate of 0.1.

Finally, the models were evaluated according to the following metrics: *accuracy*; *recall*, *precision* and *F1* score; and *ROC AUC* value. The ROC curve (receiver operating characteristic curve) plots the true-positive rate and false-positive rate to determine whether the model can distinguish among the classes. Higher values of the AUC (area under the curve) describe higher degrees of separability.

Visualization and interpretation of the models

Interpreting the behavior of a prediction model can help understand the motivations behind its decisions, and thus detect eventual biases [159]. To achieve interpretability, we can either use interpretable models, such as decision trees, decision rules, and linear regression) or define an additional layer for the model explanation [159]. This explanation is meant to be human-readable and can be obtained either by using surrogate models or visual representations. Examples of a surrogate are *global surrogate models*, which threats the non-interpretable model as a black box and approximates their behavior. However, the conclusions drawn from those surrogates refer to the model but not to the data. Differently, local surrogate models explain each prediction rather than the model as a whole.

Random Forest is a non-interpretable model, hence the need for an additional layer to understand its behavior. We decided to use partial dependence plots (PDP or PD plot) [159], which are model-agnostic. PDP plots show the marginal effect of one or two features on the predicted variable [71]. Their relationship is visually shown. Specifically, the plot displays how the average prediction in the dataset changes with the value of the feature analyzed [159].

Results

In the following section, we report the outcomes of both the regression and classification analyses. To reduce the bias introduced by the record selection, in particular random sampling, we conducted repeated experiments (100). Hereby, we present the outcomes and the value of the evaluation metrics as an average of the results obtained.

First, we replicated our previous study on ANN over 1-week timeframes and 1-day timeframes, whose performance was used as a baseline. The results are shown in

Table 4.3: Churn time (regression) and churn (classification) prediction using the ANN defined in [143] with the original settings, but one one timeframe of history, and with shorter timeframes.

	Regression				Classification				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Weekly tf	0.08	0.27	0.18	0.37	0.79	0.79	0.78	0.78	0.72
Daily tf	0.08	0.3	0.19	0.28	0.78	0.78	0.76	0.77	0.62

Table 4.4: Churn time (regression) and churn (classification) prediction using Random Forest using the same setting of the ANN.

	Regression				Classification				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Weekly tf	0.08	0.28	0.2	0.43	0.77	0.77	0.78	0.78	0.77
Daily tf	0.07	0.27	0.19	0.44	0.76	0.76	0.76	0.77	0.77

Table 4.3 and Table 4.4. Although the classification performs better in the ANN with weekly timeframes, the RF models hold better results for what concerns the regression analysis. RF is also better performing than ANN when the length of the timeframes is reduced to one day.

Instance Selection. After having compared ANN and RF, we analyzed the different approaches for instance selection, previously described (Random, Gaussian, and Uniform selection). Figure 4.14 shows the distribution of the churn-time values when using the three instance selection methods in RF prediction analysis on the Play&Go data.

Both in the regression and classification analysis, the models trained using the data sampled using the Gaussian Selection method achieved the best performance(Table 4.5). Yet, in the validation analysis, the models trained on the data sampled through Uniform Selection performed better than the other sampling methods (Table 4.6). Hence, in the following analyses, we opted for the Uniform sampling approach to produce more generalizable models.

Raw behaviors or impact values. Once we decided on the sampling method, we analyzed different ways to process data. Specifically, we compared the usage of raw participation behaviors and the impact values, derived from the impact function previously define. Our hypothesis was that the inclusion of additional information on performance improvement or worsening could enhance the prediction power of the model. Table 4.7 presents the results for both the classification and the regression model. The accuracy of the outcomes trained on the Play&Go data slightly decreases when using the impact values (Table 4.7), rather than the raw behavioral data (Table 4.5). However, in the validation analysis on the TagPro dataset, the

Table 4.5: Results of the churn prediction using the Play&Go data.

	Regression				Classification				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Random Selection	0.07	0.27	0.19	0.44	0.76	0.76	0.77	0.77	0.77
Gaussian Selection	0.05	0.22	0.14	0.64	0.85	0.85	0.85	0.85	0.83
Uniform Selection	0.1	0.32	0.26	0.4	0.74	0.74	0.74	0.74	0.74

Table 4.6: Results of the churn prediction using the TagPro data as validation set.

	Regression				Classification				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Random Selection	0.12	0.38	0.29	0.13	0.73	0.73	0.74	0.73	0.7
Gaussian Selection	0.13	0.37	0.25	0.16	0.73	0.73	0.73	0.73	0.72
Uniform Selection	<u>0.11</u>	<u>0.34</u>	<u>0.27</u>	<u>0.29</u>	<u>0.73</u>	<u>0.73</u>	<u>0.73</u>	<u>0.73</u>	<u>0.71</u>

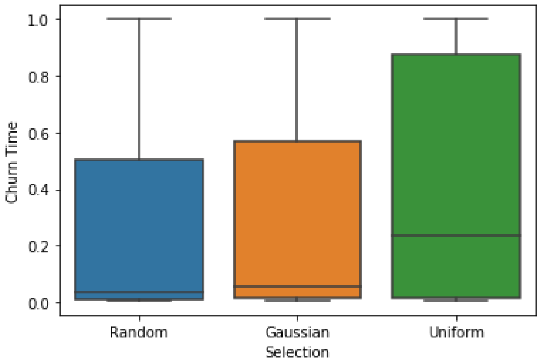


Figure 4.14: Distribution of the churn-time values using the different instance selection methods.

Table 4.7: Results of the churn prediction using the Play&Go impact values and TagPro as validation set.

	Regression				Classification				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Play&Go	0.13	0.36	0.31	0.24	0.69	0.69	0.69	0.69	0.68
TagPro	0.17	0.35	0.3	0.19	0.71	0.71	0.71	0.71	0.7

TABLE 4.8: Results of the churn prediction training the models on TagPro data

	Regression				**Classification**				
	MSE	*RMSE*	*MAE*	*R2*	*ACC*	*PREC*	*REC*	*F-SCORE*	*ROC-AUC*
Raw Behaviors	0.04	0.2	0.12	0.19	0.75	0.73	0.73	0.73	0.73
Impact values	0.08	0.27	0.2	0.31	0.79	0.79	0.79	0.79	0.75

TABLE 4.9: Summarizing table, comparing the performance of the regression and classification models trained on Play&Go (first two columns) and TagPro (last column). The models were all built using RF, and the Uniform Selection sampling method.

		Test set (Play&Go)		**Validation set (TagPro)**		**Ad-hoc trained on TagPro**	
		Raw	*Impact*	*Raw*	*Impact*	*Raw*	*Impact*
Regression	*MSE*	0.1	0.13	0.11	0.17	0.04	0.08
	RMSE	0.32	0.36	0.34	0.35	0.02	0.27
	MAE	0.26	0.31	0.27	0.3	0.12	0.2
	R2	0.4	0.24	0.29	0.19	0.19	0.31
Classification	*ACC*	0.74	0.69	0.73	0.71	0.75	0.79
	PREC	0.74	0.69	0.73	0.71	0.73	0.79
	REC	0.74	0.69	0.73	0.71	0.73	0.79
	F-SCORE	0.74	0.69	0.73	0.71	0.73	0.79
	AUC	0.74	0.68	0.71	0.70	0.73	0.75

model trained on the impact values behaved better (Table 4.7) than in the raw behavior model (Table 4.5). Therefore, impact values appear to better accommodate generalizability.

Training on TagPro. We also analyzed the models' behaviors when directly trained on the TagPro data, using the Uniform Selection method for instance sampling. Table 4.8 show the performance of the models trained on both raw behaviors and impact values.

In contrast to our expectations, the regression model from the raw behavioral data trained ad-hoc on TagPro data was less accurate than the model trained using Play&Go data and TagPro as validation set (Table 4.9). On the other hand, the classification model obtained similar results in both the validation set and in the ad-hoc TagPro model. Models trained on the impact values, on the other hand, behaved better in the ad-hoc TagPro model than in the validation analysis. The improvement is more visible in the regression than in the classification analysis.

Visualization and interpretation of the models. In the last phase of our prediction study, we visualized the behaviors of the models in partial dependence plots (PDPs). We built a plot for each feature used—i.e., frequency of usage, number of points obtained, gameplay length, and number of game actions performed—-and discerning among raw behaviors and impact values. Figure 4.15, Figure 4.16, Figure 4.17, and Figure 4.18 show the PDPs for Play&Go and TagPro models.

Each plot highlights how *frequency of usage* (FoU) and *gameplay length* strongly affect churn predictions. In other terms, a longer permanence in the game increases

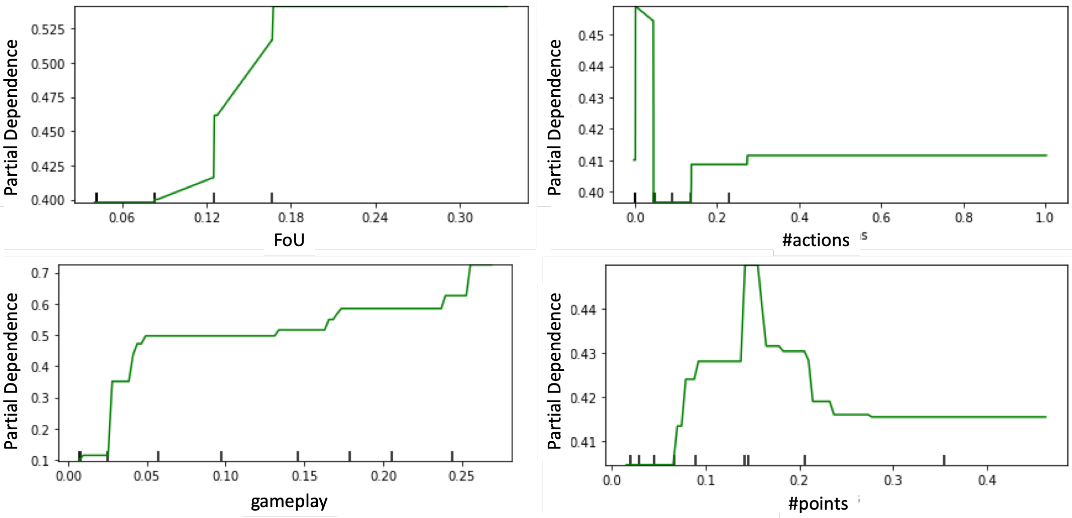


FIGURE 4.15: Visualization of the models built on Play&Go raw behavioral data.

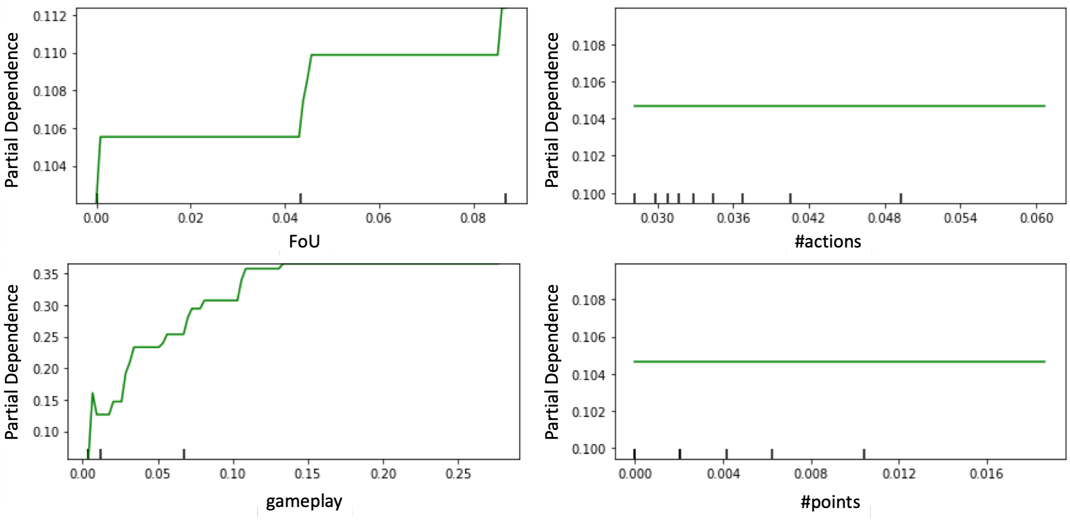


FIGURE 4.16: Visualization of the models built on TagPro raw behavioral data.

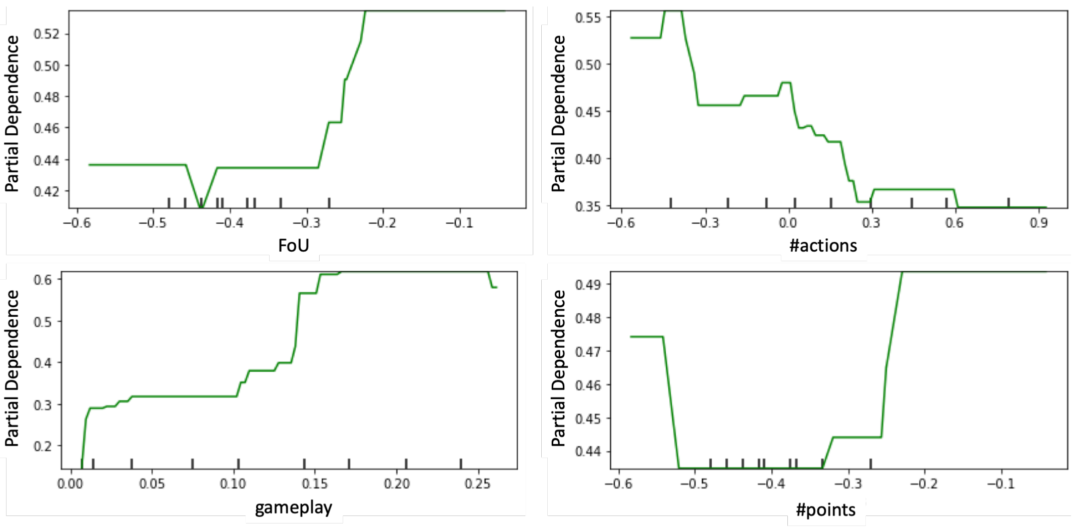


FIGURE 4.17: Visualization of the models built on Play&Go impact behaviors.

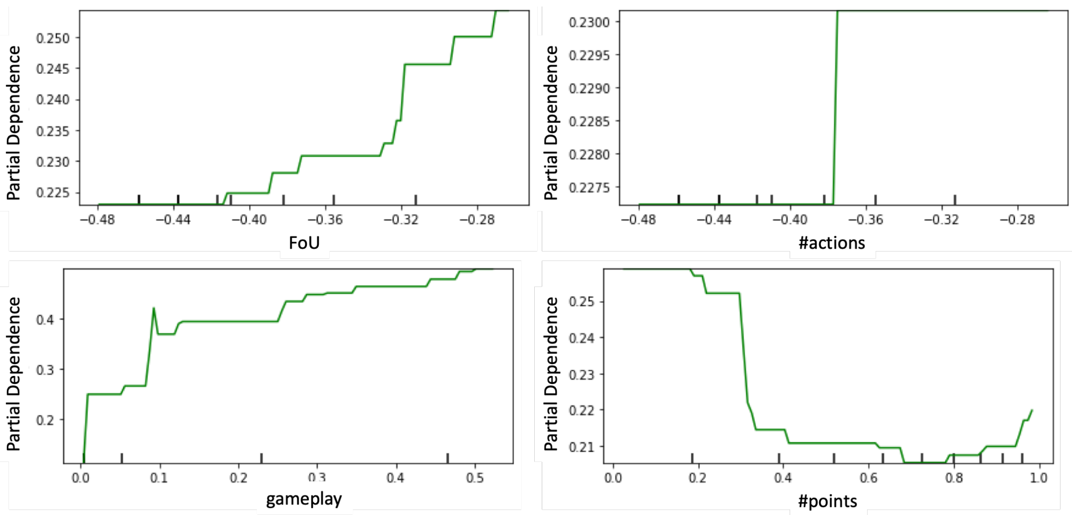


FIGURE 4.18: Visualization of the models built on TagPro impact behaviors.

the likelihood of remaining in the game for longer and being constant in the game activity.

The number of points obtained and the number of game actions performed had a different effect on each model, with the nature of the data (raw or treated with the impact function) leading to different interpretations.

For what concerns raw behaviors, the number of actions performed has a shallow impact on the prediction in both Play&Go (Figure 4.15) and TagPro (Figure 4.16) models. Conversely to the number of game actions, the number of points is indeed meaningful in the churn prediction. For example, in the Play&Go model trained on raw behaviors (Figure 4.15), the quantity of awarded points is related to players' retention until a specific threshold. Once the threshold is reached, the higher the number of points, the earlier the player will abandon the game. Thus, high activity peaks are important warnings.

The models trained on impact values, on the other hand, were very sensitive to the number of game actions performed. Higher impact values for the number of actions performed were an indicator of longer retention. Therefore, progressive increase (improvement) in the actions performed signaled players' long-term commitment. In contrast to the positive signal conveyed by the increasing number of actions, in both TagPro and Play&Go, a high impact in the number of points is indicative of players' abandonment.

4.2 Assessing Players' Preferences

In the previous section, we analyzed telemetry data to obtain information on players' population at a high level. Those coarse-grained analyses allowed, for instance, to identify whether the winning strategies actuated were in line with the gamification goal as an evaluation metric on the gameful design. The findings also highlighted the existence of a consistent subset of players lacking in (long-term) engagement. Despite using only participation data, we succeeded in predicting players' churn. However, identifying players' at risk of abandonment is not itself a solution to the problem. The likelihood of churn can be reduced by providing users stronger motivations to play, in line with their motivational affordances. In the following, we exploit players' interaction patterns and behaviors to learn their preferred configuration for a specific game element. Then, we broaden our scope to player profiling as a method to deploy ad-hoc adaptation strategies. Specifically, we investigate how the strategies differ when theory and data-driven approaches are used.

4.2.1 Learning Players' Preferences

In our first study, we researched the possibility of using players' in-game behaviors to learn their preferred variant of a specific game element. In practice, we analyzed Play&Go challenge types. We defined a Reinforcement-Learning-like algorithm to model their effect on players' in-game participation and evaluate whether preferences can be learned from the analysis of in-game activity level. The algorithm's output is a sorting of the challenge types, in decreasing preference order for each user. The algorithm's performance and accuracy are computed against ground truth, derived from players' implicit preferences. Those preferences are indirectly declared when players select one option rather than another when presented a choice.

The learned preferences are obtained from an RL algorithm. Hence the temporal information on players' behaviors is retained. Similar to previous works, we divided the gameplay into 1-week timeframes. Nevertheless, the algorithm can also be adapted to different (and even variable) gameplay structures.

Constructs

In the following, we define customization as intended in our analysis. At the lowest level of implementation, we find game elements [80], which can be seen as the game's atoms. Examples of game elements are quests, challenges, or trivial game actions. We are interested in customizable game elements, for which one or more properties can be tuned. The properties can assume a value in a finite and reasonably small set a priori defined($PARAM$). In Play&Go, the customizable game elements we analyzed are single-player challenges, and the modifiable property is the counter the challenge is evaluated upon—i.e., walk kilometers, bike kilometers, bus trips, train trips, and green leaves points. We will refer to this property as counter, property, and challenge type.

The impact of the specific challenge type on players was computed by measuring their in-game participation. Players' participation was described by modeling the intensity and the constancy of players' in-game actions [140]. The intensity of usage is calculated from the number of points and game actions performed, while the constancy of usage is translated into the frequency of usage metric, i.e., the percentage of the player's active days. Both values are normalized in the $[0, 1]$ interval and averaged to obtain a single value describing players' activity within the timeframe.

Players eligible for our study had a final level greater than two and played for at least five weeks (timeframes). This filtering was needed to retain enough information to build the ground truth, as we will later describe. As a result, our population comprised 94 players.

Implicit Preferences

As a reminder of Play&Go mechanics (Chapter 3), players can engage in weekly single-player challenges. Challenges can vary in a *parameter*, specifying the counter upon which the challenge is evaluated, among walk kilometers, bike kilometers, green leaves points, train trips, and bus trips. After players reached Level 2 in the game, each week, they are offered a selection of two or three different challenges. They can choose one option, which will be the challenge active in the following week. If a choice is not made, a default challenge is automatically assigned. The ground truth was built from players' in-game choices, using a paired comparison protocol [178]. For each player, we modeled the choices a squared matrix of dimension n, with n the number of parameters available. In the matrix, the value x of the cell c_{ij} describes how many times the counter i was chosen over the counter j. Even though the challenge types were 5, some players may have been presented with fewer counters in the selection phase. This imbalanced design derives from the functioning of the Recommender System (RS) for Play&Go single-player challenges. The RS suggests the player challenges on the counter they have already used in the past. For example, if they never tracked a bus trip, it is highly unlikely that a challenge with the bus counter will be proposed. As a consequence, users can have a different experimental design rather than the complete set of comparisons. Yet, we can still reach competitive performance, despite the imbalanced designed [28].

The preference score for each counter is computed as the number of votes received for that counter, normalized by dividing for the number of comparisons [178]. As a result, every player has a vector w of dimension n, with the preference scores for every counter. The reliability is then evaluated using the consistency factor. An inconsistency (or cyclic triad) occurs, for instance, when x is preferred to y, y is preferred to z, but z is preferred to x. The consistency coefficient [111] is computed with the formula: $\zeta = 1 - \frac{24*c}{z^3-z}$ where z is the number of choices made and c is the number of cyclic triads. Perfect consistency is described by $\zeta = 1$, which is inversely proportional to the number of circular triads. To account for human error, generally, a value of 0.75 is considered good [129].

We analyzed data from 115 players. Only 1 had a $\zeta = 0.35$, while the others had $\zeta \geq 0.79$. When we included information on challenges' difficulty, the values of ζ was higher than 0.92

To further evaluate the ground truth, we performed a prediction analysis on the last player choices, using the preference scores in the $N-1$ weeks in which they made a choice. The selection for the N^{th} week predicted and compared with the true choice. For 70% of the 115 players, the forecast was correct. The percentage increased to 71% when players performing less than ten choices were excluded (21 players). Finally, the prediction accuracy further increases to 82% when also the difficulty is included in the parameter selection. This improvement is counterintuitive as embedding information also on challenges' complexity increases the number of counters available, despite using the same amount of data. Consequently, we had less information on players.

Learning Preferences with RL

In Play&Go, challenges are the main driver of engagement [151]. Therefore, we hypothesized a relationship between in-game activity and the type of challenge assigned to the players. In other words, our intuition was that a higher in-game activity was an indicator of the player's appreciation for the challenge type. We saw a similarity with the Reinforcement Learning (RL) paradigm, in which the agent learns from experience from the reactions (feedback or reward) obtained after their action. In this trial-and-error cycle, the agent's goal is to maximize a long-term reward [208]. Analogously, we thought of learning players' preferences from their in-game behaviors in reaction to different challenge types over the gameplay. The algorithm follows an

Algorithm 2 The Offline Gradient Bandit Algorithm

```
procedure OfflineGB(B, R, k, n, α)
    H(a) ← 0, ∀1 ≤ a ≤ k ∈ ℕ
    H(a) ← 1/k, ∀1 ≤ a ≤ k ∈ ℕ
    avg_reward ← 0
    i ← 1
    while i ≤ n do
        A_t = B_i
        r = R_i
        avg_reward = avg_reward + (1/i)(r − avg_reward)
        H(A_t) = H(A_t) + α(r − avg_reward)(1 − Pr(A_t))
        for all a ≠ A_t, 1 ≤ a ≤ k ∈ ℕ do
            H(a) = H(a) + α(r − avg_reward)Pr(a)
        end for
        for all 1 ≤ a ≤ k ∈ ℕ do
            Pr(a) = exp(H_t(a))/sum(exp(H))
        end for
    end while
    return H
end procedure
```

RL-paradigm, in which information obtained until time t is used to make a decision on the action to perform at time $t+1$. Conversely to other ML paradigms, we are unaware of the correct action to take. Rather, we can exploit the data obtained to **evaluate** which our best option is. The algorithm was designed to produce players' preference scores for each challenge type, and thus a sorting from the preferred to the least appreciated type. Towards this, the gameplay was divided into weekly timeframes, where a week is the lifetime of a challenge.

In the following, we review the RL paradigm and the class of problems that can be resolved through it. In its simplest form, RL algorithms can tackle problems known as the *bandit problems*. In those problems, we have one or more "bandit", usually visually displayed as a slot machine. Every machine has its—unknown—distribution, from which it picks its reward. Pulling an arm results in a different outcome, depending on the chosen machine. The final goal is to maximize the reward in the long run by assessing the best action at each step. Therefore, the faster the optimal selection strategy is found, the higher the final gain. As the agent interacts with the bandit, they learn the Q-value, the expected reward for each action. The Q-values are updated at every interaction as the agent's experience increases.

In our problem, the goal is to estimate a preference score for every action (bandit) instead of finding a single optimal choice. Therefore, we used a variant of the multi-armed bandit problem: *the gradient bandit.* Gradient algorithms compute the value of an action in relation to the other possible actions. The outcome is a preference score $H_t(a)$, which guides the action selection. Initially, $H_0(a) = 0, \forall a$, with a uniform probability of choice for every action. As the iteration progress, the $H_t(a)$ value is updated, for each action, using a stochastic gradient ascent (SGD). When an action is taken, the new value for $H_t(A_t)$ is computed by weighting the reward returned by the probability of the action and the learning rate coefficient.

$$H_{t+1}(A_t) = H_t(A_t) + \alpha(R_t - \overline{R}_t)(1 - \pi_t(A_t)), and$$
$$H_{t+1}(a) = H_t(a) + \alpha(R_t - \overline{R}_t)\pi_t(a), \forall a \neq A_t$$

$\alpha > 0$ is the learning step-size parameter, and $\overline{R}_t \in \mathbb{R}$ is the baseline averaging the rewards obtained until time t. If the actual reward is lower than the baseline, in the following iteration, the probability of choosing that action decreases; it increases otherwise. Higher preferences result in a higher chance to select that action. The probability of taking the action a at time t follows the soft-max distribution.

$$\pi_t(a) = Pr\{A_t = a\} = \frac{e^{H_t(a)}}{\sum_{b=1}^{n} e^{H_t(b)}} \tag{4.1}$$

In those problems, the actions are not associated with the state of the environment—i.e., different situations. The learner research the single optimal strategy (if stationary) or the optimal strategy as it changes over time (otherwise). This condition well connects to our original problem. First, we use only data on players' activity without embedding data on the game's state. The reward is based on the player's level of activity in the week that the challenge was active. Second, the changeable nature of players' preferences is accounted for, thanks to the iterative nature of the RL paradigm.

In this study, we operated with offline gameplay data: we knew which challenge was active (the bandit) instead of having to choose it. RL-algorithm, on the other hand, resolves online problems. Hence, our offline algorithm (Algorithm 2) is *inspired by* the RL paradigm, aimed at assessing the *feasibility* of using in-game behaviors to learn players' preferences. The algorithm silently observes players' activity and learns the preference scores for each challenge type without actively taking part in the decision process. The algorithm computes the preference scores knowing, at each iteration, the current bandit (Line 7 in Algorithm 2) and the value of the activity behavior (Line 8 in Algorithm 2). Such values are used to adjust the preferences for each bandit (Lines 10-13 in Algorithm 2). In a future step, the algorithm can easily be translated into a real RL algorithm by changing Line 7 (Algorithm 2) to select the option depending on the preferences scores computed (Equation 4.1).

For each player, Algorithm 2 takes as input the challenges list active during the player's n weeks of gameplay (B), the values of the activity behaviors for those weeks (R), the number of types of challenges they have engaged with (k), and the learning coefficient (α). The output is a vector of dimension k with the preference scores for each challenge type k_i.

Results

In the following, we present the results of our study. First, we investigated the relationship between the challenge type and players' in-game activity and built the ground truth using players' in-game choices. Then, we computed the preferences scores using our Algorithm 2 and evaluated its outcomes against the ground truth. Finally, we characterized players for whom the algorithm performed particularly well.

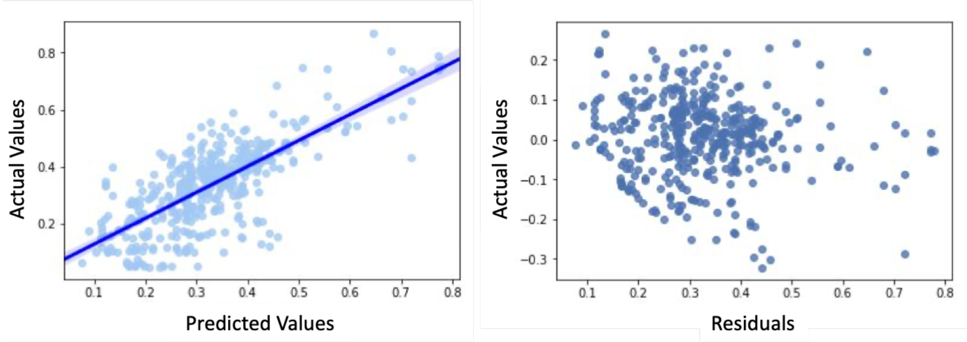


FIGURE 4.19: Linear Regression The test-set values of the test plotted against the predicted values in (a) and the residuals plotted against the predicted values in (b).

Preliminary Analysis. Before using players' in-game activity as feedback for learning their tastes, we analyzed whether their activity was indeed a predictor of their preferences.

We conducted a regression analysis in which the features, or independent variables, were the player identity (the username, as a categorical variable) and the challenge type, and the dependent variable was players' in-game activity level. We built a Multiple Linear Regression model to verify whether how predictable the level of activity in the game was given the player identity and the type of challenge they had assigned. The dependent variables—i.e., the behaviors' values— represent the repercussion the challenge had on the player's participation level.

In the model, the features were 159: 155 dummy variables representing players' nicknames and four dummy variables representing the challenges types. Our dataset was built of about 2k entries and 156 players who played between 5 and 23 challenges. The outcomes (Figure 4.19) were positive in showing that the activation of a specific challenge type for a specific player could be predictor of their in-game activity [(F(155,1943) = 161, p < 0), with an R^2 of 0.93]. Hence, we proceeded with our experimentation.

Ground Truth. In the challenge customization phase, Level 2 Play&Go players can choose among two options, while higher levels can make their selection among three options. Accepting a challenge implies refusing the other options presented. Therefore, each selection produces one or two comparisons: the chose option over the other(s) challenge(s) proposed. The ground truth was built using this information. Specifically, we used the number of accepted and refused challenges for each player and challenge type [138]. Those selections players made during the gameplay were used to infer their preferences. From our population, as we anticipated, we retained players playing for at least five weeks and that reached level 2. Those 94 users were the players involved in at least ten in-game comparisons. The conversion from in-game selection to preference score was performed through the paired comparison protocol. As typically done in paired comparison studies [178], we build a squared matrix of dimension k for every player, with k the number of properties available. In such a matrix, a value of x for the cell c_{ij} means that the counter i was chosen x times over the counter j. The preference score was computed by counting the votes received by each counter, normalized by dividing by the number of comparisons per counter [178]. This process produced, for each player, a vector w_c of dimension k, with their preference scores.

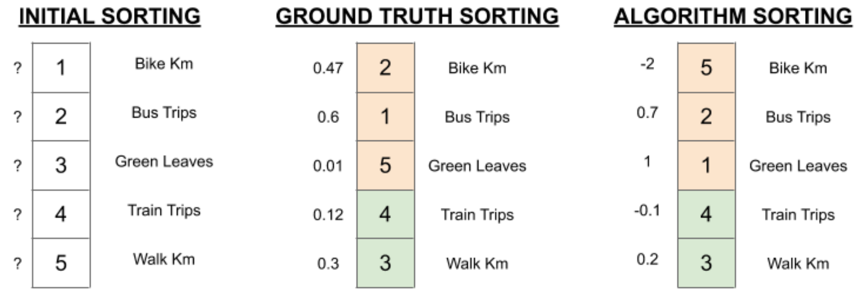


FIGURE 4.20: Example of properties sorting for a player, resulting from their in-game choices (ground truth) and the algorithm's outcomes.

Evaluation. Having assessed a dependence (and predictability) among players' activity and the challenge they have activated, we run Algorithm 2 to learn players' preference scores. We obtained 94 vectors w_p of dimension k_i, one per player. The elements of the vector correspond to the counters available for the specific user.

The two vectors w_c and w_p, of the ground truth and player preferences, are not directly comparable. First, they are obtained through different processes, and then they also belong to different value ranges: w_c is in the range $[0;1]$ and w_p in $[-\infty, +\infty]$. Nevertheless, our goal was to retrieve a preference ordering of the challenge types, from the most to the least liked. Hence, we compared the indexing of both vectors w_c and w_s, which were initialized to the same mapping (e.g., 1 = Bike Km and 2 = Bus Trips, as shown in Figure 4.20). The array of the challenge types is independently set and is not modified. Then, the updated indexing for both the ground truth and the algorithm is produced (Figure 4.20), using players' in-game choices and in-game behaviors, respectively. The output sortings s_c and s_p describe where the counters are positioned in the player's personal rankings.

Figure 4.20 shows the following example. Given $w_c = [0.47, 0.6, 0.01, 0.12, 0.3]$ and $w_p = [-2, 0.7, 1, -0.1, 0.2]$, being the vector of the scores in the ground truth and the scores computed by the algorithm respectively, their decreasing sortings are the following: $s_c = [2, 1, 5, 4, 3]$ and $s_p = [5, 2, 1, 4, 3]$. Thus, the actual ranking is:

- $BikeKm$ in position 2 for the ground-truth with value 0.47 and position 5 for the algorithm with value -2;
- $BusTrips$ in position 1 for the ground-truth with value 0.6 and position 2 for the algorithm with value 0.7;
- $GreenLeaves$ in position 5 for the ground-truth with value 0.01 and in position 1 for the algorithm with value 1;
- $TrainTrips$ in position 4 for the ground-truth with value 0.12 and position 4 for the algorithm with value -0.1;
- $WalkKm$ in position 3 for the ground-truth with value 0.3 and position 3 for the algorithm with value 0.2.

We compared the sortings, and thus the algorithm's performance, using the following measures. The selected error metric was $NRMSE$ (Equation4.2), representing the

TABLE 4.10: Evaluation of the Algorithm. Distribution of the error and similarity measures over the 94 eligible players.

	min	*mean*	*median*	*max*
$NRMSE$	0	0.27	0.24	0.8
R^2	-3	0.014	0.6	1
Similarity	0.2	0.83	0.93	1

percentage of the error obtained, easier to interpret respecting to $RMSE$. $NRMSE$ penalizes divergent sortings, in which the actual position - ground truth - of a counter in the ranking is far from the position computed from the learned preference scores. Following our example (Figure 4.20), $NRMSE = 0.57$. Such a high error value well describes the situation. The counters *TrainTrips* and *WalkKm* have been successfully positioned in the ranking (in both orderings, they have positions 4 and 3, respectively). However, *BikeKm* and *BusTrips* and *GreenLeaves* have been ranked poorly. *BikeKm* should be in position 2, but in the algorithm's sorting s_s is in position 5; *BusTrips* in the algorithm's outcome is in position 2, but should be in position 1, and *GreenLeaves* should be in the last (position 5), but in s_s assumes the first position.

$$NRMSE = \frac{\sqrt{\frac{1}{k}\sum_i^k (x_i - \overline{x_i})^2)}}{x_max - x_min} \tag{4.2}$$

Then, we used the cosine similarity metric to assess the similarity between the two (non-zero) arrays. The cosine similarity measures the cosine of the angle between two vectors projected in a multi-dimensional space. Smaller angles translate into higher similarity values. We computed the R^2, as an additional similarity metric. R^2 describes the proportion of variation in the dependent variable that can be attributed to the independent variable.

Players Characterization. The performance of the algorithm were very variable (Table4.10). The predictions were very accurate (on average $NRMSE = 0.13$, $R^2 = 0.81$, and $sim = 0.97$) for 51% of the players the prediction, but its accuracy was low for the rest.

To research the causes of this variability, we analyzed some players' properties that might explain the divergent performance. We divided our population of eligible players into two groups: players for which the algorithm achieved good and bad performance. The groups were formed considering the $NRMSE$ and R^2 values, but not the cosine similarity, which was overly optimistic (see Table4.10). Being one an error metric and the other a similarity metric, we set two thresholds. In the *good* group, players had $NRMSE \leq 0.2$ (an error of 20% is admitted), and $R^2 = 0.5$. Since data were not normally distributed, according to the D'Agostino-Pearson K^2 test, the two groups were compared using the Mann-Whitney W test.

H1. *The number of options (bandits) influenced the accuracy of the learning.*
Players may have interacted with a smaller set of challenge types for how the challenge system is designed. Our hypothesis was that the fewer the challenge types for which the preferences had to be learned, the more accurate the algorithm. Hence, we compared how the number bandits (smaller k) differed among the good and bad performance groups. Since both groups (players for which the algorithm had good

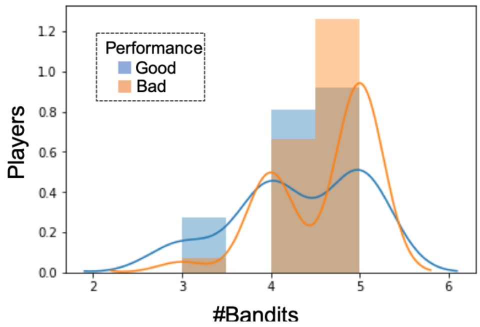


FIGURE 4.21: Distribution of the number of bandits over the two group of players.

performance and players for which the algorithm had bad performance) were not normally distributed (for the 1^{st} group $K^2 = 3.9$, p-value $> .05$; and for the 2^{nd} group $K^2 = 8.9$, p-value $< .05$), we computed the Mann-Whitney W test. We found a statistically significant difference (U = 840.5, p-value $< .05$) between players in the first group (M = 4.32, SD = 0.70) and players in the second group (M = 4.60, SD = 0.56). Despite the number of bandits affecting the algorithm's accuracy, the number of challenge types per player in the groups is still well-distributed (Figure 4.21).

H2. *The number of challenges - i.e., iterations - influenced the quality of the outcomes.*
The algorithm, although offline, follows the RL paradigm. Hence, the learning is adjusted as more iterations are performed. In our case, each iteration corresponds to a game week—i.e., a new challenge and relative in-game behavior as feedback. Hence, the number of iterations per player depends on their gameplay length. We hypothesized that the number of challenges impacted the accuracy of the learned preferences. Since both of our group were not normally distributed (for the 1^{st} group $K^2 = 2.7$, p-value $> .05$; and for the 2^{nd} group $K^2 = 3.2$, p-value $> .05$), we computed the Mann-Whitney W test. We found a non-significant difference (U = 975, p-value = .27) between players in the good performing group (M = 14.78, SD = 4.40) and players in the bad performing group (M = 15.26, SD = 04.79). Therefore, in our dataset, the number of iterations did not impact the algorithm's performance.

H3. *Players that manifested a much stronger preference were easier to model.*
We hypothesized that the algorithm was inaccurate when players' preferences were already shallow: players explored among different challenge options. Thus, it was impossible for the algorithm to learn preference scores, when they were negligible. To test the hypothesis, we measured the strength of players' preferences using the relative standard deviation (RSD) of the ground truth preference scores. We computed the relative version of the SD for comparability reasons. Since both for our group were not normally distributed (for the 1^{st} group $K^2 = 4.1$, p-value $> .05$; and for the 2^{nd} group $K^2 = 0.9$, p-value $> .05$), we computed the Mann-Whitney W test. Players in the good performance group (M = 104, SD = 34.49) showed a higher relative variance in the preference scores (U = 1339, p-value $< .01$) than players in the bad performance group (M = 88.5, SD = 22.6). Therefore, we concluded that, for the algorithm, it was easier to learn preferences for players with higher RSD—i.e., stronger preferences.

H4. *Player personality represented by their player types scores affected the modeling power of the algorithm.*
During the registration, users needed to fill survey, containing the items of the Hexad

Table 4.11: Results of the test of hypothesis (H4.), investigating whether player types characterize the two groups of players (good performance of the algorithm vs bad performance of the algorithm).

User Type	*U*	*p-value*	$mean_1$	std_1	$mean_2$	std_2
Free Spirit	889	0.4	23.12	3.15	23.3	2.87
Achiever	989	0.7	22.56	3.75	22	3.63
Player	865	0.33	17.82	5.98	18.6	5.47
Distruptor	931	0.56	14.74	5	14.26	4.7
Philanthropist	1060.5	0.89	24	2.9	23	3.67
Socializer	816.5	0.19	21.36	4.39	21.92	5.11

User Types [219]. Using the results of the survey, we modeled players' personality and researched whether for some types preferences were easier to learn. Thus, we analyzed the scores of the 6 player types (Socializer, Player, Free Spirit, Disruptor, Philanthropist and Achiever) to explain the performance variability. We computed the Mann-Whitney W test among the good and bad group of players for each player type in the Hexad model:

- **Free Spirit.** For the 1^{st} group $K^2 = 0.54$, p-value $> .05$; and for the 2^{nd} group $K^2 = 1.38$, p-value $> .05$.
- **Player.** For the 1^{st} group $K^2 = 1.7$, p-value $> .05$; and for the 2^{nd} group $K^2 = 1.6$, p-value $> .05$.
- **Achiever.** For the 1^{st} group $K^2 = 1.5$, p-value $> .05$; and for the 2^{nd} group $K^2 = 7.5$, p-value $< .05$.
- **Distruptor.** For the 1^{st} group $K^2 = 3.4$, p-value $> .05$; and for the 2^{nd} group $K^2 = 1.8$, p-value $> .05$.
- **Philanthropist.** For the 1^{st} group $K^2 = 3$, p-value $> .05$; and for the 2^{nd} group $K^2 = 6.2$, p-value $< .05$.
- **Socializer.** For the 1^{st} group $K^2 = 3.7$, p-value $> .05$; and for the 2^{nd} group $K^2 = 14.8$, p-value $< .05$.

We concluded that the Hexad User Type scores are unrelated to the performance of the algorithm (Table 4.11).

4.2.2 Modeling Players' Tastes

Players' telemetry data is a powerful source of information. In this chapter, we showed how players' in-game behaviors provide meaningful insights on the progress of the gameplay, design choices, and game strategies [145]. Behavioral information can also be used to predict players' premature churn [143, 139] and learn their preferences for what concerns the personalization of a specific game element [138, 141]. In this final study, we broaden our research on player modeling without limiting it to the customization of a single element. Instead, we study how to produce ad-hoc adaptation strategies and how self-report and telemetry data impact the creation of player models.

Table 4.12: Overview of the in-game behaviors computed.

ID	Name	Description
R	Reactivity	Describes players' velocity in answering to game events.
S	Sociality	Describes players' tendency to be initiators of multiplayer challenges.
IU	Intensive Usage	Describes players' velocity in levelling up.
W	Winning Social	Describes players winning rate in multiplayer challenges in contrast to single-player challenges.
F	Full-usage	Describes the percentage of features used.
Cm	Competitive	Describes players' preference of competitive over cooperative multiplayer challenges.
A	Active	Describe players' tendency to choose rather than having automatically assigned single-player and multi-player challenges.
Cs	Constancy	Describes the percentage of active days throughout the gameplay.
Po	Purpose-oriented	Describes the ratio of green (walk/bike) trips and kms over the trips and kms tracked.
St	Striving	Describes the tendency to engage in difficult challenges.
Si	Self-improvement	Describes the tendency to increase and improve the personal performance over the gameplay.

Constructs

In the following section, we detail how we modeled and computed players' idealized and contextualized preferences. Then we describe the approach used to compare players' models.

Idealized Preferences: User Types Theory-driven approaches often rely on surveys and self-assessment to obtain players' profiles. Hence, they produce idealized preferences, as we named them. Idealized preferences are the results of whom players think they are and what they usually enjoy, which can vary between real-life and games and across different games. In this study, we used the Hexad User Types model [153] and its validated survey [219] to obtain our participants' idealized preferences. This model, based on self-determination theory [190], describes users' motivations and interaction styles within gamified environments. The Hexad taxonomy proposes six user types, characterized by their intrinsic and extrinsic motivations: Philanthropists, Socialisers, Free Spirits, Achievers, Players, and Disruptors.

(A) **Achievers** seek to challenge themselves in difficult task and to master skills.

(Ph) **Philanthropists** seek to have purpose by helping others, regardless of rewards.

(R) **Players** seek to collect in-game goods, rewards or incentives and are willing to engage in any activity to obtain those.

(S) **Socialisers** seek social connections and interactions with other players.

(D) **Disruptors** seek being rebellious and testing the system's boundaries to achieve change.

(Fs) **Free Spirits** seek being free and independent. They enjoy exploring and experimenting with the system to satisfy their creativity.

Contextualized Preferences: Player behaviors Data-driven approaches, as the name suggests, exploit telemetry data to extract in-game behaviors, and thus, to model players. Gameplay data is generally collected in the form of datalogs, in which game events are stored. Hence, this data needs to be processed before proceeding to further analysis.

For Play&Go, we obtained data of players' game actions, such as trips tracked, levels obtained, interactions with the invitation system, and information on single-player and multiplayer challenges. However, in-app interactions (e.g., visit a specific app page) were not recorded.

In collaboration with Play&Go designers, we defined a list of relevant behaviors for the environment examined. Those behaviors (summarized in Table 4.12) describe players' activity from different perspectives and different granularity levels. The behaviors model presents: in-game activity (*Intensive Usage*, *Constancy* and *Full-usage*), impact and existence of social interactions (*Sociality*, *Competitive*, and *Winning Social*), tendency to customize (*Reactivity* and *Active*), will to challenge one-self (*Striving* and *Self-improvement*), and predisposition to act in line with the gamification ulterior motive (*Purpose-oriented*).

(R) **Reactivity** measures players' velocity in answering to game events (e.g., choice of single-player challenges and reply to invitations). Gameplay is divided into weeks, and each week players have four days (Tuesday to Friday) to select single-player or multiplayer challenges. Reactivity for a week is measured as the percentage of the four days spent before making a choice, if any. The Reactivity value is the average of all the weekly Reactivity for active weeks. A week (or a day) is active if the player performed at least one game action within that week (or day).

(S) **Sociality** measures players' willingness to play with others, either in competitive or cooperative multiplayer challenges. This is manifested through in-game invites to multiplayer challenges. Hence, for these behaviors, we measured the number of invites sent by the players. Measuring the number of invites is different from measuring the number of multiplayer challenges, as players may want to play with somebody, hence sending the invites, but may have no invites accepted, and therefore cannot complete the challenges.

(IU) **Intensive Usage** measures how important players' in-game activity is. The more game actions they perform, the more points they get, and thus the sooner they level up. In practice, *Intensive Usage* is the average of the time spent to reach a new level, weighted for the maximum level reached to avoid penalizing higher levels (harder to reach).

(W) **Winning Social** measures whether players tend to win more in multiplayer rather than single-player challenges. This tendency is measured as players' win ratio in multiplayer challenges over their win ratio in single-player challenges.

(F) **Full-usage** measures the percentage of game features used by the player. The complete set of the features available is: each transportation means (walk, bike, bus, and train), the invitation system (send/accept invites), customization of single-player challenge, the unlocking mechanism for multiplayer challenges. Some features are made available to players as they advance in the game. Thus, for each player, the Full-usage behavior is evaluated upon the features available to them.

(Cm) **Competitive** measures players preference towards competitive challenges. The value is obtained as the ratio of invites to competitive challenges over the cooperative challenges, both sent and accepted.

(A) **Active** measures players' inclination towards customization. In practice, this value is computed as the ratio of customized challenges (e.g., chosen single-player challenges and multiplayer challenges derived from invites) over challenges automatically assigned by the system - i.e., how active a player was in their tailored gamification experience.

(Cs) **Constancy** measures whether players were constant in their participation rather than having peaks of activity and many non-active days. Constancy is, thus, the percentage of active days within players' gameplay (from their registration day to the last active day).

(Po) **Purpose-oriented** measures how ecological players transportation behaviors are, considering the trips tracked. The purpose-oriented behavior evaluates how in-line players are with the gamification's ulterior motive (or purpose): sustainable mobility. It is computed as the number of green (walk and bike) trips over the trips tracked.

(St) **Striving** measures players will to challenge themselves through difficult tasks. Challenges have associated to them a difficulty value computed according to the challenge's target and the players' history and skills. Striving is computed as the ratio of difficult challenges the player choose.

(Si) **Self-improvement** measures whether players improved their performance over time in terms of green kilometers tracked. The self-improvement value represents the slope of the plot of players' activity: the higher the value, the more drastic the improvement.

Conceptually, the behaviors present some overlap. Thus, we performed Exploratory Factor Analysis (EFA) [238, 68], to ensure data independence and remove redundant information. For instance, the *Active* behaviors models players' tendency to actively customize their challenges, while the *Reactivity* behaviors models players' velocity to reply to game events. Although these events are mostly customization events for challenges, the velocity of players' reaction to them is evaluated. On the other hand, if we find that players either respond very quickly to those events or do not reply at all, Reactivity and Active behaviors become very similar.

Please note that the behaviors do not evaluate behavioral change - i.e., shift from a least to a more sustainable transportation means. Such information cannot be inferred as players can track a subset of their actual trips and could omit some non-green movements. On the other hand, we can measure a usage increase. Despite an increase in green mobility not implying fewer kilometers with other means, getting players used to move by walk or bike may result in preferring them whenever possible.

Proxy

This paper aims to compare idealized and contextualized players' preferences by studying a gamified system. However, this comparison cannot be performed directly. Idealized preferences from Hexad User Types and contextualized preferences from in-game behaviors are computed using different methodologies. Besides, there is no known relationship or correlation between the two preference models. The ultimate goal of player profiling (or modeling) is to understand how to adapt the gaming experience towards the players, which can be achieved by modifying specific game elements. Therefore, comparing idealized and contextualized preferences can be translated into the comparison of the adaptation strategies produced for each player. To perform

Player		Agents	
Representation	*Action Space*	*Adverserial*	*Non-Adverserial*
Characteristics	Skills/Stats	Individual	Companion
Personality	Controls	Managerial	Non-companion
Environment		**System**	
Physical	*Narratological*	*Goals*	*Rules*
Layout	Structure	Explicit	
Appearance	Content	Implicit	

TABLE 4.13: PEAS framework hierarchical structure [206]

this comparison, we used an existing adaptation framework as a proxy: the PEAS framework [206].

The PEAS framework [206] is a design framework informed by a literature review on both games and gamification research. Specifically, they analyzed papers on design, personalization, and player modeling between 2001 and 2018. The framework is hierarchical and is divided into four macro components: player, environment, agents, system. *Player* describes every facet of the player and their character, including their appearance and the actions allowed. *Environment* describes both the aesthetic and narratological aspects of the virtual world. *Agent* describes the game characters not controlled by the players, e.g., enemies or companions. *System* describes functional aspects of the game, as the game dynamics and rules.

The authors also provided a list of guiding questions to assist the framework's instantiation in specific applications.

The PEAS framework has been extended [207] in a generalized model for player profiling to obtain a homogeneous representation of players' preferences over different game elements. Thus, this generalized model allows combining player and personality approaches to produce a single adaptation strategy, assuming that there are no conflicts among the representation. The model is structured in phases, one of them being the model translation. While blending different player profiling methods is out of the paper's scope, the translation function is indeed relevant. Specifically, we provided a translation for both idealized and contextualized preferences into the PEAS representation in the form of a numerical vector.

In the PEAS framework, the model's definition is guided by the elements that can be personalized within the specific application domain. In Play&Go, challenges represent the game element that can be adapted towards the users. We will analyze different challenge adaptation strategies for players, derived from their (1) idealized preferences (Hexad User Types) and (2) contextualized preferences (in-game behaviors).

For the definition of the PEAS representation, the authors define a list of guiding questions [206] revolving around the game element(s) that can be modified.

GQ1. Why are you personalizing your game system?

GQ2. Why did you choose to personalize these game aspects?

GQ3. How will the chosen aspects be personalized?

GQ4. What game aspects have you chosen to personalize? How did you personalize?

We decided to personalize players' experiences to meet their preferences, either idealized or contextualized (GQ1). The game element chosen for the personalization is a challenge, as it is the only game element complex enough to be modified and adapted (GQ2). The adaptation strategy will be computed from (a) Hexad User Types and (b) players' in-game behaviors (GQ3). Finally, we defined the aspect that can be personalized in the challenges by choosing among the elements available in the hierarchical structure of the framework (Figure 4.13). The dimensions of the PEAS representation are the following:

(PC) **Player-Control** refers to players' action space and the control they have over the game. In Play&Go, *Player-Control* can be associated with players' control over challenge assignment mode. Challenges can either be chosen by players from a pool of options or can be automatically assigned in case an explicit choice was not performed. Thus, this adaptation aspect governs the possibility of customizing (player's choice) or personalizing (system's choice).

(SRD) **System-Rules-Difficulty** refers to specific rules of the system. The *System-Rules* sub-component can be further specified in each application domain. With *System-Rules-Difficulty*, we model the desired difficulty level for the challenge, which can either be easy, medium, or hard.

(SRS) **System-Rules-Social** refers to the social aspect of challenges, which can either be present or not. This sub-component governs the challenge and whether it is single-player or multiplayer.

(SRC) **System-Rules-Competition** refers to the social mechanic used for multiplayer challenges: cooperative or competitive. This dimension is relevant only if the System-Rules-Social decides for the challenge to be multiplayer.

(SRG) **System-Rules-Green** refers to the target of the challenge. Specifically, this sub-component decides whether the challenge will be focused on green transportation means (walk and bike).

Please note that in the original paper [207], the authors represented players on four axes, corresponding to the model's components. In this study, we decided to keep the element of the *System* macro-component decoupled. They refer to very diverse concepts in Play&Go and condition different aspects of the adaptation strategy.

Evaluation Metrics

For each player, we will have an adaptation strategy produced by the idealized preferences (player types) and an adaptation strategy produced by the contextualized preferences (in-game behaviors), with the PEAS framework's support. As previously discussed, each strategy will be represented as a numerical vector. Hence, the strategies can be compared using distance and similarity metrics: *euclidean distance*, *cosine similarity*, *hamming loss*, and a modification of *intersection over union*.

The **euclidean distance** is one of the most common distance metric to compute the dissimilarity of objects described by numeric attributes [85]. Its values range from 0 (absolute identity) to the maximum possible discrepancy value, whose upper bound varies with the vectors domain space. As the components of our vectors will be normalized in the range $[0;1]$, for each component, the maximum difference is 1. Given the euclidean distance formula ($d(x,y) = \sqrt{\sum_{i=1}^{d}(x_i, y_i)^2}$), the maximum

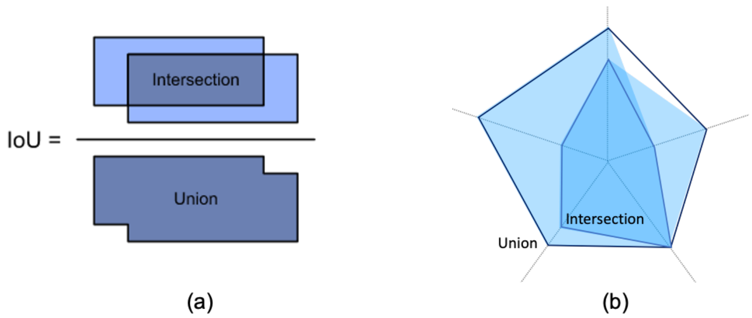


FIGURE 4.22: Intersection over Union

value is $\sqrt{d * 1^2}$, where d is the length of the vector. Hence, the value lies in the range $[0; \sqrt{d}]$.

The **cosine similarity** [85] metric measures how similar two numerical vectors are by computing the cosine of the angle between two vectors. Hence, this value determines if the two vectors are pointing in the same direction. When the vector elements are binary, the function can be interpreted as the number of shared attributes. The cosine similarity values lie in the range $[0; 1]$.

The **hamming loss** [234] is a metric designed for multi-output classification tasks. This metric computes the average difference between the predicted and the true value. Prediction and omission errors are included. The hamming loss values lie in the range $[0; 1]$. A desirable value for the hamming loss is closer to 0.

The **intersection over union** (IoU) [234] metric is widely used in Machine Learning and Deep Learning to evaluate the performance of models for image recognition. In those analyses, recognized images are generally delimited by boxes. IoU computes to what extent the boxes overlap. As the name suggests, IoU is computed as the boxes' intersection area over the union (Figure 4.22a). IoU of 0 means no overlap, and thus bad performance, while IoU of 1 means perfect overlap. Hence, the values lie in the range $[0; 1]$.

We adjusted this metric to our needs, keeping the same idea. We will represent each adaptation strategy as a polygon, where the sides of the polygon are the length of the adaptation vector - i.e., the dimension of the PEAS representation. For instance, Figure 4.22b shows visualize the concept if the dimension of the PEAS representation is 5. Each axis, represented as a dotted line, is one of the five dimensions. Each section of the figure, delimited by two axes, is a triangle, whose area can be computed using the formula $A = a * b * sin\gamma/2$. In the formula, a and b are the values of the triangle's sides adjacent to γ. Each side is the value of the PEAS dimension for the adaptation strategy (i.e., vector) and $\gamma = 360/5$. The area of the irregular polygon is the sum of the triangles' areas.

The IoU follows the same formula, as shown in Figure 4.22a. The area of the polygons' union is obtained by considering for each axis the maximum value between the two strategies. On the other hand, the intersection uses the minimum value for each axis (or dimension). Also in this version, the IoU values lie in $[0; 1]$.

Preliminary Analysis

In the following section, we present the results of the analyses introduced in the previous section.

TABLE 4.14: Factor analysis (structure matrix) for 11 player behaviors in Play&Go (N = 127). The elements in bold represent the behaviors kept.

behaviors	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
R		**-0.83**			
S	**0.75**				
IU			0.52		
W	0.72				
F	0.76				
Cm				**0.54**	
A		0.87			
Po					**0.58**
St					
Si			**0.71**		
Cs					

First, we investigated the existence of correlations between player types and player behaviors through canonical correlations. Then, we instantiated the PEAS framework and provided translations for both player types and behaviors. Finally, we compared the PEAS representation obtained using the similarity and distance metrics previously defined.

Before proceeding with the core study, we conducted (i) an exploratory factor analysis (EFA) to remove redundancy in the player behaviors defined and (ii) researched correlations among player types and the remaining behaviors.

Prior to EFA, we verified the sample size adequacy. Empirical rules suggest having 10–15 participants per variable [68]. In our case, the variables are player behaviors (#11). Hence, we have $\sim$ 13 participants ($N = 127$) per variable. Besides, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .81, meaning that the sample was large enough to perform the analysis, and Bartlett's Test of Sphericity was significant ($\chi^2_{55} = 612.38$, $p < .001$), indicating that the correlations between the variables were large enough. In conclusion, our sample of $N = 127$ was adequate.

The factor analysis was computing using the R *psych* package. We employed an Oblimin rotation because we expected that the components could partially overlap. Moreover, we considered factor loadings $> .512$ (in absolute value) as significant, as suggested in [68] for a sample size of $\sim$ 100 and $\alpha = .01$. An inspection of the scree plot showed a large drop in the eigenvalues after the fifth factor. Hence, we retained five factors.

As Table 4.14 shows, Factor 4 and Factor 5 represent the `Competitive` (Cm) and the `Purpose-oriented` (Po) behaviors, respectively. For factors from 1 to 3, we consulted with the designers to choose the behaviors to retain. Factor 1 was composed of `Sociality` (S), *Win Social* (W), and `Full-feature` (F). The choice fell on the Sociality behaviors, as all three behaviors manifested a strong social component. Win Social's high values meant that players (1) participated in many multiplayer challenges, and (2) they were more motivated to win them. Whereas, Full-feature behaviors, in this specific gamified system, also measured whether players engaged with the social mechanics of the game. Hence, sociality was the characterizing aspect of the factor. Factor 2 was built of `Reactivity` (R) and *Active* (A) behaviors. Both values computed players' will and velocity in making choices within the game. The results of the EFA can be interpreted as: either player made quick choices, or they made no selection at all. Hence, the two behaviors were interchangeable. We decided to keep Reactivity to retain information of player velocity. For Factor 3, we retained `Self-improvement` (Si) over *Intensive Usage* (IU) as the factor loading for *Si* was higher. Besides, the designer hypothesized that the intensity of usage could be related to self-improvements, as players performed more actions over time to achieve better performance.

In conclusion, among the 11 player behaviors, we retained: Reactivity, Sociality, Competitive, Purpose-oriented, and Self-improvement. Those behaviors were then translated into the PEAS representation.

To analyze whether the idealized and the contextualized preferences (user behaviors) generally align, we started by conducting a canonical correlation analysis "CC"). We used the five in-game behaviors as predictors of the six Hexad user types measured by the Hexad user types questionnaire. A CCA can be used to assess the association strength between two sets of variables and allows to compute the multivariate shared variance between them [199]. At its core, the method combines the set of predictor variables (in-game behaviors) and the set of criterion variables (Hexad user types) into separate synthetic variables [199]. The canonical correlation itself is the correlation between these synthetic variables [199]. These pairs of synthetic variables represent canonical functions (CF), which can be seen as an extension of principal components in Principal Component Analyses (PCA), since the CFs are composed of two different variable sets [227]. This process of deriving canonical functions is repeated until either no residual variance is left to be explained or there are as many canonical functions as there are variables in the smaller variable set. Although a CCA has been shown to be able to accommodate variables without relying strictly on multivariate normality [227], we inspected univariate Q-Q plots, skewness, and kurtosis of each variable. The Q-Q plots mainly supported the assumptions of normality, whereas some variables were shown to be slightly skewed. However, all skewness and kurtosis values were within the acceptable thresholds of skewness < 3 and kurtosis < 10 [118].

The resulting model across all CF was not statistically significant using the Wilks's λ=.70 criterion, $F(30.00, 418.00) = 1.28, p = .153$. Consequently, this result does not support the hypothesis that user types and player behaviors are associated. However, the result should also not be seen as supporting evidence for the absence of a relationship. When interpreting Wilks's λ, it seems like the association between types and behaviors is smaller than we were able to detect, given our sample size. Wilks's λ is .70, which hints at a potential shared variance between the variable sets of 30%, i.e. a potential r^2 type effect size of .30. Considering that the recommended threshold for strong effects is $r^2 = .64$ [64], it seems that the user types and in-game behaviors are not strongly associated but might have a small to moderate association [64].

Results

Player Types Translation Players' idealized preferences were modeled through the Hexad User Types [153, 219].

To translate a person's scores for the Hexad model into the PEAS representation, we followed the approach defined in [207]. We defined a mapping between the Achiever score (A), the Philanthropist score (Ph), the Player score (R), the Socialiser score (S), the Disruptor score (D), and the Free Spirit score (Fs) to the components of the PEAS framework, and the resulting weight vector.

TABLE 4.15: Overview of the correlations found among the game elements and strategies relevant to our context and the player types. Cells are blank when the correlation among game element and player type was not investigated in the paper. A dash represent a non-significant or nonexistent correlation. We considered correlation scores >=.2.

(A) Socializer

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		.48		.34			.16
Compet.	.25		.2		.14		.22
Cooperat.	.29		.31		.27	.29	
Meaning. Contrib.		.23				.22	
Customiz.	.31		-	-.3	-		-
Personaliz.	.17		-	-	-		
Rewards	-	-	-	-	-		-
Meaning. Goals		-					
Challenge			-		-	-	-
Learning							-

(B) Philanthropist

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		-		-			-
Compet.	-		-		-		-
Cooperat.	-		-		.24		
Meaning. Contrib.		.34				.-22	
Customiz.	-		-	-	-		-
Personaliz.	-		-	-	-		
Rewards	-	-	-	-	-		-
Meaning. Goals		.17					
Challenge			-		-	-	-
Learning							-

(C) Free Spirit

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		-		-			-
Compet.	-		-		-		.25
Cooperat.	-		-		-	-	
Meaning. Contrib.		-				-	
Customiz.	-		.13	-	-		.2
Personaliz.	.13		-	-	-		
Rewards	-	-	-	-	-		.14
Meaning. Goals		-					
Challenge			-		-	-	.41
Learning							.39

(D) Achiever

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		.28		-			-
Compet.	-		.1		-		.16
Cooperat.	-		.14		-	-	
Meaning. Contrib.		.18				-	
Customiz.	-		.21	-	-		-
Personaliz.	-		.21	-	-		
Rewards	-	-	.2	-.16	.12		.17
Meaning. Goals		.19					
Challenge			.2		.12	-	.45
Learning							.22

(E) Player

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		.26		-			.17
Compet.	.26		.37		.22		.24
Cooperat.	.14		.22		-	-	
Meaning. Contrib.		-				-	
Customiz.	-		-	-	-		.16
Personaliz.	-		-	-	-		
Rewards	.15	.35	.17	-	.15		.3
Meaning. Goals		-					
Challenge			-		-	-	.32
Learning							

(F) Disruptor

	[173]	[217]	[7]	[79]	[8]	[161]	[219]
Scl. Interac.		-		-			.18
Compet.	.11		-		-		.32
Cooperat.	-		-		-	-	
Meaning. Contrib.		-				-	
Customiz.	.14		-	-	-		.14
Personaliz.	.15		-	-	-		
Rewards	-	-	-	-	-		
Meaning. Goals		-					
Challenge			-		-	.34	.21
Learning							

For the mapping definition, we leverage the definition of the player types and the game elements that motivate them [153]. *Player-Control* (PC) is positively linked to Free Spirit and Disruptors, the first characterized by the will to customize their content and the latter by the desire to express their individuality and voice. Thus, having control over the challenge assigned, rather than having them automatically assigned, would accommodate Free Spirit and Disruptor motivational drivers to customization and self-expression. Disruptors *System-Rules-Difficulty* is positively linked to Achiever, as this player type is motivated by a sense of mastery. Achiever's need to feel challenged can be meet by assigning them difficult tasks. *System-Rules-Social* is positively linked to Socialisers, who enjoy playing in teams and engaging with other players. Therefore, multiplayer-challenges are preferred for Socialisers, in contrast to single-players challenges. *System-Rules-Competitive* is also positively linked to Socialiser, who are motivated by competition. *System-Rules-Green* is positively linked to Philanthropist, who seek to make a meaningful contribution. In this context, the meaningful contribution consists of embracing the gamification ulterior motive and preferring green transportation means. Figure 4.23a summarized the translation of player types into the PEAS representation, following the Hexad definition.

The Hexad User Types model has found remarkable consensus in the Games User Research field. Thus, many researchers investigated the correlations among the player types scores and different game elements or mechanics. Some of those correlation analyses produced unexpected results, associating types to other game elements than the ones specified in the definition [153]. For this reason, we produced a second translation deriving from those correlation studies. We collected the published papers using the Hexad User Types questionnaire [219]. As for July 2020, we found 215 papers referring to the survey in [219]. Among those papers, we selected papers written English, investigating at least one game element or strategy useful in our application domain through statistical correlation analyses. This filtering resulted in 7 relevant papers. Table 4.15 visualizes our findings. Each row represents a game element or mechanic, while each column refers to a user type. Each player type column is further divided into seven sub-columns, one for each paper analyzed. Cells report correlation scores, if available, for the game elements and the player types in each paper. Following from Table 4.15, we built a new translation model.

We confirmed the Socializer desire to engage in social interactions [217, 79, 161], but they were both interested in competition [173, 7, 8, 219] and cooperation [173, 7, 8, 161]. As a consequence, *Socialiser* was linked to System-Rules-Social but not to System-Rules-Competition. Socialiser both correlated to personalization and customization, and thus Socialisers were not linked to Player-Control. Finally, despite showing a positive correlation to Meaningful Contribution, no connection with Meaningful Goals or Rewards was found. Hence, Socializer was also not linked to System-Rules-Green. *Philanthropist* desire to make a contribution also appears from the works analyzed, as it positively correlates to Meaningful Goals [217] and meaningful contribution [217, 161]. Therefore, Philanthropist was linked to System-Rules-Green. Moreover, this user type wants to collaborate with others [217, 7, 8, 161], resulting in Philanthropist also being negatively connected to System-Rules-Competition. From the literature, *Achievers*' will to socialize emerges [217], leading to them being linked to System-Rules-Social. They do not show a preference towards personalization over customization [7], but they are positively correlated to learning and overcoming challenges [217]. Hence, Achiever was linked to System-Rules-Difficulty. *Free Spirit* manifested a preference towards customization (user-defined) rather than personalization (system-tailored/automatic [7], translated into a positive link to Player-Control. This type also correlated to competition [219], hence a positive link with System-Rules-Competition. Free Spirit also correlate to Leaning and Challenges, leading to a positive connection with System-Rules-Difficulty [219] *Player* are interested in virtual rewards [173, 217, 7, 79, 8, 161], and thus are negatively linked to System-Rules-Green. They also enjoy social mechanics [217] with a preference for competitive environments [173, 217, 7, 79, 8, 161], translated into a positive link with System-Rules-Competition. Besides, they correlate with challenge [217], resulting in a connection to System-Rules-Difficulty. *Disruptor* not only enjoy competition [217] but also desire to be involved in challenges, gambling, and the rewards that come from winning [161, 217]. Hence, Disruptors link to System-Rules-Competition and System-Rules-Difficulty. Figure 4.23b shows the second translation for player types, using correlation studies existent in the literature.

$$f_{PT-Def}(A, Ph, Fs, D, R, S) = \begin{array}{l} \textbf{PC} = Fs+D \\ \textbf{SRD} = A \\ \textbf{SRS} = S \\ \textbf{SRC} = S \\ \textbf{SRG} = Ph \end{array}$$

(a)

$$f_{PT-Lit}(A, Ph, Fs, D, R, S) = \begin{array}{l} \textbf{PC} = Fs \\ \textbf{SRD} = A+Fs+R+D \\ \textbf{SRS} = S+Ph+R+A \\ \textbf{SRC} = -Ph+R+Fs+D \\ \textbf{SRG} = S+Ph-R \end{array}$$

(b)

$$f_{Behav}(R, S, Cm, Po, Si) = \begin{array}{l} \textbf{PC} = R \\ \textbf{SRD} = Si \\ \textbf{SRS} = S+Cm \\ \textbf{SRC} = Cm \\ \textbf{SRG} = Po \end{array}$$

(c)

FIGURE 4.23: Translation functions

Player behaviors Translation Players' contextualized preferences were modeled through in-game behaviors, as previously defined. For the mapping definition, we associated with the elements of the PEAS representation one or more behaviors, exploiting the designers' suggestions. *Player-Control* is positively related to Reactivity, as we can assume that players performing (quick) choices are interested in customizing their game element. *System-Rules-Difficulty* is positively correlated to Self-improvement, as the will of improving the own performance can be translated into increasingly difficult tasks. *System-Rules-Social* is positively correlated to Sociality and Competitive, as the first behaviors highlights players' will to play with others and the second, in Play&Go, relates to (competitive) multiplayer challenges. *System-Rules-Competitive* is positively correlated to Competitive, for the coherence of the definitions. *System-Rules-Green* is positively correlated to Purpose-oriented, as the gamification goal of the app (purpose) is assuming green transportation behaviors. Figure 4.23c summarized the translation of player behaviors into the PEAS representation.

Comparing the Adaptation Strategies The two translation functions for player types and the translation function for players' behaviors produced 3 PEAS representations for each player. We compared those representations through distance and similarity metrics: euclidean distance, cosine similarity, hamming loss, and (our modification of) intersection of union. Each metric is example-based. Hence, it computed for each entry (player) and then averaged over the whole dataset. Alongside the average, we also report the standard deviation.

First, we investigated to what extend the adaptation strategies for the two player types translations differ. We found that the vectors produced by player types definition and literature correlation are fairly similar, with an euclidean distance of .38 (std = .11), a cosine similarity of .97 (std = .07), a hamming loss of .04 (std = .10), and an intersection of union of .69 (std = .08). Hence, both translations produced similar adaptation strategies, suggesting the model's robustness for idealized preferences (Hexad User Types).

Then, we compared both player types translations to the PEAS representation built from in-game behaviors. This analysis answers the question of whether idealized and contextualized preferences lead to similar adaptation strategies. We observed how both player types PEAS representations differed from the player behaviors representation: euclidean distance of 1.11 (std = .24), cosine similarity of .55 (std = .18), hamming loss of .66 (std = .15), and intersection over union of .23 (std = .09). Considering that we have $d = 5$ dimensions for the PEAS model and that the maximum value for the euclidean distance is 2.24 ($\sqrt{d}$), the player types and behaviors adaptation strategies very divergent. This outcome is also confirmed by the other evaluation metrics, as they present a low cosine similarity and a very high error in the hamming loss metric. Besides, the models overlap only for the 23% (intersection over union). Finally, we compared the the PEAS representation from the player types literature to

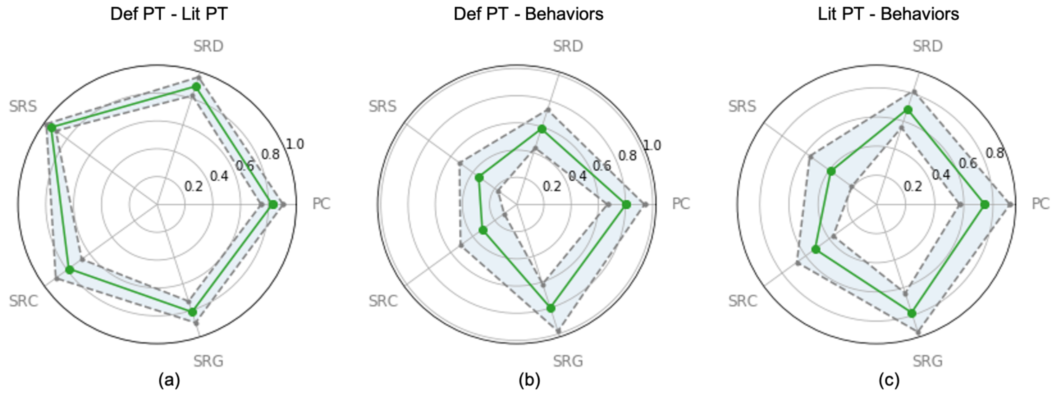


FIGURE 4.24: The spider (or radar) chart compare the three PEAS translations against one another. For each chart, the axes represent a dimension of the PEAS model defined. For each axis, the green dot represent how similar the scores in the axis were for each entry, on average [mean(1 - |difference of the values|)]. The two gray dotted perimeters delineate the mean (+/-) standard deviation of the values. In other words, the closer the value is to 1 for an axis, the higher the similarity of the models for that specific axis, and vice-versa. From the left, we have the comparison of the translations of player types deriving from the definition versus the literature findings. Then, we have the comparison of the player types translation from the definition versus the behaviors. Finally, we have the comparison of the player types translation from the literature versus the behaviors

the PEAS behaviors model. Despite results showed a slightly higher similarity than in the previous test, the magnitude of the change is very small (euclidean distance = .94 (std = .19), cosine similarity = .56 (std = .15), hamming loss = .64 (std = .17), and intersection over union = .27 (std = .10)).

Therefore, players' idealized preferences derived from the Hexad Player Types and players' contextualized preferences extracted from in-game behaviors lead to a diverse PEAS adaptation strategy. In other words, choosing one player profiling model rather than the other would lead to a different tailored content, despite considering the same player. For example, in Play&Go, the user type model could suggest single-player challenges whereas the player behaviors model could suggest multi-player challenges.

From Figure 4.24a, we can see how the PEAS models for both player types translations (definition and literature) produce very similar models. The social aspect is almost identical, while we can observe small divergences for the control, purpose-driven aspect, and the competitive side. Nevertheless, they are unlikely to condition the final adaptation. In the comparison of the player types translations and the behaviors translations (Figure 4.24b and 4.24c) something interesting emerges. First, the translation from the literature of player types is closer to the behaviors than the translation from the definition of player types. Second, both models are more similar for the control and purpose-oriented axis, while they are very different. Although the player type literature translation is closer to the behaviors for the difficulty and competition aspect, there is still a considerable difference. On the other hand, both player types translations differ greatly for the social axis.

4.3 Summary

Telemetry data is a gold mine of information on players' experience, which, to be fully exploited, needs to be processed with care, using the correct tools. While a plethora of research in games made this assertion obvious, the usefulness of gamification data is yet to be investigated, which can be questioned as objectively more modest than the game counterpart. In this chapter, we showed how versatile gamification telemetry data can be in conveying information on how gameplay progresses, both on a global and individual level.

On a global level, gameplay datalogs can be used to achieve a general understanding of the behaviors of the player population and whether they play as intended by the designers. In this regard, we compared two types of behavioral analysis using aggregated and temporal data to understand how the type of analysis answers different questions. Our outcomes showed that the aggregated analysis emphasized the winning strategies successful in the long term, while temporal analysis identified short-term rewarding behaviors. Therefore, if temporal analysis allowed the detection of abnormal, specific behaviors, aggregated analysis shows whether winners behave in line with the designers' goal, essential in gamification. Hence, 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. For instance, by modeling players' behavioral evolution, we found that a considerable portion of the users remained in a low-activity limbo, whereas another part of the population alternated among different but high-activity behaviors. Those players superficially involved in the game (a) can be indicators of a design fault, suggesting that some players' features or preferences are neglected in the games, and (b) are likely to be at risk of churn. Players' retention, on the other hand, dictates the success or the failure of a software system, in that not only acquiring users is more expensive than keeping the current users, but a satisfied player helps in building up the reputation of the product itself. Players' long-term retention in gamified systems is crucial, especially when the ulterior motive is to promote a positive behavioral change. Predicting players' churn helps in identifying users at risk of abandoning the game to allow automatic or manual intervention to occur. Towards this, we provided evidence that data describing players' activity can be used to predict churn, although not as complex and multifaceted as it happens in games. Our results show that players' time investment in the game, like length and constancy of previous gameplay, is a predictor of retention, not only in our gamified example but also in a free-to-play game (TagPro), used as a validation set. This slight analogy in the results supports the idea that the same motivational affordances move players in both games and gamified systems. Therefore, this encourages research to investigate the re-use and adaptation of models and approaches used in games research. Besides moving an additional step towards Gamification Analytics, we continued to show data preprocessing and the methods used impact the results, hence the need to gain a greater awareness also when it comes to gamification data.

Despite achieving a global understanding of players is essential to assess the progress of the gameplay, adaptive gamification also requires more player-specific information to deliver tailored content. Following along the lines of gamification providing limited data, and thus resources, to learn players' preferences or profiles, we investigated whether the available information was sufficient for the purpose. Hence, we contributed with an RL-like algorithm exploiting simple metrics like the level of activity to learn players' preferences. The outcomes lay good premises for future works, as they assess the feasibility of using participation data to inform players' preferences in terms of weekly Play&Go challenge. The results also show that data-driven approaches are successful when players have a strong preference in the merit of the type of game element they are interacting with. While this algorithm is relevant when there is a single adaptable element with a finite set of types to choose among, deriving a complete profile or adaptation strategy is more challenging. Towards player profiling, especially in gamification, the use of theoretical or data-driven approaches is still debated, supported by psychology literature on perception of the self and contextualized preferences but yet to be verified in practice. Hence, we took a step back and systematically researched whether they lead to divergent adaptation strategies. Hence, theory-driven and data-driven approaches do not move on parallel lines. Instead, they may encounter and merge into shared results. Nevertheless, researchers and practitioners should be aware that the choice of using player types or player behaviors to tailor game content will most likely impact the final results. Our results also prove how the discussion and research on adaptive gamification are still opened. Further studies should be aimed at producing concrete evidence or guidelines to inform the choice of the approach to use to model players.

Chapter 5

Understanding Players as Part of the Community

In the previous chapters, we have analyzed players' experiences from different perspectives. We presented methods and approaches to reach a global understanding of the state and progress of the gamified application, as well as a closer view of each player's gameplay. Hence, we can now estimate their engagement by analyzing users' in-game activity and predicting their likelihood of abandoning the game. We can learn and extract their preferences by studying their interaction patterns. Yet, those perspectives are still focused on the individual, considered as an isolated unit. Games and gameful environments, on the contrary, build a virtual world where, at least, indirect social connections are formed. Consequently, a more accurate description of players' identity and personality should also include information on how they affect and are affected by the virtual society.

The following chapter presents our study on social influence in gameful environments. First, we illustrate the algorithm used to identify influencers. Second, we introduce our preliminary study on the algorithm conducted on the online multiplayer game Destiny, in which we detected retention influencers. Then, we replicated and extended the preliminary analysis using data from a game provider (Steam). Those studies allowed investigating a novel approach to identify in-game social influence, availing of large and heterogeneous datasets. Finally, we researched influence in gamification, exploiting data from multiplayer challenges in Play&Go. The study of social influence in gamification further extends the influence(rs) analysis by researching other types of influence, other than retention.

5.1 Measuring Social Influence

The SNA literature features two lines of thought to characterize and identify influencers, either by considering the network's structural nature or the evolution of users' behaviors. We will refer to the first approach as *structural*, and the second as *semantic.*

The structural approach consists of computing centrality metrics on the player social network. Generally, degree centrality, closeness centrality, and betweenness centrality are considered the most robust. Therefore, structural influencers are nodes well-positioned in the network, with many connections and placed along many short paths. Central nodes are supposedly visible and popular across the network. The pioneering study on retention influencers in gameplay adopts this method [37]. Specifically, they combine degree, closeness, betweenness, and eigenvector centrality with PageRank.

In the semantic approach, rather than relying on the network's topology, users' behaviors are analyzed and monitored over time. Those behaviors need to be defined in advance and describe the type of influence that will be computed. The choice of the behavior(s) highlights the main difference and strength compared to the structural method. Structural influencers are inherently assumed to affect any choice and habit of others. Hence, those methods consider the status of *influencer* as portable. On the contrary, semantic methods enable the possibility of having different sets of influencers, varying the behavior(s) analyzed. The behavior can describe high-level properties, such as players' activity to study retention or represent context-dependent features. For instance, in gamification, designers may be interested in propagating behaviors in line with the gamification goal. Besides being malleable, the semantic approach allows more fine-grained analyses. The evolution of specific features is exploited rather than simply the skeleton of the social network - i.e., the graph.

This thesis analyzes players' behaviors from different sides and angles, including the impact of social relatedness. The semantic approach better integrates with this vision. Therefore, informed by the SNA literature, we designed an algorithm to compute semantic influence within players' networks. Despite semantic influence being our main focus, we also evaluate our findings against the structural influence results to better comprehend the differences among the two methods in our application domains: games and gameful environments.

Following the definition of influencers as users conditioning others' behaviors, our algorithm [144] computes the semantic influence exerted in the network by measuring changes in nodes' properties successive to social interaction. The algorithm (Algorithm 3) requires a dynamic graph and a temporal dataset of players' properties. A dynamic graph is built of a series of snapshots of the network taken at regular intervals. Hence, dynamic graphs embed information on the network topology evolution - e.g., appearance and disappearance of links and variations in links weights. The temporal dataset holds information on players' attributes and how they evolved across the same snapshots. It follows that, prior to the analysis, a time window must be defined. The analyst must decide the level of granularity desired, such as daily, weekly or monthly. The choice is context-dependent.

Allowing the choice of the players' properties and the time window makes the algorithm easily portable and adaptable to various application domains. Even within the same use case, several types of influence can be analyzed by tuning these parameters.

Alongside the definition of the algorithm, we provided a *Python* implementation, publicly available on Github, named `sinfpy`[1].

Semantic Influence Detection (`sinfpy`) computes the influence scores as values in $[-1; 1]$. The lower the value, the more susceptible the node is to influence. The higher the value, the more the node exerts influence on its neighbors. The algorithm is built upon the concept of influence as an increase of similarity over time. In other words, influence over an edge $e(v, w)$ occurs if v is more similar to w on time t, and they got connected on $t - 1$. The influence, in this case, is positive for w and negative for v. The similarity increases determine the magnitude of the influence. Timed information of players' properties is needed to compute the influence. The properties are user-defined, and thus, the concept of influence is tied to the properties of interest. The network can either be dynamic or static in terms of the connection among the

[1]Further details and a Python implementation are available at `https://github.com/enrlor/sinfpy`, as the Pypi package.

Algorithm 4 Influence Algorithm

1: **procedure** InfluenceScore(E, X)
2: **for all** $e(i,j) \in E$ **do**
3: **for all** $t \in \{t_1, ..., t_k\}$ **do**
4: $value = EdgeInfluence(X_i^{t-1}, X_i^t, X_j^{t-1}, X_j^t)$
5: $inf_i(e) = InfluenceAdj(value, w(e))$
6: $inf_j(e) = -inf_i(e)$
7: **end for**
8: **end for**
9: **end procedure**

nodes. In the case of dynamic connections, the persistence of a node's disappearance is taken into consideration.

The algorithm develops into two phases, which are implemented in two modules: the `EdgeInfluence` and `NodeInfluence`.

Phase 1: *Edge Influence.* In the first step, the algorithm computes the influence score for each edge (Algorithm 3) through the `EdgeInfluence` module. The module computes the influence exerted on each edge of the network as a result of the nodes' behaviors and interactions over the observation period. In the `sinfpy` implementation, the module foresees a number of optional parameters to meet context-specific requirements. While the class encloses a default function to compute the edge influence, it can be customized to better adapt to the use case.

Listing 1. Default influence function.

```
def influence_score(xi_old, xi_new, xj_old, xj_new, prev_inf):
    influence = 0
    threshold = 0.8

    sim_i = similarity(xi_old.values,
                       xi_new.values)
    sim_j = similarity(xj_old.values,
                       xj_new.values)
    sim_ij = similarity(xi_new.values,
                        xj_new.values)

    if (prev_inf > threshold) or \
        (sim_i <= threshold and sim_j > threshold) or \
        (sim_i > threshold and sim_j <= threshold):
            influence = sim_ij if sim_i > sim_j else -sim_ij

    return influence
```

The default function assumes that (i) the values are already normalized, (ii) all the properties in the dataset X are all relevant for the computation of the influence score, and (iii) the similarity function is *cosine similarity*, if not specified otherwise in the initialization phase. For each interaction, modeled through an edge $e(i,j)$, the function (Listing 1) analyzes the similarity of the nodes with their past behaviors (*sim_i* and *sim_j*) and between one another (*sim_ij*). The last condition behaves as follows. If only one among i and j maintains a coherent behavior with its past (e.g., *sim_i* $\geq$ threshold), whereas the other changes (e.g., *sim_j* $<$ threshold), influence is exerted and the magnitude of such influence is *sim_ij*. Influence is also assumed to occur when the influence has already been detected in the past (*prev_inf* $>$ *threshold*). In this case, we do not expect one node to consistently behave while the other emulates them, as this already occurred in the past. Rather, we inspect the eventual persistence (or decay) of influence by computing the current similarity among the two nodes. The edge influence score in magnitude refers to both nodes, yet one is the additive inverse of the other. Thus, a convention is applied. The sign of the edge influence value for $e(i,j)$ refers to i, where $id(i) < id(j)$ - i.e., the id label of i is lexicographically smaller than the id label of j. The control in the final assignment ensures that the edge's value refers to i, the node with the lower id label by convention.

The user-defined function must consider said convention and comply with the following signature:

- x_i^{t-1} properties values for node i prior the connection
- x_i^t properties values for node i at the time of the connection
- x_j^{t-1} properties values for node j prior the connection
- x_j^t properties values for node j at the time of the connection
- threshold; as defined in the instantiation phase
- $influence^{t-1}$ the influence score, if any influence was exerted in the past
- similarity_fun; as defined in the instantiation phase, used to compute the magnitude of the influence

Listing 2. Optional adjustment function.

```
def balance_influence(influence, weight, penality = 0.1):
    penalized_inf = influence * penality
    return float(influence - (penalized_inf * (1 - math.log(weight +
    ↪  1, 2)/weight)))
```

Line 4 of Algorithm 3 allows the definition of a penalty factor, adjusting influence scores for long-term interactions among nodes. This adjustment is optional but can be used to space users that produced sudden changes in others' behaviors from players impacting others after repeated interactions. For instance, the default adjustment function (Listing 2) is designed to expect increasingly higher influence values with more interactions over time. In other words, it models a reinforcement in player influence using a logarithmic function.

Phase 2: *Node influence.* Finally, we computed the influence score for every node. The class `NodeInfluence` computes the final value of the influence score for every node as an aggregate of the influence of the edges they were involved in. We computed influence on the edges for all snapshots. Then, we computed the influence score of each node as the average influence they have on or are subjected by their neighbors.

$$Influence_i = \frac{\sum_{j:e(i,j)\in E} inf_i(e)}{\sum_t deg(i^t)} \tag{5.1}$$

Similarly to the edge influence scores, the node influence values lie in the interval $[-1;1]$. High positive values identify influencers, while low negative values depict susceptible users. Neutral scores represent a lack of influence, either exerted and perceived.

5.2 Understanding Social Relationships within Games

The algorithm presented in the previous section (Algorithm 3) reflects the definition of social influence as a force impacting users' behaviors and choices. Yet, neither the existence of those influencers or their impact on others' retention can be assumed to occur in games or gamified systems. Therefore, we conducted a preliminary investigation on game telemetry data [144] to verify whether (a) a set of highly influential users exists and (b) they have a stronger impact on others retention compared to non-influencers. Then, we investigated if influencers impacted retention and activity across multiple game titles and if they conditioned the choice of playing a new game by analyzing data from the game provider Steam [137].

5.2.1 Influencers in Online Multiplayer Games: Destiny

Our first use case scenario is the online multiplayer game Destiny, from which we analyzed PvP Crucible matches. The primary purpose of this investigation is to compare the structural and semantic approaches to find influencers. Destiny data suits our study as it is comparable to the game "Tom Clancy's: The Division", for which the first investigation on structural influencers has been conducted [37]. In this section, we first present the structure of the player social network (PSN). Then, we present the features used in the semantic influence computation, as well as our evaluation metric. Finally, we describe and compare the set of central users, retrieved through the structural approach, and the set of influential users, identified by the semantic algorithm.

Constructs and Measures

In this study, we were interested in retention influencers, impacting others' in-game activity in the game. As a result, the behavior monitored by the semantic influence algorithm is players' participation. In practice, we exploited the following features: number of PvP Crucible matches, the time spent between matches (on average), the time spent in a match (on average), and the percentage of matches completed. Those properties describe players' activity over time and can be used to model eventual behavioral change. Yet, they provide no information on long-term retention. A player may be very active in a specific moment but abandon the game soon after. As a consequence, users with high influence values, according to our algorithm, cannot be assumed to impact others' permanence in the game. Rather, for those individuals, the algorithm found that players getting in contact with them modified their activity to resemble them at a specific point in time.

Therefore, we defined a custom metric to evaluate influencers' impact on long-term retention, which we named `retention transfer`. The word transfer has a specific purpose, as the metric is designed to model both a positive and negative impact on retention. Hence, good values in the `retention transfer` metric mean that the influencer transfers their retention to others: players leave when the influencer abandons the game. As a result, the following scenarios are modeled. When the influencer remains in the game for long periods, the influence is positive and long-term retention is achieved. On the other hand, when the influencer churns prematurely, other players are also drawn to abandon the game. Although the goal is to promote long-term retention, we cannot ignore that influence can also be negative. Being an influencer does not imply that they appreciate the product or, in our case, are engaged in the game. However, the semantic algorithm is designed to find influencers, regardless of them exerting positive or negative influence. As this metric aims at evaluating the algorithm performance in finding individuals conditioning others' retention, both negative and positive influence should be considered. Nevertheless, this does not represent a limitation, rather a strength. Finding negative influencers is even more important, as strategies to isolate or limit their effect may be actuated.

The `retention transfer` metric is computed as follows:

$$rt_i = \frac{\sum_{j \in \mathcal{N}} |gameplay_i^t - gameplay_j^t|}{|\mathcal{N}|} \tag{5.2}$$

where $\mathcal{N}$ is the set of i's neighbors, $gameplay_i^t$ and $gameplay_j^t$ are the length of the gameplay of the nodes i and j, respectively, after they first connected at time t. The values of $gameplay_i^t$ and $gameplay_j^t$ are the relative gameplays length of i and j expressed as the number of days they played from the moment they connect to their last active day. Hence, the formula compares the abandonment point of player i with each of the other users playing with i at least once, and then averages the differences.

The value of the metric lies in the interval $[0, \infty)$. Values close to zero represent a perfect retention transfer, whereas the higher the value, the more players' retention is unrelated. In other terms, smaller values show that, on average, the player's neighbors remained in the game as long as the player kept playing.

TABLE 5.1: General Network Properties

Property	Value
Nodes	10K
Edges	26K
Average Degree	4.60
Average Weighted Degree	68.18
Diameter	18
Modularity	0.97
Connected Components	1.5k
LCC	42% of the network
Second LCC	41% of the network
Average Clustering Coefficient	0.55
Average Path Length	6.61

Network Structure

The Destiny player social network is built from the PvP Crucible matches. The network nodes represent players connected through an edge if they were teammates in at least a match. Conversely to teammates, opponents cannot be chosen by players; rather, the game automatically matches two teams. Therefore, opponents were not connected in the network. The edges are undirected and weighted, where the weight describes the number of matched nodes played together. Table 5.1 reports basic statistics of the player network (See Chapter 3 for further details).

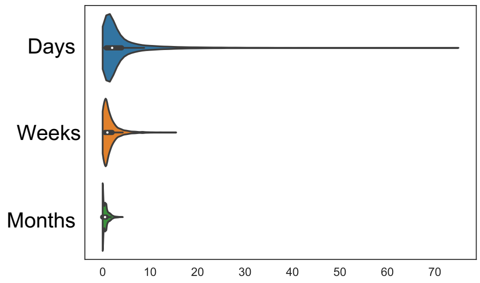


FIGURE 5.1: Peaks

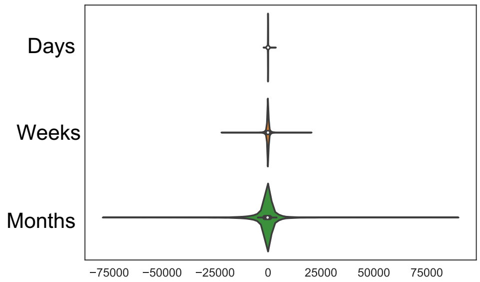


FIGURE 5.2: Slopes

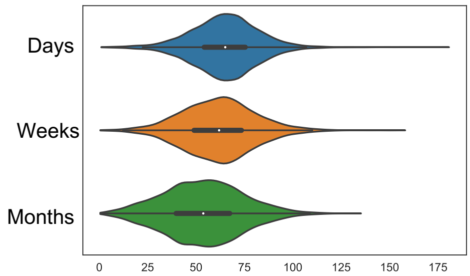


FIGURE 5.3: RSTD

FIGURE 5.4: Players' participation stability throughout their gameplay.

For this study, we built two versions of the PSN: static and dynamic. The static network was needed to find central users using the structural approach. The dynamic network, on the other hand, was required for the semantic algorithm (See Section 5.1).

A dynamic network can be modeled as a sequence of snapshots, taken at a regular time interval, without information loss [197]. Consequently, before the influence analysis, we had to identify the best time window for our dataset, as there is no universal definition. We studied three levels of granularity: daily, weekly, and monthly snapshots. The time window choice was dictated by the need to have data representing players' behavioral change without excessive noise. In other words, snapshots should neither be distant enough to produce a data flattening or too close to include too many details. Our decision was guided by a preliminary analysis conducted on players' in-game activity - i.e., the same features extracted for the semantic algorithm. Specifically, we studied the activity behavior by analyzing sudden changes, tendency, and variability. Sudden changes were modeled as drastic oscillations between high and low values in the activity features, visually represented as peaks in the distribution plot. The tendency is measured as the slope of the distribution - a lack of slope represented a lack of change, while positive and negative slopes represented a gradual variation in the behavior. Finally, the variability is computed using the relative standard deviation (RSD).

Table 5.2: Distribution of the centrality measures values

	Distribution				
	min	*25%*	*50%*	*75%*	*max*
DC	1	2	4	7	92
CC	0.08	0.14	0.16	0.17	1
BC	0	0	523	8.9k	634k
EC	0	0.005	0.01	0.03	1
PageRank	2.4e-5	6.6e-5	9.1e-5	1.1e-4	1.2e-3

Figure 5.4 shows the results. Players' participation level, expressed as frequency and number of matched played, is very variable across different days and quite stationary across months. Weeks, on the other hand, are less sensitive to change than days but retain more information than months. Therefore, we decided to represent the dynamic network using weekly snapshots.

Structural Approach

First, we researched structural influencers – i.e., players well-positioned in the PSN. Those users, or nodes, are crucial to keeping the community connected and, generally, have an important role. In the SNA literature, degree centrality (DC), closeness centrality (CC), betweenness centrality (BC), and eigenvector centrality (EC) are considered the most robust centrality metrics (CM) [197]. DC measures the number of connections a node is involved in. CC assesses how accessible a node is from the others. BC computes the number of the shortest paths passing through the node. EC evaluates how many of the node's connections are *important*, where a connection is deemed important if the other node has a high centrality value. We also included PageRank to ease the comparison to the first work on structural influencers in games - i.e., [37]). PageRank is computed considering the portion of the network directly accessible in a single step.

We computed the centrality metrics for each node of the network (Table 5.2). As expected, a minority of the nodes had high values in the CM. As a result, the distribution plots presented a long right tail, exception made of CC, which is bimodal and peaks at 0.1 and 1 - with the second peak being much lower than the first.

The structural influencers, or central nodes, were selected as the nodes with high values in all CM. We defined a threshold for each CM, considering the top 10%, top 1%, and top 0.1% of the distribution of the scores. Then, we retrieved the intersection of all the sets of the most central players according to each CM. This resulted in a set of 51 central players. When considering the intersection of the players in the top 1% and 0.1% of each centrality measure, the seats were empty.

Semantic Approach

In our second analysis, we identified semantic influencers as players affecting in-game behaviors of the individuals they connect to - i.e., they play in a match as teammates. We used the *Semantic Influence Algorithm*, as defined in Section 5.1. The algorithm allows us to specify which kind of influence we want to investigate. In this work, we researched retention influencers, impacting others' permanence in the game. The choice of computing retention influence derived from Canossa et al.'s [37]

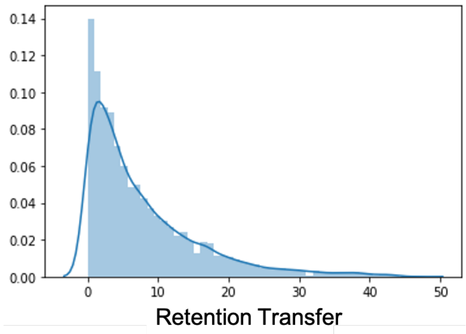


FIGURE 5.5: Population

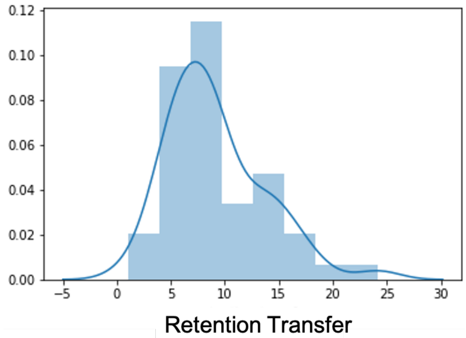


FIGURE 5.6: Structural Influencers

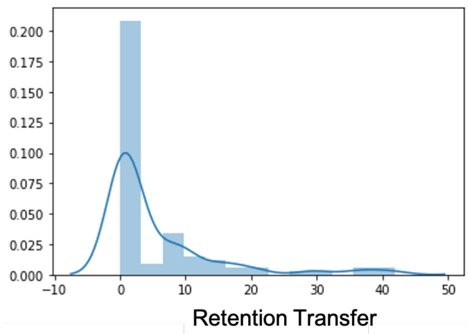


FIGURE 5.7: Semantic Influencers

FIGURE 5.8: Distribution of the retention transfer scores for all players (a), central players (b) and top 1% influential players.

work also studying retention. Hence, investigating the same type of influence eases the comparison to the state-of-the-art findings.

Since we are interested in retention and participation, we computed the similarity on participation metrics - i.e., number of matches, the time between matches, completion rate, etc. Similarly to the structural approach, we investigated three thresholds to select the subset of semantic influential players, considering the top 10%, 1%, and 0.1% of the influence distribution [110, 229]. This led us to a sample of 1000, 100, and 10 influential players, respectively.

Results

From the analysis of the PSN built of the *Destiny* PvP Crucible matches, we retrieved two groups of players: the structural and semantic influencer.

To characterize the groups, we conducted some exploratory analyses using the Mann-Whitney U test. The null hypothesis for the test is that there is a 50% probability that a randomly drawn member of the first population will exceed a member of the second population. The alternative null hypothesis can be double-tailed (the two samples come from the same population - i.e., both have the same median) or single-tailed (the values in one population are higher than the other). If the data sample is small or if the data do not follow a normal distribution, the Mann–Whitney U-test, rather than the t-test, provides the most accurate estimates of significance [205].

First, we found that central players showed significantly lower scores in the semantic influence value in contrast to the semantic influencers ($U = 1281, p = 2.3e - 08$). Similarly, we found that for almost all the CM computed, the influential players manifest lower values than structural influencers (DC $U = 0, p = 3.3e - 27$; BC $U = 0, p = 2.5e - 32$; CC $U = 2450, p = 0.38$; EC $U = 0, p = 3.5e - 24$; PageRank $U = 0, p = 1.9e - 24$). An exception is closeness centrality, for which the difference was not statistically significant. We also observed that the intersection of the structural and semantic influencers was found empty. We also found that semantic influencers were involved in few recurrent links (many matches over time), conversely to central users. The average weighted degree value was statistically higher for semantic influencers than central nodes ($U = 3k, p = 0.26$). Following this result, we also verified whether the values of the standard deviation of the edge influence differed in the two groups of players. The hypothesis was that structural influencers, by definition, are in contact with many more other players, and only part of them might have been susceptible to their influence. On the other hand, we expected this variability to be lower for semantic influencers, as they were in contact with fewer users. From the test, central players showed a significantly higher variability that semantic influencers ($U = 1281, p = 1.3e - 08$). This finding suggests that the strength, or even existence, of influence also varies according to the other user, which may not be susceptible to influence.

Finally, we computed and compared the two sets of influencers and the population through the `retention transfer` values, used to assess the influencers' impact on long-term retention. Figure 5.8 shows the distribution of the values in the population, among the 51 central nodes and among the 101 influential users. The distribution of the values for the structural influencers (Figure 5.8b) skewed away from 0. Thus, the permanence of central players in the game was unrelated to the permanence, or churn, of their connections. For semantic influencers, on the other hand, the distribution of the `retention transfer` for the influential players peaked at the value of 0, with a long right tale (Figure 5.8c). Therefore, when an influential player left the game, on average, their neighbors followed, and they tended to remain if the influencer kept playing.

5.2.2 Influencers across Games: Steam

Our second use case scenario is the game provider Steam [137]. Having assessed retention influencers' existence within a specific game, we researched whether influencers have a broader range of action. Steam fits our purposes as it allows players to store multiple games in their private library, where information on their activity in each game is also stored. Besides, a social network is formed as users can add friends, of which they can visualize games and progresses, and join groups. Steam allows identifying influencers at a higher granularity level - i.e., impacting activity and retention in multiple games, rather than one specific title. Moreover, being Steam a game library, we can study whether those influencers, if any, also condition players' choice to play new games.

In this section, we present the social network, the features analyzed, and our characterization metrics. Similar to our previous study, we researched both central nodes and influential nodes. Finally, we compared and characterized the two sets.

Constructs and Measures.

In this study, similarly to the study on Destiny PvP Crucible matches (See Section 5.2.1), we also researched retention influencers. Conversely to the previous work, instead of collecting detailed data on players' in-game activity, each user was characterized by a features vector of variable dimension. Each element of the vector represents an owned game, and the value is the time spent playing that game within the observed time window.

To evaluate and understand the properties of the influencers sets obtained, we studied *influentiable behaviors* in Steam. As in our previous analysis on retention influencers (e.g., [144]), we measured the impact influencers had on their neighbors using the `retention transfer` metric. In other terms, this metric shows to what extent nodes emulate the influencer they were connected to, both in case of retention and abandonment of the game. Besides monitoring players' retention, Steam, which was born as a game library, allows us to track the acquisition of new games. Therefore, we measured whether influencers affected the choice of playing a new game by defining the `new games` metric. This value is computed as the number of players who played at least one game after the influencer played it. Long-term retention in those newly acquired games is also desirable. Hence, we also defined and computed the `retention in new games` metric to determine the average time the influencer's neighbors spent in the new games. For each new game, the value is calculated as the days spent playing over the days that the new player could have remained in the game in the absence of a churn. Small values indicate an early churn, whether a value of 1 indicates that the users were retained in the game until the end of the observation period.

Finally, we computed the following characterization metrics to model influencers' behaviors in Steam and detected any differences among the influencers set and players' population. The metrics are summarized in Table 5.3 and can be organized in subgroups: *players' activity* and *players' library composition.*

Players' activity was studied as the time they invested playing using the number of hours they played, the hours spent playing multiplayer games, the hours spent weighted for the popularity of the game, and hours spent playing popular games. The popularity of a game is measured using the review ratings. *Players' library composition* was modeled through the ratio of owned popular games over the unpopular, and by computing, for each game genre, the ratio_*genre* as the percentage of the games owned corresponding to that genre.

In the remainder of the section, we used the Mann–Whitney U-test hypothesis testing methodology to compare the results of the methods for the influencers detection.

Network Structure

The Steam PSN was built from the friendship relationships existing in the platform. The Steam friends network was modeled as an undirected graph in which the nodes are the $39,354$ players resulting from the data cleansing, and the edges represent a

Table 5.3: Behaviors' description and distribution over the population of players analyzed.

Behavior	Description	Mean	Std
timeplayed	Describes player' level of activity as a gamer and the time investment they make to Steam.	4269.23	3814.71
weighted_time	Describes players' time investment in popular games, which are games gaining a high consensus in the Steam gamer community.	3763.67	3674.18
ratio_time_popular	Describes players tendency to invest time in popular (or mainstream) games.	0.84	0.25
avg_game_popular	Describes players tendency to buy popular (or mainstream) games. While the previous behavior evaluates time investment, this refers only to the purchases. The player may play very little.	0.85	0.08
ratio_Action	Describes the preference for the Action genre, as the percentage of owned games labeled as action games.	0.69	0.34
ratio_Adventure	Describes the ration of owned Adventure games.	0.24	0.30
ratio_Casual	Describes the ration of owned Casual games.	0.06	0.16
ratio_FreetoPlay	Describes the ration of owned F2P games.	0.30	0.37
ratio_Indie	Describes the ration of owned Indie games.	0.30	0.33
ratio_MMO	Describes the ration of owned MassivelyMultiplayer games.	0.09	0.21
ratio_RPG	Describes the ration of owned RPG games.	0.19	0.29
ratio_Racing	Describes the ration of owned Racing games.	0.06	0.18
ratio_Simulation	Describes the ration of owned Simulation games.	0.16	0.27
ratio_Sports	Describes the ration of owned Sports games.	0.06	0.19
ratio_Strategy	Describes the ration of owned Strategy games.	0.17	0.28
ratio_EarlyAccess	Describes the ration of owned EarlyAccess games.	0.05	0.15
ratio_time_mp	Describes how social the player is, and thus, whether they prefer multiplayer or single-player games.	0.09	0.21
retention_transfer	Describes the whether player's affected the long-term retention of their neighbors. Values closer to zero indicate that others left the game if and when the player abandoned.	2.91	4.96
ratio_neigh_new_games	Describes the local influence of players on their neighborhood, as it measures the percentage of neighbors that played a game AFTER the user played that game.	0.24	0.32
avg_retention_ngames	Evaluates the time investment in those new games the neighbors played after the player.	0.58	0.34

friendship status among the two nodes. In the graph, we have $218,432$ edges, resulting in a network density of 0.00028 and a clustering coefficient of 0.11.

Each player is characterized by a features vector of a variable dimension, describing the time spent playing each game they own. Conversely to our study on *Destiny* [144], the network's dynamicity cannot be modeled into its topology but only through players' properties or features. The network's topology is static as friendship relationships are a more permanent connection than in-game interactions - i.e., shared PvP matches. Additionally, our observation period is relatively small, hence the lack of major changes in the friends' graph. Changes in nodes' features, on the other hand, can be tracked, as we collected daily updates of players' activity. Hence, we could rebuild users' play habits throughout the 5-week time window we analyzed.

Structural Approach

We identified structural influencers in the Steam friends network using a combination of centrality measures, similarly to previous works [37, 144]. Table 5.4 shows the distribution of the centrality measures computed across the population of players. While the degree and the betweenness centrality values tended to be very small in the whole population, the closeness centrality was slightly better distributed.

The 394 players exhibiting high values (top 1%) in the CM scores were labeled as structural influencers. They showed a significantly higher retention transfer value than the population ($U = 5885905.0, p - value = 1.15e - 16$). Hence, the neighbors of semantic influencers are indifferent to their permanence in the game, on average. Central players also have a higher *new games* metric values than the remainder of the population ($U = 1407981.5, p - value = 5.30e - 196$). However, the number of connections they have is generally larger than the other users as they are central, by definition. When the percentage of the neighborhood that played the new games is measured, the value is lower than in the population ($U = 6188878.5, p - value = 0.01$). In addition, there is no statistical difference between the value *retention in the new games* metric between structural influencers and the population.

Semantic Approach

We used the *Semantic Influence Algorithm* to find semantic retention influencers, as defined in Section 5.1. Although the network topology of the Steam friends graph is static, nodes' features vector may change over time. The vectors describe the amount of time players spent playing their games, which is updated daily. In the semantic algorithm, the influence is computed by observing players' behaviors – i.e., playtime – in games both nodes owned. The final semantic influence scores showed many nodes having a neutral influence on others, with few players having very polarized values (mean = 4.8e-4, std = 0.09, min = -0.96, 25% = 0.00, 50% = 0.00, 75% = 0.00, max = 0.96).

Semantic influencers, also in this work, were identified as players with high influence scores. However, many of the top 10% users in the influence rankings were neutral to influence (value of 0). Hence, we refined our sample by selecting as influencers players in the top 1%. The resulting set counted 394 users, with influence scores in the range $[0.28; 0.96]$.

Conversely to structural influencers, semantic influencers had a significantly lower value in the `retention transfer` metric than the population ($U = 11939980.5, p -$

TABLE 5.4: Distribution of the CM used in the Steam Friends graph.

	Min	**25%**	**50%**	**75%**	**Max**	**Mean**	**Std**
DC	0.000025	0.000025	0.000076	0.000203	0.022260	0.000282	0.000687
CC	0.000025	0.156729	0.178468	0.207109	0.279998	0.174728	0.049302
BC	0	0	0.000004	0.000056	0.063842	0.000101	0.000557

$value = 2.22e - 77$). In other words, they showed a significantly greater impact on the long-term retention of their neighbors, on average. Those influencers also impacted their connections' choices in terms of new games. Although the absolute number of players influenced to engage in a new game is smaller than in the population ($statistic = U.0, p - value = 0.01$), the percentage of neighbors who tried the new game was significantly larger ($U = 6178900.0, p - value = 8.82e - 14$). The results are due to the neighborhood of semantic influencers being tendentially smaller than the network. Of the 394 influencers, 237 pushed at least one neighbor to try a new game. In addition, those influencers retained players in the new games for longer than the average node in the population ($U = 1548228.5, p - value = 3.89e - 17$).

Results

In the following section, we compare the semantic and structural influencers sets, which were completely disjointed. We observed that structural influencers exerted less influence than semantic influencers, according to the influence scores (U=154449.0, p-value=$1.33e - 133$), whose impact on their neighbors showed a higher variance than in the central users (U=109616.0, p-value=$1.19e - 27$). Coherently to previous findings, semantic influences also showed significantly lower values in the centrality measure values (DC (U=1006.0, p-value=$6.40e - 129$), CC (U=5866.0, p-value=$1.23e - 111$), and BC (U=0.0, p-value=$6.367e - 133$)).

This work researched retention influencers and used the `retention transfer` metric to evaluate players' impact on permanence in the game. Semantic influencers had lower – i.e., better – values than the population (U=35495.0, p-value<.001) and the set of semantic influences (U=35495.0, pvalue=.001). They also influenced more neighbors to try new games than the population (U=9358528.0, pvalue<.001) and the set of central users (U=86651.5, p-value<.01). In addition, users that played a game after their semantic influencers tended to be retained in the game for longer than users conditioned by structural influencers (U=2887234.0, pvalue<.001).

Finally, we observed users' behaviors within Steam to detect any dissimilarities between the set of influencers and the population (summarized in Table 5.5). In terms of activity and playtime, semantic influencers were more active than both the average user in the population (U=10317643.5, pvalue<.001) and semantic influencers (c=90035.0, pvalue<.001). They also were keener to play games with positive reviews for longer (U=10075956.5, pvalue<.001, and U=89812.0, pvalue<.001). Semantic influencers, in percentage, owned more popular games than structural influencers (U=93115.0, pvalue<.001, and U=91595.5, pvalue<.001). Yet, the values do not significantly differ from the populations' values.

Players' sets also differed in their preferred game genres. Semantic influencers collected more Action, F2P, Massively Multiplayer, RPG, Strategy, and Early Access games than the players' population (pvalue<.001). In comparison, structural influencers possessed more Adventure, Casual, Indie, RPG, Racing, Simulation, Sports,

TABLE 5.5: Characteristics and behaviors of Steam's influencers. For the columns *Semantic* and *Structural* we can have 3 answers: Yes, No, and empty cell. Yes answers indicate that the value is significantly higher for Semantic (or Structural) players than for the population. Conversely, No answers state that the value is significantly lower that for the population. Empty cells indicate non-significant differences. For the last column shows the answer to the question: "Which among semantic and structural influencers has significantly higher values for that behavior?" Also in this case, empty cells indicate non-significant differences.

Behaviors	**Semantic**	**Structural**	**Semantic vs Structural**
timeplayed	Yes	No	Semantic
weighted_time	Yes	No	Semantic
ratio_time_popular	-	No	Semantic
avg_game_popular	-	No	Semantic
ratio_Action	Yes	No	Semantic
ratio_Adventure	No	Yes	Structural
ratio_Casual	No	Yes	Structural
ratio_FreetoPlay	Yes	-	Semantic
ratio_Indie	-	Yes	Structural
ratio_MM	Yes	-	-
ratio_RPG	Yes	Yes	Structural
ratio_Racing	-	Yes	Structural
ratio_Simulation	-	Yes	Structural
ratio_Sports	-	Yes	Structural
ratio_Strategy	Yes	Yes	Structural
ratio_EarlyAccess	Yes	-	-
ratio_time_mp	Yes	-	-
retention_transfer	No	Yes	Structural
ratio_neigh_new_games	Yes	Yes	Semantic
avg_retention_ngames	Yes	-	Semantic

and Strategy games than the whole population (pvalue<.001). We also found how semantic influencers spent more time playing multiplayer than the population (U=8106927.0, pvalue=0.03), while there is no statistically significant difference with central players.

5.3 Understanding Social Relationships in Gamification

In the previous sections, we showed how influential individuals impacting other retention exist in specific games and various games. We also observed that semantic and structural approaches produce different sets of users, with the semantic influencers having a stronger impact on others' behaviors (i.e., retention and engagement in new games). In this section, we retraced to our core application domain: gameful systems. Hence, we replicated and extended those influence analyses by studying data from Play&Go [142]. This work aims to first verify the existence of retention influencers in our gamified application, with the consequence of bridging games and gamification research for what concerns social influence. Second, we investigated the existence of different types of influence, other than retention, allowed by the semantic algorithm. Specifically, we analyzed the presence of influencers promoting ecological mobility habits in line with the gamification goal.

In the following section, we present the network structure and the dataset, which is constrained to a subset of the whole population of Play&Go players (see Chapter 4). Then, we used the semantic algorithm to find, first, retention influencers and, second, mobility influencers.

Constructs and Metrics

Play&Go, as described in Chapter 3, promotes sustainable transportation habits and allows players to track their travels and obtain points in a proportional amount to the eco-sustainability of the journey.

For this study, we analyzed the last edition of the gamification campaign (2019-2020). Of the registered users, 119 players participated in multiplayer challenges, representing the only mean of in-game interaction. On average, those players were active for 11 game weeks (SD = 5.2) over the 17 analyzed. They were fairly active, with an average of 322 game actions (SD = 336.7) up to a maximum of 844 actions.

In-game interactions are logged and have been processed to measure the following behaviors: (1) *participation*, (2) *mobility habits*, (3) *green mobility*, (4) *public transportation*.

Participation describes how active players are in the game, using as features the number of game actions and the number of points collected. The number of game actions mostly consists of the number of trips tracked. However, other actions can also be performed, such as inviting other users in a challenge, unlocking a challenge mode, or blocking a player. While game actions convey information on how frequently the game is used, the number of points awarded communicates how intensive and in line with the gamification goal the usage was. Higher scores meant longer and greener travels. *Mobility habits* represents players' transportation behaviors in the form of km and trips tracked, divided for the means available (walk, bike, train, and bus). This view on players' activity is more descriptive than the participation data and is fully representative of the application domain: mobility. From this concept we refined two finer-grained behaviors: *green mobility* and *public transportation*. *Green mobility* describes players' activity concerning green trips, tracked by bike, and walking. *Public*

TABLE 5.6: Multiplayer challenges

	#challenges	**#players**	**Density**
All	495	119	0.05
Cooperative	397	111	0.05
Competitive	98	96	0.02
Injected	239	111	0.04
Voluntary	256	82	0.04

transportation, on the other hand, model users' km and trips tracked by bus and train. Prior processing data has been normalized in the range $[0; 1]$.

Network Structure

In Play&Go, players can connect through multiplayer challenges, which can be cooperative or competitive. Multiplayer challenges can also be grouped by assignment policy: *voluntary* and *injected*. *Voluntary* multiplayer challenges are created by players using the invitation system. Hence, a user invited another player to a challenge they defied, and the other player accepted the invitation. *Injected* challenges, on the other hand, are automatically assigned by the system that matches players using a naive algorithm. Players not already involved in voluntary multiplayer challenges are matched with others with a similar level of expertise. The challenge is generated to better match both their performance and habits. Multiplayer challenges can be translated into a graph. In each graph, the edges are undirected and weighted. The weight is a natural number and represents the number of occurred interactions.

Among the 495 multiplayer matches, 397 were cooperative challenges and 98 competitive. The 495 multiplayer challenges can also be grouped into 256 voluntary and 239 injected matches (Table 5.6). Each subset of multiplayer matches is a specific view of players in-game interactions. As a result, we modeled and analyzed the following five graphs.

- *Complete.* A link exists between two nodes if they participated in a challenge together, regardless of the type.
- *Cooperative.* A link exists between two nodes if they participated in a cooperative challenge together.
- *Competitive.* A link exists between two nodes if they participated in a competitive challenge together.
- *Injected.* A link exists between two nodes if they participated in a challenge together, which was automatically assigned.
- *Voluntary.* A link exists between two nodes if they participated in a challenge together, initiated by one of the two players through an in-game invite.

Each graph is dynamic; and thus, consists of a sequence of snapshots taken at regular intervals. The intervals are weeks, as challenges have a weekly validity. The graph's dynamicity affects both edges, for which we have temporal information and nodes' properties. Specifically, we conducted four rounds of analyses, one for each behavior presented. For instance, we first measured retention influencers, where nodes are characterized by the *participation* behavior. Retention influencers were researched in each of the five graph models listed. Then, we analyzed other types of influences by characterizing the nodes with another behavior, e.g., *mobility habits*.

Analyses and Results

In Play&Go, we analyzed, first, retention influencers by replicating and extending our previous works on *Destiny* [144] and *Steam* [137]. Then, we researched other types of influencers impacting in-game behaviors describing players' mobility habits, which is the gamified app's application domain. We also compared how influencers vary with the type of influence computed.

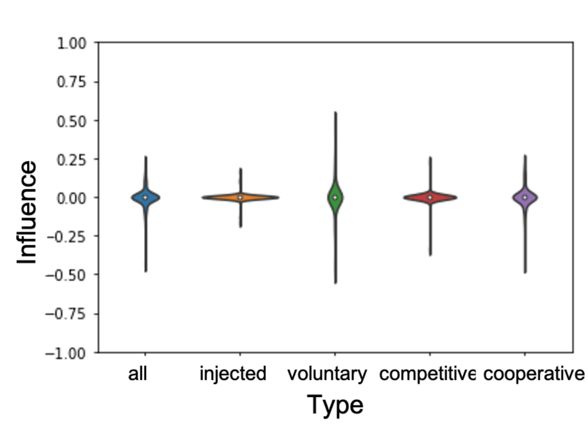


FIGURE 5.9: The violin plots represent the distribution of the participation influence scores across the five graph models: all challenges, injected challenges, voluntary challenges, cooperative challenges, and competitive challenges. The height of the violin plot represents the range of the influence scores for each graph model. The thicker part means the values in that region have a higher frequency. Vice-versa the thinner part represents lower frequency

Influence in Retention

In our previous works, presented in Section 5.2.1 and Section **??**, we argued for the existence of game influencers impacting others' long-term retention, identified through the Semantic Influence Algorithm defined in Section 5.1. We built upon that research by investigating retention influencers in our gameful application Play&Go, using the same algorithm. The Semantic Influence Algorithm, besides a (possibly dynamic) graph representation of players' social interactions, requires the nodes are characterized by a set of features describing the behavior influenced. To detect retention influencers, we analyzed players' participation behaviors. In Play&Go, we modeled participation in terms of the number of game actions and the number of points obtained. The algorithm was executed on the five graph models – e.g., all challenges, voluntary challenges, injected challenges, cooperative challenges, and competitive challenges.

Figure 5.9 shows the distribution of the influence scores across the graph representations. The strength of influence varies with the network considered, despite in each model the distribution of the values peaked at zero. The network deriving from voluntary challenges was the greatest incubator of influence, which may be interpreted as challenges formed on a voluntary base were more impactful than injected ones. The type of social mechanic, on the other hand, showed less prominent differences. The magnitude of exerted influence lies in similar intervals. Yet, the competitive network hosted more nodes neutral to influence. Then, we compared players' influence scores across the different networks to understand whether challenge types and assignment policies affected the magnitude of social contagion. Specifically, we correlated their influence scores across the models using Kendall τ [72]. The correlation among the scores in the cooperative graph and the competitive graph was very small (.12) although statistically significant ($p < 0.05$). On the other hand, the scores in the injected and voluntary graphs were unrelated.

Players' influence scores were independent of their in-game experience, described by level number and position in the leaderboard. Coherently to our prior findings, semantic influencers showed lower values in standard centrality metrics (degree centrality, closeness centrality, and betweenness centrality).

Semantic influencers impact on long-term retention was evaluated using the `retention transfer` metric, as defined in [144] (see Section 5.2.1). The smaller the retention transfer value, the higher the influencers' impact on their neighbors' retention. Semantic influencers extracted from the complete graph showed values significantly smaller than the remainder of the population (Mann Whitney $U = 451, p < .05$), very close to zero. By analyzing the other graph representation, we found similarly impactful influencers in the cooperative and voluntary networks, conversely to influencers from the injected and competitive networks. This result confirms that the network model can affect the presence and strength of influence exerted.

Other Types of Influence

Following the study on retention influencers, we investigated the existence of individuals influencing others' mobility habits, as Play&Go has the goal of promoting sustainable transportation behaviors. Specifically, we analyzed tracked kilometers and trips for each *transportation mean* available. Then, we separately measured influence on mobility behaviors using *green transportation means* and *public transportation*. Consequently, we computed three types of influence, whose distribution is depicted in Figure 5.13.

Voluntary matches were, again, the stronger conductor of social influence, regardless of the influence type. On the other hand, the influence exerted in the cooperative and competitive networks was of similar magnitude. We also observed how players' green transportation behaviors were more influenced than general mobility habits, including public transportation. Consistent with previous findings, we found a lack of correlation between the influence values in the different graph representations. Hence, the nature of the social interaction affects the existence and power of influencers across the network.

Then, we computed the `retention transfer` metric to verify whether those influencers also impact long-term permanence in the game as retention influencers. Mobility and public transportation influencers did not differ from the remaining of the population concerning the impact of retention. However, results show that green mobility influencers affected others' permanence in the game as retention influencers. Moreover, influencers in the cooperative network obtained even better values than the other networks (e.g., injected, voluntary, and competitive).

In all of our studies (i.e., [144, 137, 142]), in-game retention influencers were ordinary players residing in the periphery of the PSN. Mobility influencers match the same pattern: they are not particularly experienced users, top players, or central nodes.

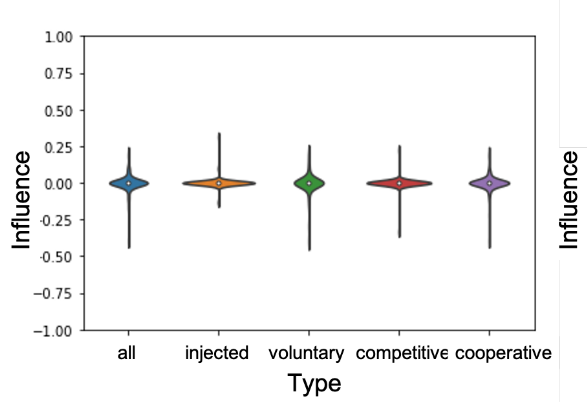


FIGURE 5.10: Transport Means

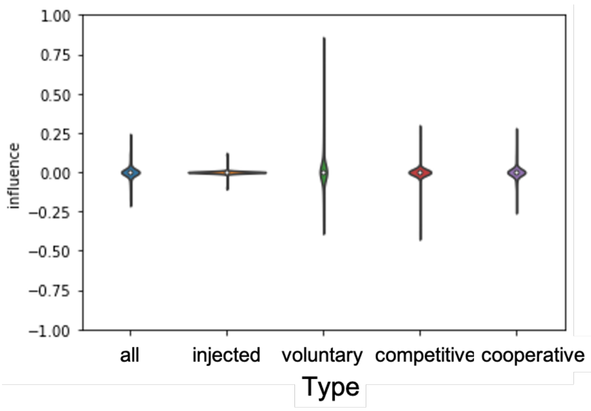


FIGURE 5.11: Active Transport Means

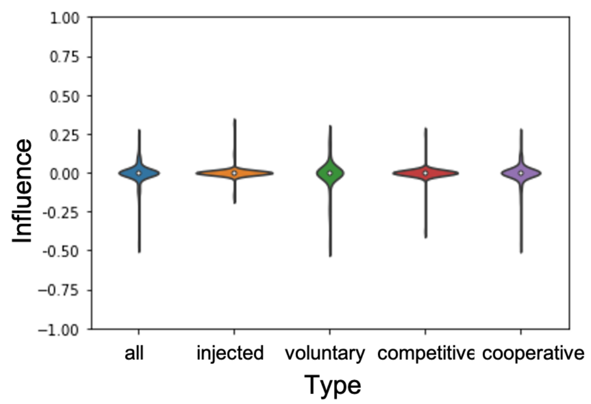


FIGURE 5.12: Passive Transport Means

FIGURE 5.13: Violinplots of the influence scores for each graph representation (all challenges, injected challenges, voluntary challenges, competitive challenges, and cooperative challenges). The different plots refer to different influence types (i.e., nodes' properties/features used): influence computed on all transportation means, green transportation means (i.e., walk and bike), and public transportation means.

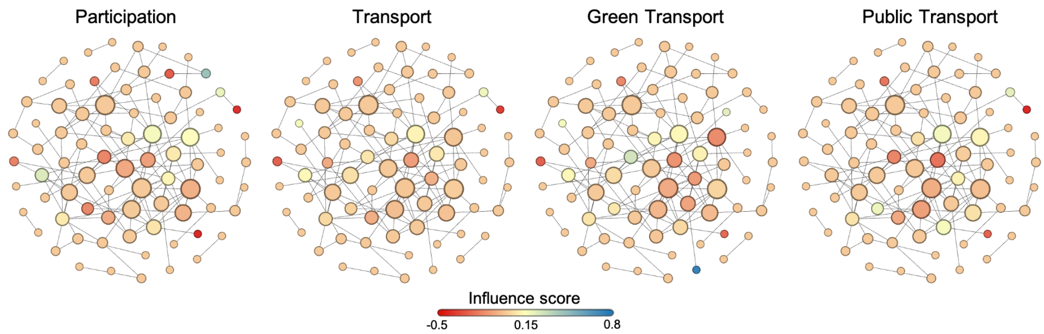


FIGURE 5.14: Graphs of the voluntary challenges. Each graph presents the same topology, where the nodes are the players and the edges represent a shared match. The bigger the node size, the higher the node's degree centrality. The graphs differ on the nodes' colors, representing the influence scores in participation, transport, green transport, and public transport influence, respectively. Please note that despite the influence laying in the range $[-1;1]$, the legend ranges between -0.5 and 0.8. Those values are the local minima and maxima across our experiments. We used this reduced range to make the colors more evident.

Finally, we visualized and compared the influence scores players' obtained in the four influence types analyzed: participation, transport, green transport, and public transport influence. Figure 5.14 shows the voluntary graph, in which the exerted influence was the strongest. Across the graphs, the position of the nodes was kept consistent to ease the comparison. The colors of the nodes in the picture evidently show how to influence, and semantic influencers, chance with the type of influence analyzed. Hence, being an influencer is not an absolute status.

5.4 Summary

People's behaviors are affected by how the network around them acts. The literature on players' social networks proves as the structure of gamer communities—either built around games [226] or inside the gameplay [37]—impacts the players' activities in the game. Some individuals, in particular, have a say in how their connections behave by exerting influence on them. Those *game influencers*, a term borrowed from social media, were initially identified through centrality metrics [37], as commonly done in other platforms. Hence, they were intended as well-connected, visible players strategically placed in the network. As the SNA literature shows a double interpretation of influencers, who are either described to possess desirable structural properties (centrality in the network) or semantic characteristics (measurable behavioral change), we researched whether the latter overlapped with the former. Our preliminary study on the *Destiny* Crucible matches showed how central (structural) influencers greatly differ from semantic influencers, allowing a more in-depth understanding of the player community. First, we found that semantic (behavioral) influence can be strengthened over time when players engage in multiple matches together. Semantic influencers, in fact, were involved in few but stronger connections, conversely to central players. Second, our results highlighted how the act of exerting influence also varies with the person receiving that influence, and thus, being an influencer is not an absolute status. Finally, we observed that semantic influencers had a stronger impact on players' retention than structural (central) influencers.

To expand further our knowledge of the players' network, we investigated how players connect on a game distribution platform, Steam, which abstracts the concept of specific games. We found how social contagion can also occur in a cross-game environment, supporting that social relationships can be transferred from one game to another [44]. Steam influencers, as game influencers, impact others' retention within the game provider but also influenced others' choice to play a new game. Additionally, the analysis of the Steam network highlighted properties and behaviors of influencers distinguishing them from the *ordinary* player, undetectable from telemetry data currently analyzed in prior works [37]. Therefore, influencers have proven to fully impact the community as they condition future trends and help maintain the network active and connected.

Finally, we studied the existence of social influence in gamification (*Play&Go*) and whether influencers impacted other players' retention and in-game activity. Our findings show how influencers do indeed exist and affect players' long-term permanence and help achieve the gamification goal. Besides, as in games, influential users are neither elite nor central players regardless of the type of influence analyzed. In conclusion, social influence can also occur in gamification with social rules similar to what emerged in entertainment games. Moreover, influence can be computed on different behaviors, and influencers vary with the influence type computed.

Chapter 6

Adaptive Gamification Framework

Players' individual preferences and the game's ability to accommodate them can determine the goodness of their gaming experience, as the game is perceived differently by each user. Hence, their participation, retention, and engagement can be impacted by whether the game elements and mechanics meet their tastes [219].

Player modeling has been tackled since Bartle's work [18] and many theoretical frameworks and taxonomies have been proposed (see Chapter 2. In a parallel and related research area, in-game behaviors have been analyzed to retrieve a data-driven representation of players instead of relying solely on self-assessments. Nevertheless, a systematic method to translate the knowledge extracted from player in-game behaviors into fully tailored game experiences is still lacking. Moreover, player profiling approaches generally focus on the individuals' identity, neglecting the social context in which they live. Information on their role and impact in the player network is never included. Finally, player modeling taxonomies, regardless of whether they are theoretical or data-driven, associate a static profile with the user. Researchers, on the other hand, found the player gaming experience to be ever-changing. Hence, iterative adaptation strategies should be preferred.

Although most research on personalization and customization has been conducted on entertainment games, gamification can highly benefit from being adaptive. The ulterior motive [96, 228] characterizing a gamified application is, by default, something that would be hard to pursue otherwise. Game elements are exploited to transform the activity into something fun and entertaining. However, engagement is very subjective and depends on players' preferences and personalities. As a result, gamification is not granted to have successful outcomes. Instead, many gamification examples can have a neutral (or even negative) effect on players [121]. Iterative design, informed by games analytics, can help to reduce this variability and ensure more successful gamified applications by allowing the system's supervision and modification as it progresses [202]. The same iterative approach should be exploited for player-centric adaptation to produce ad-hoc content for each user.

In the previous chapters, we analyzed telemetry data to study players' experiences from various angles. First, we analyzed players as individuals. Modeling their in-game behaviors was used to predict the likelihood of churn, understand whether the gamification goal was pursued, and learn and model their preferences. Then, we investigated in-game social relationships to identify drivers of positive behaviors - i.e., influencers. Most importantly, we showed how in-game activity could be used

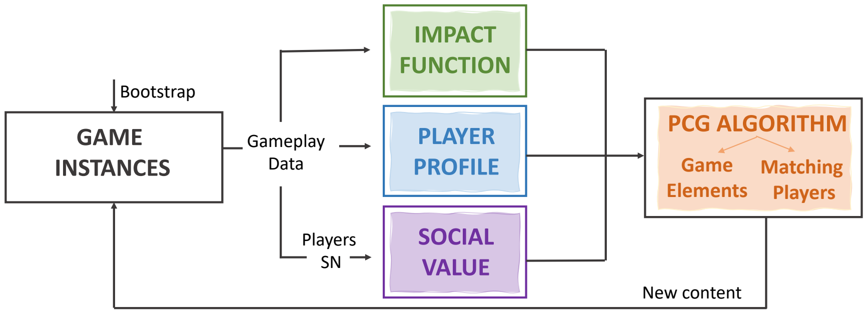


FIGURE 6.1: Framework's modules and flow.

as feedback to interpret players' responses to specific customization of the game elements with which they interact. This chapter connects those findings in a conceptual framework for adaptive gamification, described in detail in the following section.

6.1 The Conceptual Framework

We propose a data-driven conceptual framework for adaptive gamification, designed to extract information on players from telemetry data and learn to produce tailored content for each player. The framework (Figure6.1) is built of four modules: `Impact Function`, `Player Profile`, `Social Value`, and `PCG Algorithm`. The first three modules (*Extraction Modules*) take as input the gameplay data collected from the *Game Instances*. Each game instance describes the state of the game for a user. The gameplay data, collected from players' interactions with their game instance, provides each extraction module information on a specific aspect of the players' experiences. The last module (PCG Algorithm), or *Processing Module*, is fed with the outcomes of the previous modules and generates game content for each player accordingly.

The core characteristics of the framework are modular and iterative. First, modularity fosters generalizability. Each module is treated as a black box: assuming the inputs and the outputs are respondent to the definition, the implementation can vary. Moreover, the three *Extraction Modules* are optional, as long as at least one is included. Second, the iterative nature of the framework has the primary purpose of monitoring and accommodating the dynamicity of players' gaming experiences. Hence, the framework can work in real-time. The designer or analyst can define the time window length that is better suited for the use case. The framework then extracts players' data at regular intervals, processes their behaviors, and makes decisions. A more detailed description of the modules follows.

Impact Function is the module responsible for understanding how the game experience is perceived by the player - i.e., the impact the game has on them. The module elaborates data describing players' activity in the game. It produces a measure of engagement, which will be used by the *PCG Algorithm* to decide to maintain or change the current adaptation strategy for that player. In other words, the metric assesses the goodness of the decision and helps to adjust it in the following iterations. It should also be noted that this module is also in charge of signaling behavioral shifts. In case players modify their preferences or behaviors as the game advances, the impact function value will reflect this change, processed in the procedural content generation phase. As we anticipated, each module is a black box. The impact function module can also be adapted to the application domain. The default function would be to monitor players' activity and participation, indicators of the likelihood of churn [143, 139]. However, analysts may prioritize other aspects. For instance, in the serious games and gameful systems domains, they might choose to measure whether the "ulterior motive" in play is pursued (e.g., in a Persuasive Game for Health, the measure of interest would be a health-based indicator).

Player Profile is the module responsible for modeling players' behaviors from gameplay data. Hence, it takes as input players' in-game interactions and outputs player profiles. The profiles are constantly updated throughout the gameplay with real-time data. Designers and analysts are in charge of the definition of the in-game actions and the profiling model. Although the module is defined as iterative and data-driven, practitioners can decide to use static profiles as the framework itself. If they rely on questionnaire-based approaches, for example, they can replace the module with the surveys' results. The model's output will still be a player characterization fed to the *PCG Algorithm*, which remains unchanged throughout the game experience, regardless of behavioral changes.

Social Value is the module responsible for modeling in-game social relationships. This component is relevant only for gamified applications that include multiplayer elements, to some extent, such as cooperative or competitive tasks. The module takes as input players' social interactions to build a player social network. The output describes players' social value, which is a numeric representation of their role in the community. For instance, this module can be used to research influencers and influenced players after having defined the behavior of interest ([144, 142, 137]). Then, players' social value is communicated to the *PCG Algorithm* to inform player matching strategies.

PCG Algorithm is the module responsible for processing players' information to produce tailored content. The component is fed with information on players' experiences (*Impact Function*), on players' preferences and characteristics (*Player Profile*), and on the player social network (*Social Value*). In practice, the procedural content generation algorithm deploys personalized game mechanics using the player profiles and matches players according to their profiles and social values. Iteration after iteration, the generation strategy is adjusted according to the impact function value in a reinforcement-learning-fashion. In other terms, the effectiveness of the algorithm's choice is evaluated iteratively, using players' in-game activity, and the adaptation strategy is modified accordingly.

6.2 A Feasibility Study – The Play&Go Case

In the following section, we provide a concrete implementation of the conceptual adaptation framework based on our persuasive gamified application Play&Go. The purpose of this exercise is to ground our assumptions on the benefit of the framework in previous studies. Chapter 4 and Chapter 5 presented analyses on players' experiences, both individual and social. We elaborate those findings to detail the modules of the framework and show a possible implementation. As we will later discuss, in future works, the framework's implementation will be empirically validated.

Play&Go (described in Chapter 3) allows players to interact with the system in a multitude of ways. Not only can players track their movements, which is an action highly tied to the application domain, but also engage with ad-hoc challenges, the customization mechanism, and (indirectly) interact with other users. The sponsored gamification campaigns are organized in weeks: partial leaderboards dictate the winners of weekly physical prizes. Therefore, the framework will analyze the gameplay at intervals of one week - i.e., weekly timeframes.

In this example, we assume that the designers' goal is to retain players in the long term. Therefore, the **Impact Function** module will exploit players' participation data, measured by their intensity and frequency of usage. Intensity is computed as the number of game actions performed, whereas frequency describes the number of active days in the game week. Our previous findings informed the choice of those two metrics or behaviors. First, participation data is a predictor of player churn [143, 139] and a low level of activity is a warning for a possible abandonment of the game [139]. Since the purpose is to retain players for a long time, the framework will aim to maximize the values for the impact function - i.e., players' in-game participation. Second, players' level of activity can be used in reinforcement-learning-like algorithms to learn player preferences concerning a particular game element [141]. Specifically, participation data can be used, in Play&Go, to order the different types of challenges according to the users' tastes.

The **Player Profile** module can either be built from the results of a survey-based investigation - e.g., the Hexad User Types [219] - or from data-driven players' behaviors. Although in-game choices [138] and level of activity [141] are strong indicators of players' preferences, there is no evidence of the superiority of data-driven over theoretically-informed adaptation strategies [136]. However, our findings showed that choosing one approach rather than the other significantly affects the type of content generated [136]. Therefore, we opted for a hybrid solution. Player profiles are described by combining their scores in the Hexad User Types Survey [219] and their in-game behaviors. Players' in-game behaviors are defined by the designers or domain experts and describe specific interactions with the gameful environment. Examples of in-game behaviors, in Play&Go, are `Reactivity` measuring how fast players were in customizing game elements, `Sociality` evaluating the number of social interactions they are involved in, and `Self-improvement` computing whether they tended to better their performance [136] (for more details refer to Chapter 4). Existing methods can be used to merge multiple profiling techniques [206].

Having computed the game's impact on players and their profile, we then define the **Social Value** module. Although the conceptual framework was implemented to maximize players' permanence in the game, the application domain remains a persuasive gamified system pursuing a concrete goal. Towards this, this module is instantiated to identify influencers in the network of players propagating green mobility behaviors among their peers [142]. Hence, players' social value is dictated by their nature of influencer or being susceptible (or neutral) to influence.

Finally, the **PCG Algorithm** module collects and elaborates the outputs from the previous modules and procedurally generates ad-hoc content. In Play&Go, the customizable elements are weekly challenges. Challenges vary in difficulty level, computed as the player's expected effort to win it, and in the type of target (e.g., biking, walking, or points, in general). Challenges can also be single-player or among two players (i.e., multiplayer).
In the first step, the module computes the type of game elements to deploy for each user. The algorithm uses as a baseline the profiles produced by the *Player Profile* module to define the initial adaptation strategy. The baseline is integrated with information on players' experience, from the *Impact Function* module, and fed to the

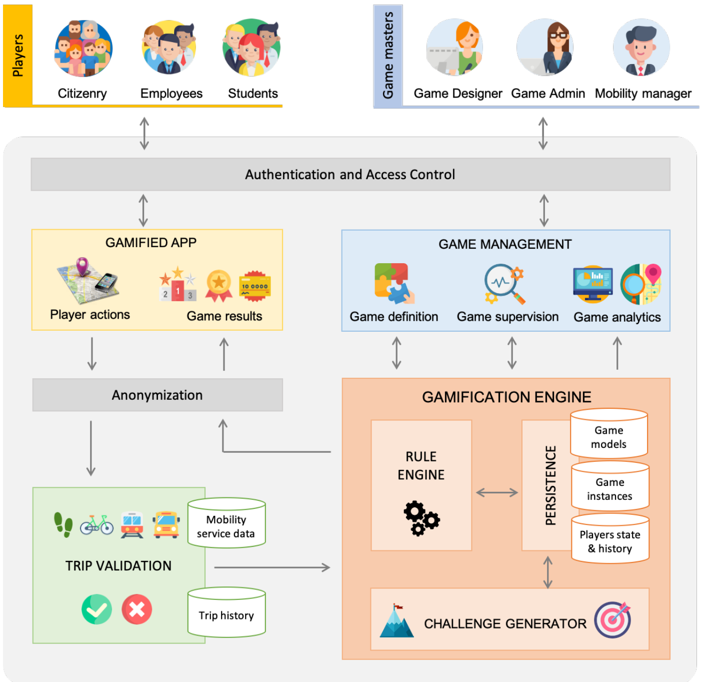


FIGURE 6.2: Conceptual Architecture of the Gamification Platform.

online version of the Reinforcement Learning Algorithm defined in [141]. The algorithm takes as input players' implicit feedback (i.e., the impact values) and learns how to order the challenge types according to the players' tastes. Iteration after iteration, the learning becomes more accurate. Moreover, the baseline, updated with the current players' behaviors, is used to ease the processing of eventual preferences (or behavioral) shifts. The second step of the module is applied to players preferring multiplayer challenges. The algorithm uses players' social value, describing their role in the community and the effect they have (or receive) on others. The value is exploited to formulate a strategy to match players in order to maximize the spreading of green mobility habits.

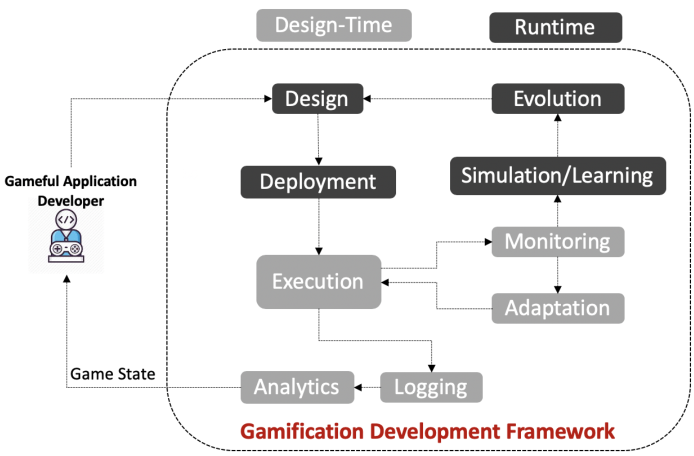


FIGURE 6.3: The Gamification Development Framework.

6.3 A Software Engineering Perspective: The Gamification Development Framework

In the following section, we discuss how the conceptual adaptation framework can be easily integrated into gameful systems' development life cycle from a software engineering perspective. Figure 6.3 shows the *Gamification Development Framework* (GDF) [34], designed to assist the *Gameful Application Developer* throughout all the phases of the design, development, and management of gamified applications. In the previous sections of the chapter, we presented the high-level conceptualization of the adaptation framework, agnostic to the concrete implementation. In this section, we tackle the problem from an implementation perspective, and thus we analyze the models from a different abstraction angle.

The GDF is grounded on the Gamification Platform (Figure 6.2), designed for the definition of mobility gamification campaigns. The Gamification Platform interconnects all the interested stakeholders, such as the *mobility managers and decision-makers*, promoting sustainable mobility policies; the *game designers*, defining the game dynamics, and the *citizens*, participating in the gameful experience. The platform is organized in four main components: the **Gamification Engine**, the **Game Management**, the **Gamified App**, and the **Trip Validation**. The *Gamification Engine* is an open-source component that supports the basic functionalities related to the design, deployment, and execution of gameful systems. The *Game Management* component supports the definition, supervision, and monitoring of on-going gamification campaigns. The information retrieved from observing the game execution can be analyzed to provide analysts data on the game impact, both for the mobility aspect and the players' engagement. The *Gamified App* is the access point for the end-users - i.e., citizens. Finally, the *Trip Validation* component allows comparing the declared transportation mode to the user's "actual" mode, based on traces of the user's position and activity sampled during their trip.

Gamification design is a complex task, especially when the goal is delivering adaptable content at runtime. To ease the definition of the gameful experience, the Gamification Development Framework [34] provides a modular approach for designing gameful systems (Figure 6.3). The modularity accommodates the gamification life cycle providing the designer and domain expert with a modeling language for each component. The framework foresees a `Logging` module in which the in-game actions to track are defined and then collected during the game execution phase. This data is fed to the `Monitoring` component that, together with the `Adaptation` and `Execution` components models the *Adaptive Gamification Framework* [135]. The `Monitoring` component is in charge of collecting and modeling players' in-game interactions. This analysis allows the definition of players' profiles and their social interactions, if any. The `Adaptation` component aims at tailoring the gaming experience for each user by removing, adding, or modifying specific game elements according to their profiles and interaction patterns. In other terms, the `Adaptation` component maps the *PCG Algorithm* module of the conceptual framework, while the `Monitoring` component is a global representation of the three *Extraction Modules* (i.e., Impact Function, Player Profile, and Social Value). The adaptations are deployed on the `Execution` component, which collects the *Game Instances* and regulates the progress of the game. The cycle is continuously retraced as the game advances, as described in the previous sections. In-game actions and behaviors are analyzed again, adapting the gameplay to each player's growth or fixing eventual miscalculation.

6.4 Summary

Adaptive gamification is a cutting-edge research field, as current one-fits-all design strategies have proved to be disadvantageous. The previous chapters' results supported the thesis by which data-driven approaches can also be exploited in gamification to produce tailored content. We integrated those findings in a modular framework, exploiting existent gameplay datalogs to (i) model players' experience and behaviors and (ii) deliver tailored content based on the players' profiles. The framework's core points are its iterative structure and its modularity. As the framework follows a cyclical flow, it accommodates the ever-changing nature of players' behaviors, whereas the division into modules is the key to its flexibility. Not only different methods and approaches can be used to treat players' data, but they also enable the integration of theory-drivel tools—e.g., questionnaires—in the adaptation process. We then showed how this conceptual framework can be contextualized in our core use case, the gamified app Play&Go, and how each module contributed to the profiling and understanding of players' in-game experience towards gameplay adaptation. Finally, we presented a more technical perspective, which supports the feasibility and the usefulness of the framework also from a developer perspective.

Chapter 7

Discussion

Adaptation, or tailoring, is a core component of interactive experiences [241], either them being games or gameful systems. Players are moved by different motivational affordances [237], are engaged by diverse game elements [103, 215], and assume various roles in the community [195, 37]. It follows that adaptive gameplay is a broad topic, as full adaptivity calls for a complete understanding of players' behaviors and personalities from both an individual and social perspective. For a time, adaptive gamification has been under-explored [84, 198], and tailored content was perceived as a complex task more suitable for full-fledged games. Soon, however, the cruciality of also making gamification adaptive emerged [173, 26, 119]. Even more than *traditional* games, gameful systems had to accommodate players' preferences, as, generally, the purpose was to engage users in actions that they would unlikely perform otherwise [84, 198]. Crowdsourcing, behavioral change, and education are only a few examples of application domains that can greatly benefit from tailored content to make the user more willing to engage in the platform, hopefully in the longer term. The main critique moved to the gamified application is their simplicity [52, 121], as most implementations rely on few game elements hardly covering the vast spectrum of players' preferences. Those limited features may be the solely available resources to make the inference on players' profiles and their role in the community, if existent.

In this thesis, we addressed the problem of adaptive gamification, which we structured in (a) the analysis of players' experiences focusing on their individuality and (b) the study of how they connect with others and may be exposed to social influence.

7.1 Players as Individuals

Gameplay datalogs are the players' biographies. They tell the story of how they interacted with the system, evolved with it, and are affected by its rules. Globally, they narrate the population history of that small virtual world, which the governor—i.e., designer—can examine to monitor its signs of progress and decide when an intervention is needed. Individually, those biographies pose a magnifying glass on the citizens of the world to better their lives and connections but also detect issues that might be irrelevant to the general public. In its small, gamification telemetry data also tells a story, and, in this study, we dissected it to understand how much information it can convey.

Telemetry data provides a global overview on gameplay

In their raw form, usually, datalogs are a collection of activities, reflecting players' in-game actions. Therefore, a choice is presented on whether to analyze the data in its aggregated versus its temporal form (`RQ_I_Monitor`). To elaborate on this question, we built players' archetypes (using Archetypal Analysis) from both aggregated and temporal data. Our comparison highlighted similarities and differences in terms of the information conveyed by the two data analysis approaches. Although we detected an overlap in terms of the low-activity archetypes, the temporal archetypes provided a more detailed view of the most active profiles. For example, we obtained the *Wannabe Competitive Achiever*, proving the interest of few players in the competitive multiplayer challenges, which were entirely ignored by most of the users. From a designer perspective, the aggregated analysis led to the belief that competitive challenges were unappealing for the Play&Go population and supported the hypothesis of players perceiving competition as a risk in their path to dominate the leaderboard, and thus, preferred cooperation. Conversely, the temporal analysis sheds light on a minority of players, willing to compete with others but unable to find a match. Uncovering this design fault allows the designers to conceive better matching systems for multiplayer challenges. However, the aggregated analysis was crucial to identify the identikit of players successful in the game in the long term—i.e., dominating the final leaderboard. Finding those winning strategies helped assess whether the gamification goal was pursued, as identified the rewarded interaction patterns. Hence, the aggregated analysis provided great design feedback, as designers' expectations on players' behaviors cannot be assumed to be met.

While the gamification goal dictates the semantic success of the app, a portion of the population may be less invested in the system. Temporal analysis on Play&Go, for instance, emphasized a problem of engagement by modeling players' behavioral transitions. Specifically, we found a sharp division between low-activity and high-activity archetypes. In other words, users would unlikely shift from low activity to a high activity level. This finding is also a symptom of a design fault: part of the players were never fully engage in the game, probably because their specific preferences were not met. Having assessed an engagement problem, we investigated whether we could timely identify those uninterested players who would eventually abandon the game. When gamification is used as persuasive technology, long-term engagement indulges internalization of the behavior in the own routine. Therefore, being able to detect churners promptly could help in preventing such abandonment by actuating contingency strategies. Churn prediction in gameful systems is challenging due to the limited amount of available data (`RQ_I_Churn`). Thus, we build our study under the constraint of using solely participation data, which is inherently produced once the players interact with the system. Constraining the data to participation behaviors may limit the models' performance as a complete view of players' experience is lacking. For instance, in more complex systems, other behaviors can be elaborated describing players' flow, interactions, and preferences. The reasons leading players to churn are multifaceted. In churn prediction analyses performed on games, we witness a very high performance, but also many context-dependent variables modeling players' in-game interactions [147, 76]. On the other hand, in many gameful systems, a reduced set of behaviors can be tracked, as the types of interactions allowed are fewer. Thus, studying whether participation data holds enough information to predict churn can help to monitor the progress of the gaming experience, also in those systems that only implement the *blueprint triad* [121]. In practice, we studied how to process player participation behaviors to improve the prediction accuracy, as well as compared the interpretation and performance of the model in an unrelated F2P game (TagPro). We found that participation data was sufficient to make a reasonable prediction of churn in Play&Go. In addition, interpreting the prediction models provided interesting insights. Players' time investment in the game (gameplay length) and constancy in the activity (frequency of usage) was an important predictor of churn. In other words, the longer the players stay in the game, and the more frequently they play, the less likely it is for them to abandon the game. Conversely, players' activity, computed in terms of the number of points and actions, was not always reliable, and its importance varied across the models built (Play&Go vs. TagPro). This may be due to the policy for point assignment being extremely context-dependent, and thus, difficult to compare.

Despite the multiple shades of knowledge, all those analyses showed how insightful gamification telemetry data could be. Hence, investigating players' datalogs can highly ease and assist the designers' and developers' jobs. While a large-scale, global overview is essential to understand the general trend of the gameful system, modeling players' experiences on an individual level can help tackle issues at a low (player) level. Consequently, delivering customized, tailored content contributes to raising enjoyment for each user, leading to the sought large consensus.

Telemetry data provides player-specific information

Player profiling, informed from telemetry data, is commonly done in GUR. Prior research on games and our findings supported the value of gamification gameplay datalogs. Consequently, we inquired whether player participation data can also be used to learn their preferences for a specific game element (`RQ_I_Learn`). We conducted a feasibility study by investigating the usage of participation behaviors to evaluate player preferences regarding the Play&Go weekly challenges. Towards this, we developed an algorithm following the reinforcement learning paradigm and used explicit in-game choices as ground truth. The results showed that for a portion of the users, the inference was very accurate, whereas others were poor. When we researched the motivation of this dichotomy, we found that it was more connected to the strength of players' preferences rather than personal properties. This finding emphasizes an aspect inherent in the problem tacked. The goal of the algorithm was to learn players' preferences. Yet, it was unsuccessful where there was a lack of thereof. Therefore, the algorithm succeeded in detecting players' tastes, assumed that they actually preferred some challenge types.

While our algorithm can be used to learn how to customize a game element, a full-fledged profile informing the tailoring mechanism is built of multiple shades, calling for more complex methods. Entirely data-driven profiling methods are still an under-explored topic in gamification, which has been mostly dominated by theoretical approaches [119]. Hence, the debate on whether theoretical or data-driven techniques should be used subsists only if, in practice, they lead to diverse adaptation strategies (`RQ_I_Pref`). Otherwise, we could use both methods interchangeably. Therefore, we aimed to shed light on the complex interactions between idealized (self-reports) and contextualized (telemetry data) preferences to outline the benefits, drawbacks, and biases of adaptation strategies, helping researchers and practitioners make informed, conscientious decisions. Towards this, we compared the Hexad User Types [153, 219] and in-game players' behaviors using as a proxy a state-of-the-art model: the PEAS framework [206]. Before analyzing and comparing idealized and contextualized preferences, we conducted a more in-depth analysis of player types. The Hexad User Type model has found remarkable consensus among researchers and practitioners [79]. Therefore, many empirical works analyzed correlations among the player types scores and several game elements and dynamics. Consequently, in our translation to the PEAS representation, we could have used (a) the model definition or (b) the empirical results. Despite our hypothesis of the models leading to different adaptation strategies, we observed how the Hexad model's definition is very robust. Hence, the discrepancies between the original definition and the literature findings are not impactful in the personalization of challenges in Play&Go. Finally, we also compared the two Hexad translations to the player behavior adaptation strategies. We found that the translation informed by the literature review was slightly more similar than the other to the behavior PEAS representation. Specifically, the models were more similar in the difficulty and competitive axis. Although the difference was minimal, this tendency might highlight how empirical outcomes move closer towards data-driven findings than entirely theoretical approaches.

These results are also in line with the previously performed CCA, where no significant association between Hexad User Types and behaviors was found. Nevertheless, the outcomes hint at a potential small to medium-sized effect. The debate on whether using self-reports or telemetry data was still built upon the assumption that they led to different adaptation strategies, without a systematic comparison. Hence, we moved a step towards "reconciling theory-driven and data-driven models" in gamification [241]. The comparison confirmed that those two approaches differ, and provides details on when it is more likely to happen. Specifically, we found similarities in the control and purpose-oriented adaptation axes, whereas the data-driven and self-report-driven models varied in the difficulty, sociality, and competitive axes. In the case of control and purpose orientation, the two Hexad translations (definition and empirical results) were similar, hinting that those aspects may be easier to model than the others. Difficulty and sociality, on the other hand, are more nuanced. The desire to challenge oneself and improve may vary with the importance that the user gives to the task - i.e., it should be contextualized. Sociality can be perceived in very different ways. Multiplayer challenges may be enough for somebody, while others need a stronger social presence. Contextualization is again important as the will to socialize may vary with the specific domain. For the competition aspect, we can make a similar argument.

In conclusion, while our study supports the validity of the Hexad User Types taxonomy, it also highlights discrepancies with actual in-game behaviors. Hence, especially for some adaptation aspects, the context may modify players' preferences. In those cases (i.e., when difficulty and sociality are analyzed), data-driven models are likely to be more accurate, besides better lending themselves to dynamic tailoring.

7.2 Players' as Part of the Community

Players' profiles, however, should also account for the community they are part of. Well-designed social game mechanics can foster a sense of connectedness and belonging [117, 181], which often are feeling prioritized over the will to win [9]. Feeling part of a community of players does not necessarily mean a preference for multiplayer mechanics, but instead being—directly or indirectly—affected by it. Sociality has also been discussed in gamification, where cooperative and competitive dynamics have proven to be beneficial towards pursuing the gamification goal [83, 48, 231]. Players are positively stimulated by the presence and interactions of other individuals, with whom they form an (implicit) social network.

Social contagion and peer influence exist in games

From the SNA literature, we know that online social networks host key users whose opinions have a substantial impact on the digital market [124], namely influencers. Social media influencers are generally identified as users assuming a privileged position in the network: central nodes [74, 14], hence having a broad range of action. Not only social media (i.e., Twitter) influencers exist in the context of games [225], but also game guilds host figures (moderators) positively impacting the activity and performance of the group [195]. Influencers, assuming a central position in the player network, have also been found in the analysis of telemetry data [37]. Players who connected with those central influencers were more likely to stay in the game for longer. However, their favored position increases the likelihood of being connected to more players, including those retained in the game. Therefore, we extended the knowledge on game influencers by developing an algorithm grounded in previous works in social network analysis [197], adopting the definition of influence as an increase in similarity over time (`RQ_S_Inf`). Then, we studied how those semantic influencers encouraged (or hindered) others' retention and how they differ from central (or structural) influencers. The advantage of analyzing semantic influence over centrality metrics is in avoiding the assumption that only one type of influencer exists. Rather, it allows defining the behaviors (features) upon which the influence can be exerted. In this context, we analyzed players' participation in terms of in-game activity. To evaluate the impact on retention, we defined a custom metric, *retention transfer*, which also account for negative influence —i.e., a premature churn of an influencer may have repercussions on the retention of users influenced by them. Our results show a dichotomy between structural (central) influencers and semantic influencers, suggesting that centrality does not imply influence within games. Instead, semantic influencers were the players impacting others' retention. They were involved in few strong connections, persistent over time, suggesting that influence can be reinforced. Central players also exerted influence to some extend. However, they conditioned the behaviors of only a part of their connections, hinting at influence not being a status but rather a condition requiring both influencers and an individual susceptible to their influence. Those findings partially agree with previous research on influencers in games by Canossa et al. [37], in that social influence is exerted through implicit gameplay and affects long-term retention. However, this also highlights how different types of influencers may exist, and their properties may vary according to the context they are placed in, hence the importance of semantic approaches.

Social contagion and peer influence exist across games

Game influencers, detected from the analysis of telemetry data, result from a context-specific investigation, limited to the virtual world—i.e., game—under investigation. However, the player community extends across different game titles, and different types and genres can attract players. The Steam social network is peculiar, as it retraces, reinvented, and adapted characteristics of social media translated into games. Nevertheless, Steam greatly differs from social media. As a game distributor, registered players can ignore its social features and only use the system as a personal game library. The context is also very different from a multiplayer game, where interactions with other players are necessary and, to some extent, controlled. Also, unlike specific games, Steam social relationships are explicitly manifested, whereas game interactions are implicit and can be casual. Hence, the semantic behind players' connections skews the platform slightly closer to social media. Therefore, we investigated the presence of influencers in Steam and compared them to gameplay and social media influencers (`RQ_S_Steam`).

Structural (centrality-based) approaches work well in social media. These influencers communicate explicitly with their followers and may condition them in real life. Game influencers, on the other hand, impact in-game behaviors rather than daily actions. Steam is a hybrid that combines explicit and direct social tools and includes high-level information on implicit in-game interactions (e.g., time spent playing and achievements awarded). It also differs from other game-related platforms, such as Twitch, which encourages direct forms of communication. This dichotomy raises the question of whether centrality-based methods should be used to identify steam influencers, as social networks [14, 85], or by semantic approaches, as games [144]. Our outcomes confirm semantic influencers being more impactful on players' retention than central players, connecting low-level social relationships (in games) to high-level social relationships (in Steam). Such a result is not obvious since multiplayer games' social dynamics are entirely different from Steam's social interactions. Besides, they partially collide with previous analyses on the Steam network showing how players connected to users similar to them [170]. The semantic algorithm controlled homophily by modeling influence as a behavioral change: friends' behaviors were different before one of the two players (influenced) emulated the other (influencers). Therefore, the similarity observed in prior studies [170] discards specific (low-level) gaming habits. Moreover, our findings strengthen the concept of games connecting players at a deep level [31, 58, 69], even when social interactions are offered as an optional feature instead of being deeply embedded in the platform. While we are unaware of the importance that players attribute to the SN-based mechanics, making influencers more SN-like (or central), Steam's focus on games might sensitize players to their friends' actions acting as social motivators to play [232]. Our findings support the latter hypothesis. Steam influencers, coherently to games research, are (a) better identified through the semantic approach, and (b) are far from being central (or popular) players. Therefore, regardless of the relationships being implicit or explicit, the platform domain—i.e., games—plays a crucial role in the types and modalities of players' interactions and how they affect one another. This finding also strengthens the conception of games as incubators of intimate, social relationships and retention being positively affected by connectedness. While SN influencers may build a bond with a wide audience through social media, games pose a stricter constraint. The reach is dictated by the number of players with which influencers can physically engage (and play). This reflection opens many further inquiries on who those influencers are, how they differ from other social roles (e.g., team leaders), and whether some of them might contribute to the spreading of toxic behaviors within the community [41].

Steam semantic influencers impact long-term retention, as occurs in games [144]. However, we found that they also promoted new games within their neighbors, unlike ordinary players in the population. Central players also were copied by their neighbors in their game choices. However, players' retention in new games was significantly higher among influential nodes' neighbors than the other players (i.e., central). The ability to disseminate new games among their peers links Steam influencers to social media influencers and their recognized power to spread products in the network [239]. Although this might also support the conception of games as material goods rather than social platforms, there is a subtle difference. Social media influencers, which are central, visible individuals, are commonly identified through structural approaches [14], like our popular (central) players in the Steam network. Yet, while users connected to central players are keen to try new games, players in contact with semantic influencers are more likely to be motivated to also stay in the game for longer. In other words, although central players stimulate other players' curiosity, semantic influencers' effect is stronger and more persistent, reconnecting to the more intimate relationships that semantic influencers can form. Such a finding strengthens the social worth of semantic influencers. We showed how crucial their role in the community is, determining its evolution and behavioral choices (i.e., play a specific game). This phenomenon opens many points for further investigations and discussion. The longer retention in new games can be explained by the presence of the influencer in the game. Following up on the results regarding long-term retention both in specific games and in the Steam players' community, the influential users may compel others to play the new game for as long as they keep playing it. In other terms, it should be researched whether influencers at a community level (Steam) are also influencers at a game level.

Within games, influencers are regular users and not part of an elite [144, 37]. However, Steam influencers showed peculiar properties, which distinguish them from other players. Semantic influencers were particularly active, as they spent many hours playing and invested their time, especially in games with good reviews. They also favored specific genres: they owned more MMO, F2P, and RPG games, often requiring a higher commitment than casual games (which they tend to avoid). We can connect this to the GUR literature, where Destiny influencers were characterized by strong repeated connections [144]. This commitment might be interpreted as a characteristic of influencers, whose constant presence in the community contributes to building a bond with influenced players. This constancy is also a characteristic proper of social media influencers, who are active posters and contributors to the social platform [239]. In contrast, central influencers, despite also being fairly active, showed a preference for casual games, in which social connections are less likely to perpetuate over time. We also found that semantic influencers owned more early access games. This can be interpreted as an influencers' interest to try new games and being updated with the current trends. Nevertheless, their games library is not wholly composed of popular games, rather comprises many titles. Considering that they own many but only play the most popular games, they may be drivers of the global consensus. This suggests that, even if they may own less acclaimed games, they abandon them pretty quickly. Of course, the latter argument is speculative but calls for more analyses on the matter. Hence, researchers should investigate whether (a) they acquired the game despite the bad reviews to find it to be not of their liking, or (b) they obtained the game before the bad reviews, and their abandonment contributed to the games' unpopularity. This may be a reflection for further research: *How do influencers impact the future and success of games?*

Social contagion and peer influence exist in gamification

The research on game influencers conducted on games, and game-related platforms, enabled the investigation on the method and approaches to identify influencers, thanks to the consistent amount of data and population size. However, our core use case remains the gameful system Play&Go, in which players can also connect socially. Building on the belief that the motivational affordances that are drawing people to play in multiplayer video games are also present within gamified environments [96], we researched the presence of influencers within Play&Go (`RQ_S_Gamif`).

Influencers in games attract much interest from the industry because they proved to have power over a game's player social network, specifically regarding the retention of players in the game. Although the motivations are less tied to sales in gamification, influencers still take on an important role within a gameful application's community. They can, for instance, help achieve the gamification goal (e.g., a positive behavioral change) and support other users to maintain their habit changes indefinitely. For positive behavioral change to occur, long-term retention is needed to allow the users to internalize the new behavior. Hence, if those influential players drive long-term retention, the gameful system is more likely to be successful. Our analysis proved the existence of retention influencers in Play&Go, where users can indirectly interact through two-player challenges. Therefore, we showed how influence could be exerted in gameful environments similar to games. Besides—similar to games— influential nodes are neither central or elite players (e.g., veterans or top-players).

Once having assessed the existence of influential users in Play&Go, we studied the effects of the type of social interaction on social influence. Specifically, in Play&Go, the two-player challenges can be cooperative or competitive. From the literature, there are discordant opinions about the impact of cooperation and competition on social play and influence [97, 2, 87]. Researchers have debated how, on the one hand, cooperation builds a more constructive and less frustrating environment that leads to constructive social growth, free of toxicity [87, 16, 97]. Yet, on the other hand, researchers argue that competition results in compelling challenges, leading players to push their limits further [2]. A third observation in the related literature is that players respond differently to cooperation and competition, because of their personality. Hence, cooperation and competition are different social dynamics and may condition the type and strength of the social connection formed between two individuals. In addition to the social dynamics, challenges can vary in their assignment policy: either they are automatically assigned by the game or can be created through a voluntary invitation from players. Hence, we have an artificial (or injected) connection on the one hand or the creation of a natural and consensual link on the other hand. Given the great difference between willingly choosing the opponent and the task itself over an automatic assignment, we can assume that the social influence emerging within the network could have been affected. Consequently—following the analysis of the complete player network—we studied influence in the (a) cooperative, (b) competitive, (c) voluntary, and (d) injected graphs separately. Our results show that influence is stronger (in magnitude) when the interaction is voluntary rather than injected. This can happen either because the current matching algorithm is too simple or because influence occurs only when an interaction is "natural." However, the social dynamic does not greatly impact the strength of the computed influence. We also found that the influence scores computed in the cooperative and competitive graphs were unrelated. Different influencers were found in each model. A similar result was obtained when comparing the influence scores in the injected and voluntary graph. Therefore, a player's social connection type leads to different influence scores.

Researchers argue how influence can potentially be exerted on any behavior [197, 14]. We studied whether, in Play&Go, other behaviors are susceptible to influence—other than in-game participation. As we used a semantic algorithm to identify influencers, we could specify the behaviors we wanted to study. In our analysis of influence in Play&Go, we found that not only participation influence is exerted, but also *context-specific* influence existed, tied to the system ulterior motive. From the literature, we know how gamification is not guaranteed to produce successful behavioral change outcomes [121]. Whereas, the existence of influencers that help to drive a gamification's end goal may help in having more successful examples, assuming that multiplayer elements can be included in a gameful design. In Play&Go, the gamification goal is to promote sustainable transportation habits. Hence, we computed the influence scores by analyzing players' mobility behaviors: This was done using the number of trips and kilometers tracked with the app using the transportation means available to the user (walk, bike, train, and bus). We also distinguished among green (walk and bike), and public transportation means for a total of new tree types of influence scores. Not only did we find that players influencing others' mobility behaviors existed, but also that the type of influence computed lead to uncorrelated influence scores, and thus, different influencers. Nevertheless, despite influencers varying when different behaviors are analyzed, in all models, (1) nodes with higher influence scores were among the least central nodes, and (2) highly influential nodes were a small minority of the network.

Our results hint that player retention and behavioral change might be fostered by encouraging connections with influencers. So, retention influencers exist and are characterized by intense, repeated contacts over time. This connects with the concept of relatedness and how players are keener to be committed [46], retained, and engaged [183] in the game when they perceive a sense of belonging. At the same time, the presence of influencers promoting the gamification goal is in line with sociality benefiting performance [123, 126, 182]. Thus, including social elements in gamification is beneficial. The findings are also in line with previous research on influencers in games [37, 144], far from being central and well-connected users. This similarity connects further gamification and games, supporting the thesis. The motivational drivers are the same both in games and gamification [96]. Like [144], our outcomes show how variable influence scores are, hinting that influence is not a status but rather a condition deriving from the specific connection. Therefore, the priority is to find good influencers-influenced matching rather than attracting influencers.

7.3 Towards Automatic Adaptive Gamification

Dissecting players' gameplay provided an understanding of the information this data can convey and how to exploit its potential better. Telemetry data showed us how the game was progressing, whether the gamification goal was achieved, and identify users at risk of churn. It also helped in detecting players' tastes and how their behaviors connected with their idealized preference. Finally, telemetry data modeled in-game social relationships and provided information on the social dynamics occurring within the player network. Therefore, this project's last step was to combine those fragments in a unique adaptation mechanism (`RQ_Fram`): our conceptual adaptation framework.

In the GUR literature, the deployment of the adaptive gamification context has been particularly prolific in the field of education. Generally, the adaptation models are grounded on theoretical player taxonomies [160], sometimes combined with motivational analysis [78]. Adaptation based on players' profiles can produce conflicting results [78], especially when a dominant player type is assumed to exist [79]. Towards the usage of more objective metrics, gamification analytics represents the set of tools and methods to measure and evaluate gameful systems [89], as machine learning is deemed a promising field in the path towards automatic adaptation [26]. In this regard, a few recommender systems for persuasive gamified systems exist. Yet, either those are strictly focused on a single game elements [112], inefficient in the generation of a complete adaptation strategy, or still partially rely on player profiles and taxonomies [218]. In this thesis, we answer researchers' calls for automatic, dynamic, adaptive methods for tailoring gameful systems, which do not necessarily rely only on player types or profiles [119]. In the literature, also a design framework exist [25]. It is built of four interconnected macro-modules, depicting the adaptation strategy—i.e., adaptation purpose, adaptation criteria, adaptive intervention, and adaptive game elements and mechanics. From this framework, several design paths can be identified, which represent the process practitioners and designers decide to follow to define the adaptation strategy. Conversely to our solution, this adaptation framework includes several design strategies aimed at producing tailored gameful applications. However, the adaptation aspect is still treated at a higher level, whereas we focused on the possibility of adaptation strategies guided by players' telemetry data and behaviors. Additionally, we pose higher importance on the social aspect, with a dedicated module investigating the player social network, lacking in the design framework. Hence, one of our pillars is the social value gameful environment holds, not only functional in satisfying designers' goals but also in building cohesive, positive communities, benefiting players' well-being [105]. Not only we supported our conceptual model by showing how our previous findings, in terms of individual and social in-game behaviors, connect to the framework, but also by showing how that can be integrated and implemented in a gameful system like Play&Go. The adaptation framework is also connected to our study on the reinforcement learning paradigm applied to player profiling in gamification [141], in which we show how players' participation behaviors can be used to learn their preferences. This study can be seen as a rudimentary version of the whole framework, which extends the adaptation on multiple axes. The core intuition, however, remains the same: using players' in-game implicit feedback—at runtime—to deploy and adjust tailored content.

7.4 Implications and Contributions

A better understanding of players' individual and social interaction patterns in gameful environments is fundamental. Therefore, this work has implications for both researchers and practitioners. We extended the research on gamification on a methodological and conceptual level. First, our findings support the existence of similarities among games and gamification, leading to the possibility of using machine learning techniques to predict and infer knowledge about players' experiences. In other words, we showed how gamification data, although limited, can still be of use. Hence, games and gamification also host similar social phenomena—i.e., social influence —which further bridges the two fields and incentives new investigations. Second, we advanced the discussion on players' profiles and behaviors, in gamification, by systematically proving a difference between what players perceive or expect to like and their actual behaviors in the app. This finding strengthens and motivates further research on players' preferences, also showing how they can change across different contexts and application domains. Consequently, this limits the applicability of survey-based profiling. Finally, we proved the importance of sociality in gamification, supporting the argument of gameful apps needing more complexity than the implementation of the traditional blueprint triad [52, 121]. The importance of social elements in gamification is also relevant for designers and practitioners, as social influence can be exploited to ensure the pursuit of the gamification goal and retain players for longer. This thesis also holds other concrete, actionable findings for gamification practitioners. First, we raised awareness in the choice of the analytics approaches (aggregated vs. temporal), showing how different methods can convey specific views on the gameplay. Thus, they can decide whether they are interested in finding global winning strategies through aggregated analysis, detecting design faults using temporal analysis, or being meaningful in the decision to employ both. We also provided a model to predict churn using limited participation data, which can easily be implemented, as well as a function to measure the game impact at runtime. Analysts, for instance, can instantiate the function to monitor players' progress on a daily or weekly basis, rather than waiting for the conclusion of the gamification campaign to evaluate the impact of their gameful app. Finally, we designed a versatile, modular framework, with an example of implementation. Although its complete functioning is yet to be empirically validated, it provides a blueprint for adaptive gamification, accommodating the mutable nature of players' behaviors and allowing diverse player modeling points (e.g., player behaviors, profiles, and social interactions).

7.5 Limitations

This work comes with a few limitations, which are also inputs for our future works. For what concerns the analyses on Play&Go, the main weakness is the sample size, further reduced when specific constraints were needed (e.g., the study of player preferences or social interactions). Nevertheless, it is challenging to reach numbers comparable to the games industry in most gameful applications. Most popular systems are proprietary and not inclined to share their data. The selected gamified application (Play&Go) stands out from other gameful environments because of users' participation being substantial and voluntary.

Additionally, the analyses were conducted on a single gamified app, which prevents a broad generalization beyond Play&Go. Another issue is presented by some analyses still requiring the human-in-the-loop, which prevents complete automation. For instance, the definition of the impact function tuning points or the interpretation of the churn prediction models. Similarly, in the study of contextualized and idealized player preferences, a domain expert was needed to provide the models' translation. While we suggested multiple translations to reduce the possibility of having a bias, this factor should still be considered.

The generalization issue is less crucial in studying social influencers in games, as we analyzed influencers in three different domains (i.e., Play&Go, Steam, and Destiny) and obtained similar results. Although the sampling of the bigger networks (Steam and Destiny) may have affected the analysis, investigating multiple domains attenuated this bias. Nevertheless, we cannot argue that social influence will manifest in this way in every gamified system. Instead, we show that influencers can also exist in gamification, and thus, there is worth investigating their presence and effect. This finding is also an argument against those who believe in the simplistic view of gamification. We show how gameful environments can be incubators of complex and diverse phenomena, also enclosing people's social life.

Other limitations are specific to the approaches used. In the context of the visualization of the churn models, it should be noted that the interpretation is of the built model instead of the data. PDP describes the relationship that features have with the predicted variable. Whereas in the process of learning player preferences, using our RL-like algorithm, we should acknowledge that the ground truth relied on players' in-game choices. Hence, the results may be influenced by the time pressure of making a selection. Yet, filtering out players making incoherent decisions over time attenuates this limitation.

Another conceptual limitation is in the link between games and gameful systems. While we run the analyses on our dataset, the similarities found are not enough to link games and gamification. Rather, they are an incentive to research more analogies in the two application domains.

Finally, the framework proposed is only theoretical and, as such, still requires and empirical validation. This validation is left for future works. However, we lay good premises by defining its modules following in-depth analyses on players' behaviors and interaction patterns, as well as providing an implementation as proof of concept.

Chapter 8

Conclusions

Static gamification tends to replace internal motivation with external motivation [169], which is notably less persistent over time [201]. When players are well-connected to the gamification ulterior motive through intrinsic engagement, both the organization and players gain benefits [169]. Consequently, gamification is embracing the concept of adaptation to support and increase long-term engagement. However, in the race for tailored content, a balance must be found among player-centric adaptation, accounting for users' tastes and preferences, and global tailoring to pursue the gamification goal [26]. The gamification application success is dependent not only on the characteristics of the population but also on the application domain and the specific design [89]. The difficulty in gamification design, especially when matched to adaptive content, lies in the meaningful connection of players' needs to the gamification layer [26]. The research of methods to tailor gamification is still a trend [119].

Researchers call for automatic, dynamic, adaptive methods for tailoring gameful systems, which do not necessarily rely only on player types or profiles [119]. Towards this, we proposed a conceptual adaptation framework for gamification, with the core features of being iterative and modular. The division in modules structure allows defining different profiling approaches, not necessarily connected to survey-based player types. The cyclical structure connects to the need of having a dynamic, reinforced process, avoiding the "novelty effect", which may accompany a static adaptation strategy [26]. Finally, the deployment module for the procedural content generation lends itself to being fully automatic, before further investigations. Although the framework is still conceptual, we run in-depth analyses proving the operation of each gear and showing how those fragments can be integrated into the whole machine. The gameplay's impact on players can be measured at runtime, using churn prediction modules or players' participation behaviors, either in a raw or elaborated form. Players' profiles can be derived from their in-game behaviors, reflecting an eventual behavioral change, or be statically defined through state-of-the-art models. Players' social value can be computed by evaluating the (different types of) influence exerted or perceived. Finally, we supported the applicability of reinforcement-learning-like algorithms to learn and adjust players' adaptation strategies using their implicit (participation) behaviors.

In summary, we proved the feasibility of an automatic, dynamic, modular, cyclical adaptation framework feeding on telemetry data and embedding information on players conceived as both individuals and part of the community.

Future Works

As we mentioned multiple times in this manuscript, there is an extensive opportunity for future works. The obvious first step would be empirically validating a preliminary version of the framework to assess how the pieces work and can be concretely integrated. In this regard, we can choose among plenty of variations, changing the specific implementations of the framework modules. Another high-level, natural next step is the study of our model and findings in a completely different gameful application to provide information on the generalizability of our outcomes. On a lower-level, each of the studies conducted on players' behaviors and social interactions can be extended. For example, the RL-like algorithm to learn players' tastes can be tested in an online setting and adapted to learn multiple preferences, rather than being limited to a single game element. Correlated to player behaviors, we can empirically investigate how different adaptation strategies (e.g., survey-based or data-driven) impact retention end enjoyment. Future studies should also deepen the understanding of gamification influencers and contribute to players' retention and positive behavioral change by informing matchmaking algorithms.

All in all, this thesis is also meant as a stimulus for interesting future research in gamification, where the main takeaway is that the only limits that we have are the ones we build in our minds. Like every interactive experience, gameful applications are a gold mine of information that can be used to better the humans' virtual experiences.

Bibliography

[1] Sumayah Abu-Dawood. 2016. The cognitive and social motivational affordances of gamification in E-Learning environment. In *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016*. Institute of Electrical and Electronics Engineers Inc., (November 2016), 373–375. ISBN: 9781467390415. DOI: 10.1109/ICALT.2016.126.

[2] Gregory Afentoulidis, Jie Yang, Zoltán Szlávik, and Alessandro Bozzon. 2018. Social gamification in enterprise crowdsourcing. In *WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science*. Association for Computing Machinery, Inc, New York, NY, USA, (May 2018), 135–144. ISBN: 9781450355636. DOI: 10.1145/3201064.3201094.

[3] Charu C. Aggarwal. 2011. An Introduction to Social Network Data Analytics. In *Social Network Data Analytics*. Springer US, 1–15. DOI: 10.1007/978-1-4419-8462-3{_}1.

[4] Muhammad Waseem Ahmad, Monjur Mourshed, and Yacine Rezgui. 2017. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147, (July 2017), 77–89. ISSN: 03787788. DOI: 10.1016/j.enbuild.2017.04.038.

[5] Icek Ajzen. 1991. The Theory of Planned Behavior. *Organizational behavior and human decision processes*, 50, 179–211.

[6] N Aldenaini, F H Alqahtani, R Orji, and S Sampalli. 2020. Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. *Physical Activity and Sedentary Behavior: A Systematic Review. Front. Artif. Intell*, 3, 7. DOI: 10.3389/frai.2020.00007.

[7] Maximilian Altmeyer, Pascal Lessel, Linda Muller, and Antonio Krüger. 2019. Combining behavior change intentions and user types to select suitable gamification elements for persuasive fitness systems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 11433 LNCS. Springer Verlag, (April 2019), 337–349. ISBN: 9783030172862. DOI: 10.1007/978-3-030-17287-9{_}27.

[8] Maximilian Altmeyer, Marc Schubhan, Pascal Lessel, Linda Muller, and Antonio Krüger. 2020. Using hexad user types to select suitable gamification elements to encourage healthy eating. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (April 2020), 1–8. ISBN: 9781450368193. DOI: 10.1145/3334480.3383011.

[9] Chee Siang Ang and Panayiotis Zaphiris. 2010. Social Roles of Players in MMORPG Guilds. *Information, Communication & Society*, 13, 4, (June 2010), 592–614. ISSN: 1369-118X. DOI: 10.1080/13691180903266952.

[10] Álvar Arnaiz-González, Marcin Blachnik, Mirosław Kordos, and César García-Osorio. 2016. Fusion of instance selection methods in regression tasks. *Information Fusion*, 30, (July 2016), 69–79. ISSN: 15662535. DOI: 10.1016/j.inffus.2015.12.002.

[11] Yigal Attali and Meirav Arieli-Attali. 2015. Gamification in assessment: Do points affect test performance? *Computers and Education*, 83, (April 2015), 57–63. ISSN: 03601315. DOI: 10.1016/j.compedu.2014.12.012.

[12] Lars Backstrom, Ravi Kumar, Cameron Marlow, Jasmine Novak, and Andrew Tomkins. 2008. Preferential behavior in online groups. In *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM Press, New York, New York, USA, 117–127. DOI: 10.1145/1341531.1341549.

[13] Roger Bakeman and John M. Gottman. 1997. *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge University Press.

[14] Eytan Bakshy, Brian Karrer, and Lada A. Adamic. 2009. Social influence and the diffusion of user-created content. In *Proceedings of the ACM Conference on Electronic Commerce*. ACM Press, New York, New York, USA, 325–334. ISBN: 9781605584584. DOI: 10.1145/1566374.1566421.

[15] Eytan Bakshy, Winter A. Mason, Jake M. Hofman, and Duncan J. Watts. 2011. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011*, 65–74. ISBN: 9781450304931. DOI: 10.1145/1935826.1935845.

[16] Albert Bandura. 1991. Social Cognitive Theory of Self-Regulation. *Organizational behavior and Hyman Decision Processes*, 50, 248–287.

[17] Balázs Barna and Szabina Fodor. 2019. A data-driven approach to analyze user behavior on a personalized gamification platform. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 11899 LNCS. Springer, (November 2019), 266–275. ISBN: 9783030343491. DOI: 10.1007/978-3-030-34350-7{_}26.

[18] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research*, 1, 1.

[19] Gustavo E. A. P. A. Batista, Ronaldo C. Prati, and Maria Carolina Monard. 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6, 1, (June 2004), 20–29. ISSN: 1931-0145. DOI: 10.1145/1007730.1007735.

[20] Christian Bauckhage, Anders Drachen, and Rafet Sifa. 2015. Clustering Game Behavior Data. *IEEE Transactions on Computational Intelligence and AI in Games*, 7, 3, (September 2015), 266–278. ISSN: 1943068X. DOI: 10.1109/TCIAIG.2014.2376982.

[21] Roy F. Baumeister and Mark R. Leary. 1995. The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, 117, 3, 497–529. ISSN: 00332909. DOI: 10.1037/0033-2909.117.3.497.

[22] Paul Bertens, Anna Guitart, and Africa Perianez. 2017. Games and big data: A scalable multi-dimensional churn prediction model. In *2017 IEEE Conference on Computational Intelligence and Games, CIG 2017*. Institute of Electrical and Electronics Engineers Inc., (October 2017), 33–36. ISBN: 9781538632338. DOI: 10.1109/CIG.2017.8080412.

[23] Max V. Birk, Maximillian A. Friehs, and Regan L. Mandryk. 2017. Age-based preferences and player experience: A crowdsourced cross-sectional study.

In *CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, (October 2017), 157–170. ISBN: 9781450348980. DOI: 10.1145/3116595.3116608.

[24] Anita L. Blanchard. 2007. Developing a sense of virtual community measure. *Cyberpsychology and Behavior*, 10, 6, (December 2007), 827–830. ISSN: 10949313. DOI: 10.1089/cpb.2007.9946.

[25] Martin Böckle, Isabel Micheel, Markus Bick, and Jasminko Novak. 2018. A Design Framework for Adaptive Gamification Applications. In *the 51st Hawaii International Conference on System Sciences* /. ISBN: 9780998133119.

[26] Martin Böckle, Jasminko Novak, and Markus Bick. 2017. Towards Adaptive Gamification: a Synthesis of Current Developments. In *25th European Conference on Information Systems (ECIS)*, Research Papers, (June 2017).

[27] Phillip Bonacich. 1987. Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92, 5, (March 1987), 1170–1182. ISSN: 0002-9602. DOI: 10.1086/228631.

[28] Sándor Bozóki, László Csató, and József Temesi. 2016. An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248, 1, (January 2016), 211–218. ISSN: 03772217. DOI: 10.1016/j.ejor.2015.06.069.

[29] Paula Branco, Rita P Ribeiro, Luís Torgo, Bartosz Krawczyk, and Nuno Moniz. 2017. SMOGN: a Pre-processing Approach for Imbalanced Regression. In *Proceedings of Machine Learning Research*. Volume 74. PMLR, (October 2017), 36–50.

[30] Paula Branco, Luís Torgo, and Rita P. Ribeiro. 2016. A survey of predictive modeling on imbalanced domains. (August 2016). DOI: 10.1145/2907070.

[31] Barry Brown and Marek Bell. 2004. CSCW at play. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*. Association for Computing Machinery (ACM), New York, New York, USA, 350. DOI: 10.1145/1031607.1031666.

[32] Phil Brown and Junlan Feng. 2011. Measuring User Influence on Twitter Using Modified K-Shell Decomposition. In *The Social Mobile Web*. AAAI.

[33] Joachim C. Brunstein and Günter W. Maier. 2005. Implicit and Self-Attributed Motives to Achieve: Two Separate but Interacting Needs. *Journal of Personality and Social Psychology*, 89, 2, 205–222.

[34] Antonio Bucchiarone, Nelly Bencomo, Enrica Loria, Annapaola Marconi, and Antonio Cicchetti. 2020. Run-time and collective adaptation of gameful systems. In *2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*. IEEE Computer Society, 145–146.

[35] John T. Cacioppo, James H. Fowler, and Nicholas A. Christakis. 2009. Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology*, 97, 6, 977–991.

[36] Roger Caillois. 2011. *Man, Play, and Games*. University of Illinois Press.

[37] Alessandro Canossa, Ahmad Azadvar, Casper Harteveld, Anders Drachen, and Sebastian Deterding. 2019. Influencers in multiplayer online shooters evidence of social contagion in playtime and social play. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (May 2019), 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300489.

[38] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and Krishna P Gummadi. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. In *fourth international AAAI conference on weblogs and social media*.

[39] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, (June 2002), 321–357. ISSN: 10769757. DOI: 10.1613/jair.953.

[40] Shi Yi Chen, Zhe Feng, and Xiaolian Yi. 2017. A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, 9, 6, (June 2017), 1725–1729. ISSN: 20776624. DOI: 10.21037/jtd.2017.05.34.

[41] Christine Cook, Rianne Conijn, Juliette Schaafsma, and Marjolijn Antheunis. 2019. For Whom the Gamer Trolls: A Study of Trolling Interactions in the Online Gaming Context. *Journal of Computer-Mediated Communication*, 24, 6, (October 2019), 293–318. ISSN: 1083-6101. DOI: 10.1093/jcmc/zmz014.

[42] Kristof Coussement and Dirk Van den Poel. 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34, 1, (January 2008), 313–327. ISSN: 09574174. DOI: 10.1016/j.eswa.2006.09.038.

[43] David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, and Siddharth Suri. 2008. Feedback effects between similarity and social influence in online communities. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, New York, New York, USA, 160–168. ISBN: 9781605581934. DOI: 10.1145/1401890.1401914.

[44] Nicole Crenshaw and Bonnie Nardi. 2014. What's in a name? Naming practices in online video games. In *CHI PLAY 2014 - Proceedings of the 2014 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, New York, NY, USA, (October 2014), 67–76. ISBN: 9781450330145. DOI: 10.1145/2658537.2658685.

[45] Sven F. Crone and Steven Finlay. 2012. Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, 28, 1, (January 2012), 224–238. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2011.07.006.

[46] Laura Dabbish, Robert Kraut, and Jordan Patton. 2012. Communication and commitment in an online game team. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM Press, New York, New York, USA, 879–888. ISBN: 9781450310154. DOI: 10.1145/2207676.2208529.

[47] Yvonne A.W. De Kort and Wijnand A. Ijsselsteijn. 2008. People, places, and play: Player experience in a socio-spatial context. *Computers in Entertainment*, 6, 2, (July 2008), 1–11. ISSN: 15443574. DOI: 10.1145/1371216.1371221.

[48] Luis De-Marcos, Eva Garcia-Lopez, and Antonio Garcia-Cabot. 2016. On the effectiveness of game-like and social approaches in learning: Comparing educational gaming, gamification & social networking. *Computers and Education*, 95, (April 2016), 99–113. ISSN: 03601315. DOI: 10.1016/j.compedu.2015.12.008.

[49] Edward L. Deci and Richard M. Ryan. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 4, 227–268. ISSN: 1047840X. DOI: 10.1207/S15327965PLI1104{_}01.

[50] Simon Demediuk, † Peter York, Anders Drachen, † James, Alfred Walker, and Florian Block. 2019. Role Identification for Accurate Analysis in Dota 2. In

Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment number 1. Volume 15. (October 2019), 130–138.

[51] Mouna Denden, Ahmed Tlili, Fathi Essalmi, and Mohamed Jemni. 2017. An investigation of the factors affecting the perception of gamification and game elements. In *2017 6th International Conference on Information and Communication Technology and Accessbility, ICTA 2017*. Volume 2017-December. Institute of Electrical and Electronics Engineers Inc., (April 2017), 1–5. ISBN: 9781538644607. DOI: 10.1109/ICTA.2017.8336019.

[52] Sebastian Deterding. 2015. The lens of intrinsic skill atoms: A method for gameful design. (May 2015). DOI: 10.1080/07370024.2014.993471.

[53] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*. ACM Press, New York, New York, USA, 9–15. ISBN: 9781450308168. DOI: 10.1145/2181037.2181040.

[54] Anders Drachen, James Green, Chester Gray, Elie Harik, Patty Lu, Rafet Sifa, and Diego Klabjan. 2016. Guns and guardians: Comparative cluster analysis and behavioral profiling in destiny. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. Volume 0. IEEE Computer Society, (July 2016). ISBN: 9781509018833. DOI: 10.1109/CIG.2016.7860423.

[55] Anders Drachen, Diego Klabjan, Myat Aung, Simon Demediuk, Yuan Sun, Ye Tu, Yu Ang, Siva Nekkanti, Rafet Sifa, Shantanu Raghav, and Anders 2018 Drachen. 2019. The trails of Just Cause 2: spatio-temporal player profiling in open-world games The Trails of Just Cause 2: Spatio-Temporal Player Profiling in Open-World Games ACM Reference Format. In *FDG '19: ACM Foundations of Digital Game*. ACM, 1–11. ISBN: 9781450372176. DOI: 10.1145/3337722.3337765.

[56] Anders Drachen, Eric Thurston Lundquist, Yungjen Kung, Pranav Simha Rao, Diego Klabjan, Rafet Sifa, and Julian Runge. 2016. Rapid Prediction of Player Retention in Free-to-Play Mobile Games, (July 2016).

[57] Anders Drachen, Christian Thurau, Rafet Sifa, and Christian Bauckhage. 2014. A Comparison of Methods for Player Clustering via Behavioral Telemetry, (July 2014).

[58] Nicolas Ducheneaut and Robert J Moore. 2004. The Social Side of Gaming: A Study of Interaction Patterns in a Massively Multiplayer Online Game. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*. ACM Press, New York, New York, USA. ISBN: 1581138105.

[59] Nicolas Ducheneaut, Robert J. Moore, and Eric Nickell. 2007. Virtual "third places": A case study of sociability in massively multiplayer games. *Computer Supported Cooperative Work*, 16, 1-2, (May 2007), 129–166. ISSN: 15737551. DOI: 10.1007/s10606-007-9041-8.

[60] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J Moore. 2006. "Alone Together?" Exploring the Social Dynamics of Massively Multiplayer Online Games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*. ACM Press, New York, New York, USA. ISBN: 1595933727.

[61] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets.* Number 5. Volume 26. Cambridge University Press (CUP). DOI: 10.1017/s0266466609990685.

[62] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. *Game Analytics - Maximizing the Value of Player Data*. Springer London. DOI: 10.1007/978-1-4471-4769-5.

[63] Seymour Epstein. 1998. Cognitive-Experiential Self-Theory. In Springer, Boston, MA, 211–238. DOI: 10.1007/978-1-4419-8580-4{_}9.

[64] Christopher J. Ferguson. 2015. An effect size primer: A guide for clinicians and researchers. *Methodological issues and strategies in clinical research (4th ed.).*, (December 2015), 301–310. DOI: 10.1037/14805-020.

[65] Juliana Carvalho Ferreira and Cecilia Maria Patino. 2016. hat is survival analysis, and when should I use it? *Jornal Brasileiro de Pneumologia*, 42, 1, (January 2016), 77. ISSN: 18063756. DOI: 10.1590/S1806-37562016000000013.

[66] Lauren S. Ferro, Steffen P. Walz, and Stefan Greuter. 2013. Towards personalised, gamified systems: An investigation into game design, personality and player typologies. In *ACM International Conference Proceeding Series*. ACM Press, New York, New York, USA, 1–6. ISBN: 9781450322546. DOI: 10.1145/2513002.2513024.

[67] Michela Ferron, Enrica Loria, Annapaola Marconi, and Paolo Massa. 2019. Play&go, an urban game promoting behaviour change for sustainable mobility. *Interaction Des. Archit. J.*, 40, 24–25.

[68] Andy Field. 2013. *Discovering Statistics Using IBM SPSS Statistics*. (4th Edition edition). Sage.

[69] Guo Freeman. 2016. Making games as collaborative social experiences: Exploring an online gaming community. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Volume 26-February-2016. Association for Computing Machinery, New York, New York, USA, (February 2016), 265–268. ISBN: 9781450339506. DOI: 10.1145/2818052.2869076.

[70] Linton C Freeman. 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1, 215–239.

[71] Jerome H. Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 1189–1232.

[72] Andrew R. Gilpin. 1993. 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*, 53, 1, (March 1993), 87–92. ISSN: 0013-1644. DOI: 10.1177/0013164493053001007.

[73] Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. 2010. Learning influence probabilities in social networks. In *WSDM 2010 - Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*. ACM Press, New York, New York, USA, 241–250. ISBN: 9781605588896. DOI: 10.1145/1718487.1718518.

[74] Mark Granovetter. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91, 481–510. DOI: 10.2307/2780199.

[75] Mark S. Granovetter. 1973. The Strength of Weak Ties. *American Journal of Sociology*, 78, 6, (May 1973), 1360–1380. ISSN: 0002-9602. DOI: 10.1086/225469.

[76] Fabian Hadiji, Rafet Sifa, Anders Drachen, Christian Thurau, Kristian Kersting, and Christian Bauckhage. 2014. Predicting player churn in the wild. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. IEEE Computer Society, (October 2014). ISBN: 9781479935468. DOI: 10.1109/CIG.2014.6932876.

[77] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. 2017. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73, (May 2017), 220–239. ISSN: 09574174. DOI: 10.1016/j.eswa.2016.12.035.

[78] Stuart Hallifax, Elise Lavoué, and Audrey Serna. 2020. To tailor or not to tailor gamification? An analysis of the impact of tailored game elements on learners' behaviours and motivation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 12163 LNAI. Springer, (July 2020), 216–227. ISBN: 9783030522360. DOI: 10.1007/978-3-030-52237-7{_}18.

[79] Stuart Hallifax, Audrey Serna, Jean Charles Marty, Guillaume Lavoué, and Elise Lavoué. 2019. Factors to consider for tailored gamification. In *CHI PLAY 2019 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, New York, NY, USA, (October 2019), 559–572. ISBN: 9781450366885. DOI: 10.1145/3311350.3347167.

[80] Stuart Hallifax, Audrey Serna, Jean Charles Marty, and Élise Lavoué. 2019. Adaptive Gamification in Education: A Literature Review of Current Trends and Developments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 11722 LNCS. Springer Verlag, (September 2019), 294–307. ISBN: 9783030297350. DOI: 10.1007/978-3-030-29736-7{_}22.

[81] Juho Hamari. 2013. Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications*, 12, 4, (July 2013), 236–245. ISSN: 15674223. DOI: 10.1016/j.elerap.2013.01.004.

[82] Juho Hamari and Jonna Koivisto. 2013. Social Motivations To Use Gamification: An Empirical Study Of Gamifying Exercise. *Completed Research*, 105.

[83] Juho Hamari and Jonna Koivisto. 2015. "Working out for likes": An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, 50, (September 2015), 333–347. ISSN: 07475632. DOI: 10.1016/j.chb.2015.04.018.

[84] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does gamification work? - A literature review of empirical studies on gamification. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE Computer Society, 3025–3034. ISBN: 9781479925049. DOI: 10.1109/HICSS.2014.377.

[85] Jiawei Han, Micheline Kamber, and Jian Pei. 2012. Getting to Know Your Data. In *Data Mining*. Elsevier, (January 2012), 39–82. DOI: 10.1016/b978-0-12-381479-1.00002-2.

[86] John Harris and Mark Hancock. 2019. To asymmetry and beyond! Improving social connectedness by increasing designed interdependence in cooperative play. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (May 2019), 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300239.

[87] Lobna Hassan, Antonio Dias, and Juho Hamari. 2019. How motivational feedback increases user's benefits and continued use: A study on gamification, quantified-self and social networking. *International Journal of Information Management*, 46, (June 2019), 151–162. ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2018.12.004.

[88] Lobna Hassan, Jere Rantalainen, Nannan Xi, Henri Pirkkalainen, and Juho Hamari. 2020. The relationship between player types and gamification feature preferences. In *Proceedings of the 4th International GamiFIN Conference*. CEUR workshop proceedings, 11–20.

[89] Benjamin Heilbrunn, Philipp Herzig, and Alexander Schill. 2017. Gamification Analytics—Methods and Tools for Monitoring and Adapting Gamification Designs. In Springer, Cham, 31–47. DOI: 10.1007/978-3-319-45557-0{_}3.

[90] Wilhelm Hofmann, Bertram Gawronski, Tobias Gschwendner, Huy Le, and Manfred Schmitt. 2005. A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. (October 2005). DOI: 10.1177/0146167205275613.

[91] Danial Hooshyar, Moslem Yousefi, and Heuiseok Lim. 2018. A Procedural Content Generation-Based Framework for Educational Games: Toward a Tailored Data-Driven Game for Developing Early English Reading Skills. *Journal of Educational Computing Research*, 56, 2, (April 2018), 293–310. ISSN: 0735-6331. DOI: 10.1177/0735633117706909.

[92] Danial Hooshyar, Moslem Yousefi, and Heuiseok Lim. 2018. Data-driven approaches to game player modeling: A systematic literature review. (January 2018). DOI: 10.1145/3145814.

[93] David C. Howell. 2012. *Statistical Methods for Psychology - David C. Howell - Google Books*. Cengage Learning.

[94] Ji Lung Hsieh and Chuen Tsai Sun. 2008. Building a player strategy model by analyzing replays of real-time strategy games. In *Proceedings of the International Joint Conference on Neural Networks*, 3106–3111. ISBN: 9781424418213. DOI: 10.1109/IJCNN.2008.4634237.

[95] Yun Huang, Wenyue Ye, Nicholas Bennett, and Noshir S. Contractor. 2013. Functional or social? Exploring teams in online games. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. ACM Press, New York, New York, USA, 399–408. ISBN: 9781450313315. DOI: 10.1145/2441776.2441822.

[96] Kai Huotari and Juho Hamari. 2017. A definition for gamification: anchoring gamification in the service marketing literature. *Electronic Markets*, 27, 1, (February 2017), 21–31. ISSN: 14228890. DOI: 10.1007/s12525-015-0212-z.

[97] Katja Hutter, Johann Füller, Julia Hautz, Volker Bilgram, and Kurt Matzler. 2015. Machiavellianism or Morality: Which Behavior Pays Off In Online Innovation Contests? *Journal of Management Information Systems*, 32, 3, (July 2015), 197–228. ISSN: 0742-1222. DOI: 10.1080/07421222.2015.1099181.

[98] Wijnand Ijsselsteijn, Yvonne De Kort, Karolien Poels, Audrius Jurgelionis, and Francesco Bellotti. 2007. Characterising and Measuring User Experiences in Digital Games. In *International conference on advances in computer entertainment technology*. Volume 2, 27.

[99] Katherine Isbister and Noah Schaffer. 2008. *Game Usability: Advancing the Player Experience*.

[100] Tomislav Jagušt, Ivica Botički, and Hyo Jeong So. 2018. Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers and Education*, 125, (October 2018), 444–457. ISSN: 03601315. DOI: 10.1016/j.compedu.2018.06.022.

[101] Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6, 5, (January 2002), 429–449. ISSN: 15714128. DOI: 10.3233/ida-2002-6504.

[102] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies*, 66, 9, (September 2008), 641–661. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2008.04.004.

[103] Yuan Jia, Bin Xu, Yamini Karanam, and Stephen Voida. 2016. Personality-targeted Gamification: A Survey Study on Personality Traits and Motivational

Affordances. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (May 2016), 2001–2013. ISBN: 9781450333627. DOI: 10.1145/2858036.2858515.

[104] Jialun Jiang and Svetlana Yarosh. 2016. Why do teammates hate me? Cross-cultural tensions and social dynamics in online games. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Volume 26-February-2016. Association for Computing Machinery, New York, New York, USA, (February 2016), 301–304. ISBN: 9781450339506. DOI: 10.1145/2818052.2869087.

[105] Niklas Johannes, Matti Vuorre, Andrew K Przybylski, and Andrew K Przybylski Methodology. 2020. Video game play is positively correlated with well-being. *PsyArXiv*.

[106] Jina Kang, Min Liu, and Wen Qu. 2017. Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72, (July 2017), 757–770. ISSN: 07475632. DOI: 10.1016/j.chb.2016.09.062.

[107] Maurits Kaptein, Boris De Ruyter, Panos Markopoulos, and Emile Aarts. 2012. Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. (June 2012). DOI: 10.1145/2209310.2209313.

[108] Jaya Kawale, Aditya Pal, and Jaideep Srivastava. 2009. Churn prediction in MMORPGs: A social influence based approach. In *Proceedings - 12th IEEE International Conference on Computational Science and Engineering, CSE 2009*. Volume 4, 423–428. ISBN: 9780769538235. DOI: 10.1109/CSE.2009.80.

[109] Raman Kazhamiakin, Enrica Loria, Annapaola Marconi, and Mauro Scanagatta. 2021. A gamification platform to analyze and influence citizens' daily transportation choices. *IEEE Transactions on Intelligent Transportation Systems*, Special Issue on Diversity in Transportation Systems for People and Goods.

[110] Edward Keller and Jonathan Berry. 2003. *The influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy.*

[111] M. G. Kendall and B. Babington Smith. 1940. On the Method of Paired Comparisons. *Biometrika*, 31, 3/4, (March 1940), 324. ISSN: 00063444. DOI: 10.2307/2332613.

[112] Reza Khoshkangini, Giuseppe Valetto, Annapaola Marconi, and Marco Pistore. 2020. Automatic generation and recommendation of personalized challenges for gamification. *User Modeling and User-Adapted Interaction*, (May 2020), 1–34. ISSN: 15731391. DOI: 10.1007/s11257-019-09255-2.

[113] Michael D. Kickmeier-Rust, Eva-C Hillemann, and Albert Dietrich. 2014. Gamification and smart feedback: Experiences with a primary school level math app. *International Journal of Game-Based Learning (IJGBL)*, 4, 3, 35–46.

[114] Kristian Kiili, Korbinian Moeller, and Manuel Ninaus. 2018. Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators. *Computers and Education*, 120, (May 2018), 13–28. ISSN: 03601315. DOI: 10.1016/j.compedu.2018.01.012.

[115] Jun H. Kim, Daniel V. Gunn, Eric Schuh, Bruce C. Phillips, Randy J. Pagulayan, and Dennis Wixon. 2008. Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM Press, New York, New York, USA, 443–451. ISBN: 9781605580111. DOI: 10.1145/1357054.1357126.

[116] Seungwook Kim, Daeyoung Choi, Eunjung Lee, and Wonjong Rhee. 2017. Churn prediction of mobile and online casual games using play log data. *PLOS*

ONE, 12, 7, (July 2017), e0180735. Wei-Xing Zhou, editor. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0180735.

[117] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu Ting Lin, Naomi McArthur, and Thomas W. Malone. 2017. What makes a strong team? Using collective intelligence to predict team performance in League of Legends. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Association for Computing Machinery, New York, NY, USA, (February 2017), 2316–2329. ISBN: 9781450343350. DOI: 10.1145/2998181.2998185.

[118] Rex B. Kline. 2011. *Principles and Practice of Structural Equation Modeling*. (Fourth Edition edition). The Guilford Press.

[119] Ana Carolina Tomé Klock, Isabela Gasparini, Marcelo Soares Pimenta, and Juho Hamari. 2020. Tailored gamification: A review of literature. *International Journal of Human Computer Studies*, 144, (December 2020), 102495. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2020.102495.

[120] Antti Knutas, Rob van Roy, Timo Hynninen, Marco Granato, Jussi Kasurinen, and Jouni Ikonen. 2019. A process for designing algorithm-based personalized gamification. *Multimedia Tools and Applications*, 78, 10, (May 2019), 13593–13612. ISSN: 15737721. DOI: 10.1007/s11042-018-6913-5.

[121] Jonna Koivisto and Juho Hamari. 2019. The rise of motivational information systems: A review of gamification research. (April 2019). DOI: 10.1016/j.ijinfomgt.2018.10.013.

[122] Dimosthenis Kotsopoulos, Cleopatra Bardaki, Stavros Lounis, and Katerina Pramatari. 2018. Employee Profiles and Preferences towards IoT-enabled Gamification for Energy Conservation. *International Journal of Serious Games*, 5, 2, (June 2018), 65–85. ISSN: 2384-8766. DOI: 10.17083/ijsg.v5i2.225.

[123] Markus Krause, Marc Mogalle, Henning Pohl, and Joseph Jay Williams. 2015. A playful game changer: Fostering student retention in online education with social gamification. In *L@S 2015 - 2nd ACM Conference on Learning at Scale*. Association for Computing Machinery, Inc, New York, NY, USA, (March 2015), 95–102. ISBN: 9781450334112. DOI: 10.1145/2724660.2724665.

[124] Robert E. 1946 Kraut, Paul. Resnick, and Sara 1940 Kiesler. 2011. *Building successful online communities : evidence-based social design*. The MIT Press. ISBN: 9780262528917.

[125] Bartosz Krawczyk. 2016. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5, 4, (November 2016), 221–232. ISSN: 21926360. DOI: 10.1007/s13748-016-0094-0.

[126] Gregory W. Kurtzman, Susan C. Day, Dylan S. Small, Marta Lynch, Jingsan Zhu, Wenli Wang, Charles A.L. Rareshide, and Mitesh S. Patel. 2018. Social Incentives and Gamification to Promote Weight Loss: The LOSE IT Randomized, Controlled Trial. *Journal of General Internal Medicine*, 33, 10, (October 2018), 1669–1675. ISSN: 15251497. DOI: 10.1007/s11606-018-4552-1.

[127] Kwan Min Lee. 2004. Presence, Explicated. *Communication Theory*, 14, 1, (February 2004), 27–50. ISSN: 1050-3293. DOI: 10.1111/j.1468-2885.2004.tb00302.x.

[128] Nikoletta Zampeta Legaki, Nannan Xi, Juho Hamari, and Vassilios Assimakopoulos. 2019. Gamification of The Future: An Experiment on Gamifying Education of Forecasting. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences. DOI: 10.24251/hicss.2019.219.

[129] Frédéric B. Leloup, Michael R. Pointer, Philip Dutré, and Peter Hanselaer. 2010. Geometry of illumination, luminance contrast, and gloss perception. *Journal of the Optical Society of America A*, 27, 9, (September 2010), 2046. ISSN: 1084-7529. DOI: 10.1364/josaa.27.002046.

[130] Huan Liu. 2002. On Issues of Instance Selection. *Data Mining and Knowledge Discovery*, 6, 115–130.

[131] Xi Liu, Muhe Xie, Xidao Wen, Rui Chen, Yong Ge, Nick Duffield, and Na Wang. 2018. A Semi-Supervised and Inductive Embedding Model for Churn Prediction of Large-Scale Mobile Games. In *Proceedings - IEEE International Conference on Data Mining, ICDM*. Volume 2018-November. Institute of Electrical and Electronics Engineers Inc., (December 2018), 277–286. ISBN: 9781538691588. DOI: 10.1109/ICDM.2018.00043.

[132] Xi Liu, Muhe Xie, Xidao Wen, Rui Chen, Yong Ge, Nick Duffield, and Na Wang. 2020. Micro- and macro-level churn analysis of large-scale mobile games. *Knowledge and Information Systems*, 62, 4, (April 2020), 1465–1496. ISSN: 02193116. DOI: 10.1007/s10115-019-01394-7.

[133] Christian Sebastian Loh, Yanyan Sheng, and Dirk Ifenthaler. 2015. Serious games analytics: Theoretical framework. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*. Springer International Publishing, (January 2015), 3–29. ISBN: 9783319058344. DOI: 10.1007/978-3-319-05834-4{_}1.

[134] Christian Lopez and Conrad Tucker. 2020. Toward Personalized Adaptive Gamification: A Machine Learning Model for Predicting Performance. *IEEE Transactions on Games*, 12, 2, (June 2020), 155–168. ISSN: 24751510. DOI: 10.1109/TG.2018.2883661.

[135] Enrica Loria. 2019. A framework to infer player experience and inform customized content generation in gameful systems. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 21–28.

[136] Enrica Loria, Maximillian Altmeyer, Stuart Hallifax, Lennart E. Nacke, and Annapaola Marconi. Under Preparation. Like peas in a pod: the impact of theory and data-driven player profiling on adaptive gamification. In *International Conference on Human-Computer Interaction in Play (CHI Play)*. ACM.

[137] Enrica Loria, Alessia Antelmi, Johanna Pirker, and Carmine Spagnuolo. Submitted. Stay and play – modeling social influence in the steam players community. In *Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*. ACM.

[138] Enrica Loria and Annapaola Marconi. 2019. A data-driven approach to deduce players' preferences from in-game interactions in gameful systems. In *1st International Workshop on Game Analytics*. In Press. CEUR Workshop.

[139] Enrica Loria and Annapaola Marconi. Minor Revision. Exploiting limited players' behavioral data to predict churn in gamification. *Electronic Commerce Research and Applications*, Gamification Special Issue.

[140] Enrica Loria and Annapaola Marconi. 2018. Player types and player behaviors: analyzing correlations in an on-the-field gamified system. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 531–538.

[141] Enrica Loria and Annapaola Marconi. 2020. Reading between the lines–towards an algorithm exploiting in-game behaviors to learn preferences in gameful systems. In *International Conference on the Foundations of Digital Games (FDG)*. IEEE Computer Society, 1–12.

[142] Enrica Loria, Lennart E. Nacke, and Annapaola Marconi. Under Preparation. Klout like a kardashian: the power of gamification influencers in an open-world gameful system. In *International Conference on Human-Computer Interaction in Play (CHI Play)*. ACM.

[143] Enrica Loria, Francesco Paissan, and Annapaola Marconi. 2019. Exploiting general-purpose in-game behaviours to predict players churn in gameful systems. In *1st International Workshop on Game Analytics*. In Press. CEUR Workshop.

[144] Enrica Loria, Johanna Pirker, Anders Drachen, and Annapaola Marconi. 2020. Do influencers influence?–analyzing players' activity in an online multiplayer game. In *International Conference on Games (CoG)*. ACM.

[145] Enrica Loria, Jessica Riviera, and Annapaola Marconi. 2021. Do they play as intended? – comparing aggregated and temporal behavioral analysis in a persuasive gamified system. In *2021 54th Hawaii international conference on system sciences*.

[146] Vijay Mahajan, Eitan Muller, and Frank M. Bass. 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing*, 54, 1, (January 1990), 1–26. ISSN: 0022-2429. DOI: 10.1177/002224299005400101.

[147] Tobias Mahlmann, Anders Drachen, Julian Togelius, Alessandro Canossa, and Georgios N. Yannakakis. 2010. Predicting player behavior in Tomb Raider: Underworld. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, CIG2010*, 178–185. ISBN: 9781424462971. DOI: 10.1109/ITW.2010.5593355.

[148] Sasha Makarovych, Alessandro Canossa, Julian Togelius, and Anders Drachen. 2018. Like a DNA string : Sequence-based Player Profiling in Tom Clancy's The Division. In *Artificial Intelligence and Interactive Digital Entertainment Conference*, 13–17.

[149] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. 2006. A continuous and objective evaluation of emotional experience with interactive play environments. In *Conference on Human Factors in Computing Systems - Proceedings*. Volume 2. Association for Computing Machinery, New York, New York, USA, 1027–1036. ISBN: 1595931783. DOI: 10.1145/1124772.1124926.

[150] Regan L. Mandryk, Kori M. Inkpen, and Thomas W. Calvert. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour and Information Technology*, 25, 2, (March 2006), 141–158. ISSN: 0144929X. DOI: 10.1080/01449290500331156.

[151] Annapaola Marconi and Enrica Loria. 2020. 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: Emerging Research and Opportunities*. IGI Global, 211–246.

[152] Edoardo Marcucci, Valerio Gatta, and Michela Le Pira. 2018. Gamification design to foster stakeholder engagement and behavior change: An application to urban freight transport. *Transportation Research Part A: Policy and Practice*, 118, (December 2018), 119–132. ISSN: 09658564. DOI: 10.1016/j.tra.2018.08.028.

[153] Andrzej Marczewski. 2015. User Types. In *Even Ninja Monkeys like to play*. CreateSpace Indep., 69–84.

[154] David C. McClelland. 1985. How Motives, Skills, and Values Determine What People Do. *American Psychologist*, 40, 7, (July 1985), 812–825. ISSN: 0003066X. DOI: 10.1037/0003-066X.40.7.812.

[155] David C. McClelland. 1987. *Human Motivation*. Cambridge University Press.

[156] David C. McClelland, Richard Koestner, and Joel Weinberger. 1989. How Do Self-Attributed and Implicit Motives Differ? *Psychological Review*, 96, 4, 690–702. ISSN: 0033295X. DOI: 10.1037/0033-295X.96.4.690.

[157] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, 1, (August 2001), 415–444. ISSN: 0360-0572. DOI: 10.1146/annurev.soc.27.1.415.

[158] David Melhart, Ahmad Azadvar, Alessandro Canossa, Antonios Liapis, and Georgios N. Yannakakis. 2019. Your gameplay says it all: Modelling motivation in tom clancy's the division. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. Volume 2019-August. IEEE Computer Society, (August 2019). ISBN: 9781728118840. DOI: 10.1109/CIG.2019.8848123.

[159] Christoph Molnar. 2019. *Interpretable Machine Learning*.

[160] Baptiste Monterrat, Élise Lavoué, and Sébastien George. 2017. Adaptation of Gaming Features for Motivating Learners. *Simulation & Gaming*, 48, 5, (October 2017), 625–656. ISSN: 1046-8781. DOI: 10.1177/1046878117712632.

[161] Alberto Mora, Gustavo F. Tondello, Laura Calvet, Carina González, Joan Arnedo-Moreno, and Lennart E. Nacke. 2019. The quest for a better tailoring of gameful design: An analysis of player type preferences. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, New York, NY, USA, (June 2019), 1–8. ISBN: 9781450371766. DOI: 10.1145/3335595.3335625.

[162] Alberto Mora, Gustavo F. Tondello, Lennart E. Nacke, and Joan Arnedo-Moreno. 2018. Effect of personalized gameful design on student engagement. In *IEEE Global Engineering Education Conference, EDUCON*. Volume 2018-April. IEEE Computer Society, (May 2018), 1925–1933. ISBN: 9781538629574. DOI: 10.1109/EDUCON.2018.8363471.

[163] Benedikt Morschheuser and Juho Hamari. 2019. The Gamification of Work: Lessons From Crowdsourcing. *Journal of Management Inquiry*, 28, 2, (April 2019), 145–148. ISSN: 1056-4926. DOI: 10.1177/1056492618790921.

[164] Benedikt Morschheuser, Marc Riar, Juho Hamari, and Alexander Maedche. 2017. How games induce cooperation? A study on the relationship between game features and we-intentions in an augmented reality game. *Computers in Human Behavior*, 77, (December 2017), 169–183. ISSN: 07475632. DOI: 10.1016/j.chb.2017.08.026.

[165] Lennart E. Nacke, Chris Bateman, and Regan L. Mandryk. 2014. BrainHex: A neurobiological gamer typology survey. *Entertainment Computing*, 5, 1, (January 2014), 55–62. ISSN: 18759521. DOI: 10.1016/j.entcom.2013.06.002.

[166] Lennart E. Nacke and Sebastian Deterding. 2017. The maturing of gamification research. *Computers in Human Behaviour*. ISSN: 0747-5632. DOI: 10.1016/j.chb.2016.11.062.

[167] Lennart E Nacke, Matthias Klauser, and Paul Prescod. 2015. Social Player Analytics in a Facebook Health Game. In *Proceedings of HCI KOREA 2015*, 180–187.

[168] M. E.J. Newman. 2005. A measure of betweenness centrality based on random walks. *Social Networks*, 27, 1, (January 2005), 39–54. ISSN: 03788733. DOI: 10.1016/j.socnet.2004.11.009.

[169] Scott Nicholson. 2012. A User-Centered Theoretical Framework for Meaningful Gamification. In *Games+Learning+Society 8.0*. Madison, WI.

[170] Mark O'Neill, Elham Vaziripour, Justin Wu, and Daniel Zappala. 2016. Condensing steam: Distilling the diversity of gamer behavior. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*. Volume 14-16-November-2016. Association for Computing Machinery, New York, NY, USA, (November 2016), 81–95. ISBN: 9781450345262. DOI: 10.1145/2987443.2987489.

[171] Rita Orji, Regan L. Mandryk, and Julita Vassileva. 2015. Gender, age, and responsiveness to cialdini's persuasion strategies. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 9072. Springer Verlag, 147–159. ISBN: 9783319203058. DOI: 10.1007/978-3-319-20306-5{_}14.

[172] Rita Orji, Regan L. Mandryk, and Julita Vassileva. 2017. Improving the efficacy of games for change using personalization models. *ACM Transactions on Computer-Human Interaction*, 24, 5, (October 2017). ISSN: 15577325. DOI: 10.1145/3119929.

[173] Rita Orji, Lennart E. Nacke, and Chrysanne Di Marco. 2017. Towards personality-driven persuasive health games and gamified systems. In *Conference on Human Factors in Computing Systems - Proceedings*. Volume 2017-May. Association for Computing Machinery, (May 2017), 1015–1027. ISBN: 9781450346559. DOI: 10.1145/3025453.3025577.

[174] Rita Orji, Gustavo F. Tondello, and Lennart E. Nacke. 2018. Personalizing persuasive strategies in gameful systems to gamification user types. In *Conference on Human Factors in Computing Systems - Proceedings*. Volume 2018-April. Association for Computing Machinery, (April 2018). ISBN: 9781450356206. DOI: 10.1145/3173574.3174009.

[175] Ranilson Paiva, Ig Ibert Bittencourt, Thyago Tenório, Patricia Jaques, and Seiji Isotani. 2016. What do students do on-line? Modeling students' interactions to improve their learning experience. *Computers in Human Behavior*, 64, (November 2016), 769–781. ISSN: 07475632. DOI: 10.1016/j.chb.2016.07.048.

[176] Hyunsoo Park and Kyung Joong Kim. 2015. Social network analysis of high-level players in multiplayer online battle arena game. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 8852. Springer Verlag, (November 2015), 223–226. ISBN: 9783319151670. DOI: 10.1007/978-3-319-15168-7{_}28.

[177] Olena Pastushenko, Wilk Oliveira, Seiji Isotani, and Tomáš Hruška. 2020. A methodology for multimodal learning analytics and flow experience identification within gamified assignments. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (April 2020), 1–9. ISBN: 9781450368193. DOI: 10.1145/3334480.3383060.

[178] Maria Perez-Ortiz and Rafal K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv*, (December 2017).

[179] Africa Perianez, Alain Saas, Anna Guitart, and Colin Magne. 2016. Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. In *Proceedings - 3rd IEEE International Conference on Data*

Science and Advanced Analytics, DSAA 2016. Institute of Electrical and Electronics Engineers Inc., (December 2016), 564–573. ISBN: 9781509052066. DOI: 10.1109/DSAA.2016.84.

[180] Johannes Pfau, Jan David Smeddinck, and Rainer Malaka. 2020. Enemy Within: Long-term Motivation Effects of Deep Player Behavior Models for Dynamic Difficulty Adjustment. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, New York, NY, USA, (April 2020), 1–10. ISBN: 9781450367080. DOI: 10.1145/3313831.3376423.

[181] Johanna Pirker, Simone Griesmayr, Anders Drachen, and Rafet Sifa. 2016. How playstyles evolve: Progression analysis and profiling in Just Cause 2. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 9926 LNCS. Springer Verlag, 90–101. ISBN: 9783319460994. DOI: 10.1007/978-3-319-46100-7{_}8.

[182] Johanna Pirker, André Rattinger, Anders Drachen, and Rafet Sifa. 2018. Analyzing player networks in Destiny. *Entertainment Computing*, 25, (March 2018), 71–83. ISSN: 18759521. DOI: 10.1016/j.entcom.2017.12.001.

[183] Andrew K. Przybylski, C. Scott Rigby, and Richard M. Ryan. 2010. A Motivational Model of Video Game Engagement. *Review of General Psychology*, 14, 2, (June 2010), 154–166. ISSN: 1089-2680. DOI: 10.1037/a0019440.

[184] Amon Rapp, Frank Hopfgartner, Juho Hamari, Conor Linehan, and Federica Cena. 2019. Strengthening gamification studies: Current trends and future opportunities of gamification research. (July 2019). DOI: 10.1016/j.ijhcs.2018.11.007.

[185] André Rattinger, Günter Wallner, Anders Drachen, Johanna Pirker, and Rafet Sifa. 2016. Integrating and inspecting combined behavioral profiling and social network models in Destiny. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Volume 9926 LNCS. Springer Verlag, 77–89. ISBN: 9783319460994. DOI: 10.1007/978-3-319-46100-7{_}7.

[186] Luís Filipe Rodrigues, Abílio Oliveira, and Carlos J. Costa. 2016. Playing seriously - How gamification and social cues influence bank customers to use gamified e-business applications. *Computers in Human Behavior*, 63, (October 2016), 392–407. ISSN: 07475632. DOI: 10.1016/j.chb.2016.05.063.

[187] Ryan Rogers. 2017. The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach. *Computers in Human Behavior*, 73, (August 2017), 446–450. ISSN: 07475632. DOI: 10.1016/j.chb.2017.03.048.

[188] J. Niels Rosenquist, Joanne Murabito, James H. Fowler, and Nicholas A. Christakis. 2010. The spread of alcohol consumption behavior in a large social network. *Annals of Internal Medicine*, 152, 7, (April 2010), 426–433. ISSN: 15393704. DOI: 10.7326/0003-4819-152-7-201004060-00007.

[189] Michael Ruckenstein. 1991. Leisure and Ethics Reflections on the Philosophy of Leisure S Of PAR1MFN7 OF EOLICATiON 5i L A 1, fill, FO. *Leisure and Ethics*.

[190] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 1, 68–78.

[191] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, 30, 4, (December 2006), 347–363. ISSN: 01467239. DOI: 10.1007/s11031-006-9051-8.

[192] Alain Saas, Anna Guitart, and Africa Perianez. 2016. Discovering playing patterns: Time series clustering of free-to-play game data. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. Volume 0. IEEE Computer Society, (July 2016). ISBN: 9781509018833. DOI: 10.1109/CIG.2016.7860442.

[193] Jeffrey S. Saltz, Starr Roxanne Hiltz, and Murray Turoff. 2004. Student social graphs. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*. Association for Computing Machinery (ACM), New York, New York, USA, 596. DOI: 10.1145/1031607.1031709.

[194] Mirweis Sangin. 2018. Observing the player experience: The art and craft of observing and documenting Games User Research. In *Games User Research*. Oxford University Press, (March 2018), 175–188. ISBN: 9780198794844. DOI: 10.1093/oso/9780198794844.003.0011.

[195] Michael H. Schiller, Günter Wallner, Christopher Schinnerl, Alexander Monte Calvo, Johanna Pirker, Rafet Sifa, and Anders Drachen. 2018. Inside the group: Investigating social structures in player groups and their influence on activity. *IEEE Transactions on Games*, 11, 4, (December 2018), 416–425. ISSN: 24751510. DOI: 10.1109/TG.2018.2858024.

[196] Erwan Scornet. 2017. Tuning parameters in random forests. *ESAIM: Proceedings and Surveys*, 60, (December 2017), 144–162. Jean-François Coeurjolly and Adeline Leclercq-Samson, editors. ISSN: 2267-3059. DOI: 10.1051/proc/201760144.

[197] Jerry Scripps, Pang Ning Tan, and Abdol Hossein Esfahanian. 2009. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, New York, New York, USA, 747–755. ISBN: 9781605584959. DOI: 10.1145/1557019.1557102.

[198] Katie Seaborn and Deborah I. Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human Computer Studies*, 74, (July 2015), 14–31. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2014.09.006.

[199] Alissa Sherry and Robin K. Henson. 2005. Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of Personality Assessment*, 84, 1, 37–48. ISSN: 00223891. DOI: 10.1207/s15327752jpa8401{_}09.

[200] Xiaolin Shi, Jun Zhu, Rui Cai, and Lei Zhang. 2009. User grouping behavior in online forums. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, New York, New York, USA, 777–785. ISBN: 9781605584959. DOI: 10.1145/1557019.1557105.

[201] Grace Shin, Yuanyuan Feng, Mohammad Hossein Jarrahi, and Nicci Gafinowitz. 2019. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. *JAMIA Open*, 2, 1, (April 2019), 62–72. ISSN: 2574-2531. DOI: 10.1093/jamiaopen/ooy048.

[202] Rafet Sifa, Anders Drachen, and Christian Bauckhage. 2018. Profiling in Games: Understanding Behavior from Telemetry. In *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective*, 337–375.

[203] Elton Sarmanho Siqueira, Carla Denise Castanho, Genaina Nunes Rodrigues, and Ricardo Pezzoul Jacobi. 2018. A data analysis of player in world of warcraft using game data mining. In *Brazilian Symposium on Games and Digital*

Entertainment, SBGAMES. Volume 2017-November. IEEE Computer Society, (June 2018), 1–9. ISBN: 9781538648469. DOI: 10.1109/SBGames.2017.00009.

[204] David S. Siroky. 2009. Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*, 3, 147–163. ISSN: 19357516. DOI: 10.1214/07-SS033.

[205] Neil R. Smalheiser. 2017. Nonparametric Tests. In *Data Literacy*. Elsevier, (January 2017), 157–167. DOI: 10.1016/b978-0-12-811306-6.00012-9.

[206] Sam Snodgrass, Omid Mohaddesi, Jack Hart, Guillermo Romera Rodriguez, Christoffer Holmgård, and Casper Harteveld. 2019. Like PEAS in PoDS: The player, environment, agents, system framework for the personalization of digital systems. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, (August 2019). ISBN: 9781450372176. DOI: 10.1145/3337722.3337756.

[207] Sam Snodgrass, Omid Mohaddesi, and Casper Harteveld. 2019. Towards a generalized player model through the PEAS framework. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, New York, NY, USA, (August 2019), 1–7. ISBN: 9781450372176. DOI: 10.1145/3337722.3341856.

[208] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning An Introduction*. (Second Edition edition). MIT Press.

[209] Lee Taber and Steve Whittaker. 2018. Personality depends on the medium: Differences in self-perception on Snapchat, Facebook, and offline. In *Conference on Human Factors in Computing Systems - Proceedings*. Volume 2018-April. Association for Computing Machinery, New York, NY, USA, (April 2018), 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3174181.

[210] Marco Tamassia, William Raffe, Rafet Sifa, Anders Drachen, Fabio Zambetta, and Michael Hitchens. 2016. Predicting player churn in destiny: A Hidden Markov models approach to predicting player departure in a major online game. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. Volume 0. IEEE Computer Society, (July 2016). ISBN: 9781509018833. DOI: 10.1109/CIG.2016.7860431.

[211] Chenhao Tan, Jie Tang, Jimeng Sun, Quan Lin, and Fengjiao Wang. 2010. Social action tracking via Noise Tolerant Time-varying Factor Graphs. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, New York, New York, USA, 1049–1058. ISBN: 9781450300551. DOI: 10.1145/1835804.1835936.

[212] Ching I. Teng. 2017. Impact of avatar identification on online gamer loyalty: Perspectives of social identity and social capital theories. *International Journal of Information Management*, 37, 6, (December 2017), 601–610. ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2017.06.006.

[213] Sarah Kristin Thiel. 2016. Reward-based vs. Social gamification: Exploring effectiveness of gamefulness in public participation. In *ACM International Conference Proceeding Series*. Volume 23-27-October-2016. Association for Computing Machinery, New York, NY, USA, (October 2016), 1–6. ISBN: 9781450347631. DOI: 10.1145/2971485.2996739.

[214] Armando Toda, Paula Palomino, Luiz Rodrigues, Wilk Oliveira, Lei Shi, Seiji Isotani, and Alexandra Cristea. 2020. Validating the Effectiveness of Data-Driven Gamification Recommendations: An Exploratory Study, (August 2020), 763. DOI: 10.5753/cbie.sbie.2019.763.

[215] Gustavo F. Tondello, Dennis L. Kappen, Elisa D. Mekler, Marim Ganaba, and Lennart E. Nacke. 2016. Heuristic evaluation for gameful design. In *CHI PLAY*

2016 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion. Association for Computing Machinery, Inc, (October 2016), 315–323. ISBN: 9781450344586. DOI: 10.1145/2968120.2987729.

[216] Gustavo F. Tondello, Alberto Mora, Andrzej Marczewski, and Lennart E. Nacke. 2019. Empirical validation of the Gamification User Types Hexad scale in English and Spanish. *International Journal of Human Computer Studies*, 127, (July 2019), 95–111. ISSN: 10959300. DOI: 10.1016/j.ijhcs.2018.10.002.

[217] Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. 2017. Elements of gameful design emerging from user preferences. In *CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, New York, NY, USA, (October 2017), 129–140. ISBN: 9781450348980. DOI: 10.1145/3116595.3116627.

[218] Gustavo F. Tondello, Rita Orji, and Lennart E. Nacke. 2017. Recommender systems for personalized gamification. In *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, Inc, New York, NY, USA, (July 2017), 425–430. ISBN: 9781450350679. DOI: 10.1145/3099023.3099114.

[219] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The gamification user types Hexad scale. In *CHI PLAY 2016 - Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, Inc, (October 2016), 229–243. ISBN: 9781450344562. DOI: 10.1145/2967934.2968082.

[220] Luís Torgo, Paula Branco, Rita P. Ribeiro, and Bernhard Pfahringer. 2015. Resampling strategies for regression. *Expert Systems*, 32, 3, (June 2015), 465–476. ISSN: 02664720. DOI: 10.1111/exsy.12081.

[221] Amos Tversky and Itamar Simonson. 1993. Context-dependent preferences. *Management Science*, 39, 10, (October 1993), 1179–1189. ISSN: 00251909. DOI: 10.1287/mnsc.39.10.1179.

[222] Julita Vassileva. 2012. Motivating participation in social computing applications: A user modeling perspective. (April 2012). DOI: 10.1007/s11257-011-9109-5.

[223] Amy Voida, Sheelagh Carpendale, and Saul Greenberg. 2010. The individual and the group in console gaming. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. ACM Press, New York, New York, USA, 371–380. ISBN: 9781605589879. DOI: 10.1145/1718918.1718983.

[224] Richard W. Robins, R. Chris Fraley, and Robert F. Krueger. 2007. *Handbook of Research Methods in Personality Psychology - Google Books*. The Guilford Press.

[225] Gunter Wallner, Simone Kriglstein, and Anders Drachen. 2019. Tweeting your destiny: Profiling users in the twitter landscape around an online game. In *IEEE Conference on Computatonal Intelligence and Games, CIG*. Volume 2019-August. IEEE Computer Society, (August 2019). ISBN: 9781728118840. DOI: 10.1109/CIG.2019.8848079.

[226] Günter Wallner, Christopher Schinnerl, Michael Helfried Schiller, Alexander Monte Calvo, Johanna Pirker, Rafet Sifa, and Anders Drachen. 2019. Beyond the individual: Understanding social structures of an online player matchmaking website. *Entertainment Computing*, 30, (May 2019), 100284. ISSN: 18759521. DOI: 10.1016/j.entcom.2019.01.002.

[227] Hao-Ting Wang, Jonathan Smallwood, Janaina Mourao-Miranda, Cedric Huchuan Xia, Theodore D. Satterthwaite, Danielle S. Bassett, and Danilo Bzdok. 2018. Finding the needle in high-dimensional haystack: A tutorial on canonical correlation analysis. *arXiv*, (December 2018).

[228] Harald Warmelink, Jonna Koivisto, Igor Mayer, Mikko Vesa, and Juho Hamari. 2018. Gamification of the work floor: A literature review of gamifying production and logistics operations. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 1108–1117. ISBN: 9780998133119.

[229] Duncan J. Watts and Peter Sheridan Dodds. 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34, 4, (December 2007), 441–458. ISSN: 0093-5301. DOI: 10.1086/518527.

[230] Jamie Woodcock and Mark R Johnson. 2019. Live Streamers on Twitch.tv as Social Media Influencers: Chances and Challenges for Strategic Communication. *International Journal of Strategic Communication*, 13, 4, (August 2019), 321–335. ISSN: 1553-118X. DOI: 10.1080/1553118X.2019.1630412.

[231] Nannan Xi and Juho Hamari. 2019. Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management*, 46, (June 2019), 210–221. ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2018.12.002.

[232] Rongjing Xiang, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*. ACM Press, New York, New York, USA, 981–990. ISBN: 9781605587998. DOI: 10.1145/1772690.1772790.

[233] Hanting Xie, Sam Devlin, Daniel Kudenko, and Peter Cowling. 2015. Predicting player disengagement and first purchase with event-frequency based data representation. In *2015 IEEE Conference on Computational Intelligence and Games, CIG 2015 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., (November 2015), 230–237. ISBN: 9781479986217. DOI: 10.1109/CIG.2015.7317919.

[234] Donna Xu, Yaxin Shi, Ivor W. Tsang, Yew Soon Ong, Chen Gong, and Xiaobo Shen. 2019. Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 31, 7, (July 2019), 2409–2429. ISSN: 21622388. DOI: 10.1109/TNNLS.2019.2945133.

[235] Yan Xu, Xiang Cao, Abigail Sellen, Ralf Herbrich, and Thore Graepel. 2011. Sociable killers: Understanding social relationships in an online first-person shooter game. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. ACM Press, New York, New York, USA, 197–206. ISBN: 9781450305563. DOI: 10.1145/1958824.1958854.

[236] Georgios N. Yannakakis and Julian Togelius. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing*, 2, 3, (July 2011), 147–161. ISSN: 19493045. DOI: 10.1109/T-AFFC.2011.6.

[237] Nick Yee. 2016. The Gamer Motivation Profile. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery (ACM), New York, NY, USA, (October 2016), 2–2. DOI: 10.1145/2967934.2967937.

[238] An Gie Yong and Sean Pearce. 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9, 2, 79–94.

[239] Shupei Yuan and Chen Lou. 2020. How Social Media Influencers Foster Relationships with Followers: The Roles of Source Credibility and Fairness in

Parasocial Relationship and Product Interest. *Journal of Interactive Advertising*, 20, 2, (May 2020), 133–147. ISSN: 1525-2019. DOI: 10.1080/15252019.2020.1769514.

[240] Ping Zhang. 2008. Motivational affordances: Reasons for ICT design and use. *Communications of the ACM*, 51, 11, (November 2008), 145–147. ISSN: 00010782. DOI: 10.1145/1400214.1400244.

[241] Jichen Zhu and Santiago Ontañón. 2020. Player-Centered AI for Automatic Game Personalization: Open Problems. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, New York, NY, USA, (September 2020), 1–8. ISBN: 9781450388078. DOI: 10.1145/3402942.3402951.

[242] Quan Zou, Sifa Xie, Ziyu Lin, Meihong Wu, and Ying Ju. 2016. Finding the Best Classification Threshold in Imbalanced Classification. *Big Data Research*, 5, (September 2016), 2–8. ISSN: 22145796. DOI: 10.1016/j.bdr.2015.12.001.

Acknowledgements

Undertaking this path was a bet with myself, which I won on so many levels. Three years and a half ago, I thought, "why not trying?", and, now, I cannot imagine my life being different. The doctoral program has been a long journey, dense of all sorts of events, which made me grow immensely, professionally but even more so personally. The autonomy, independence, courage, determination, awareness, and creativity that I developed are something that I would have never achieved otherwise. For that, I will always be thankful to my advisor, Dr. Annapaola Marconi, who gave me this incredible opportunity.

A special thank is due to Prof. Johanna Pirker, who hosted me in her group for a wonderful semester. Your support and advice were priceless. I will be forever grateful to you, Prof. Lennart E. Nacke, and Prof. Anders Drachen. It was (and is) a pleasure and an honor to work with you. Each collaboration was a fundamental building block for my career and development.

I also want to thank all of my collaborators who shared our projects' pains and joys with me. I even thank (the vast majority of) my papers' reviewers. That was sometimes hard to take in but helped me improve incredibly. Yes, even you, R3.

It has not always been all fun and games (did you get it?), though. I experienced failure, disappointment, and anxiety as I never had before. This stronger, empowered version of myself was born only thanks to the beautiful people, friends, new and old, who surround me. First, I have to mention my long-lasting pillars: Alessia and Simone, who, despite being kilometers far away, made me always feel their presence. There is so much to say, but you already know. Then, I want to thank my incredible neighbors Gabriele, Diana, and Syafiq. I will always remember the *Sanba days* with nostalgia and affection: the late nights staying in the living room talking about everything, the unfinished Netflix series, and the last-minute, improvised nights out were what kept me sane. I have to thank Alice, companion of (mis)adventures, and the *Nabucodonosor* group. Finally, a huge thanks to my colleagues and office neighbors: Gianmarco, Francesco, Elia, and Velu, who helped me found my talent as a ping-pong player (I wish). You made coming to the office extra fun.

The biggest thanks are owed to my family, who trusted my decisions despite their concerns. Your support is a certainty but will never be taken for granted. Thank you.

Dulcis in fundo, Michel. You are my rock. This would have never happened without your continuous moral support. What I have and will achieve is also yours.