Michael H. Schiller, BSc

The Influence of Communities on Activity in Video Games

Master's Thesis

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Dipl.-Ing. Dr.techn. Johanna Pirker, BSc

Institute of Interactive Systems and Data Science
Head: Univ.-Prof. Dipl.-Inf. Dr. Stefanie Lindstaedt

Graz, August 2019

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Diplomarbeit identisch.

Datum

Unterschrift

Abstract

With video games becoming a more and more important global phenomenon, they have also evolved into an interesting field of study. Social play, be it cooperative or competitive, is an important aspect of the player experience in multiplayer games. Since social interactions with other players also shape engagement, being able to find others to play with and enjoy the game is of vital importance. To facilitate grouping and matchmaking, on one hand, games often offer in-game features that allow to find mates. On the other hand, if games lack these features, third-party websites have emerged to fill this need. In this thesis, one site that allows for social matchmaking, `the100.io` is analyzed. While the site supports different video game titles, in this thesis the focus lies on *Destiny*. The dataset combines information about demographic data, user preferences, in-game features, as well as platform-related information. A social network is constructed from friendships formed on `the100.io` and subsequently analyzed. One such analysis is community detection. Understanding how social connections form and how these relationships can foster in-game activity offers insights for building and maintaining a healthy player base which, in turn, can improve both player retention and player engagement. Correlations between preferences, network properties and in-game performance measures are presented and the impact of these metrics on activity is assessed. Furthermore, archetypal analysis is applied to both players and groups in order to identify patterns of behavior. The results show communities forming clearly around platforms. Furthermore, players are best partitioned into five distinct archetypes, while groups form four clusters. Lastly, findings indicate that group activity is strongly impacted by the group's size, as well as by the number and social connectivity of moderators.

Kurzfassung

Da Videospiele zu einem immer wichtigeren globalen Phänomen werden, haben sie sich auch zu einem interessanten Forschungsgebiet entwickelt. Soziales Spielen, sei es kooperativ oder in Konkurrenz, ist ein wichtiger Aspekt des Spielerlebnisses in Multiplayer-Spielen. Da soziale Interaktionen mit anderen Spielern das langfristige Spielerlebnis prägen, ist es von großer Bedeutung, Mitspieler zu finden, mit denen man spielen und das Spiel genießen kann. Um die Gruppierung und das Matchmaking zu erleichtern, bieten Spiele einerseits oft Funktionen im Spiel, die es ermöglichen, Partner zu finden. Andererseits haben auch Drittanbieter Websites entwickelt um diesen Bedarf zu befriedigen, falls Spielen diese Funktionalität fehlt. In dieser Arbeit wird eine Website, die soziales Matchmaking ermöglicht, `the100.io` analysiert. Während diese Seite verschiedene Videospieltitel unterstützt, liegt der Fokus in dieser Arbeit auf dem Spiel *Destiny*. Der Datensatz kombiniert Informationen über demographische Daten, Benutzereinstellungen, In-Game-Features sowie plattformbezogene Informationen. Ein soziales Netzwerk aus Freundschaften, die auf `the100.io` geschlossen wurden, wird aufgebaut und anschließend analysiert. Eine dieser Analysen ist die Erkennung von Communities. Das Verständnis, wie soziale Verbindungen entstehen und wie diese Beziehungen die Aktivitäten im Spiel fördern können, bietet Erkenntnisse für den Aufbau und die Erhaltung einer gesunden Spielerbasis, die wiederum sowohl die Spielerbindung als auch das Spielerlebnis der Spieler verbessern können. Es werden Zusammenhänge zwischen Präferenzen, Netzwerkeigenschaften und Leistungskennzahlen im Spiel dargestellt und die Auswirkungen dieser Kennzahlen auf die Aktivität abgeschätzt. Darüber hinaus wird Archetypal Analysis sowohl auf Spieler als auch auf Gruppen angewendet, um Verhaltensmuster zu identifizieren. Die Ergebnisse zeigen, dass sich Communities deutlich um Konsolen-Plattformen herum bilden. Darüber hinaus lassen sich fünf Archetypen für Spieler identifizieren, während Gruppen vier unterschiedliche Cluster

bilden. Schließlich deuten die Ergebnisse darauf hin, dass die Gruppenaktivität stark von der Größe der Gruppe sowie von der Anzahl und der sozialen Vernetzung der Gruppen-Moderatoren beeinflusst wird.

Acknowledgements

First of all, I would like to thank my thesis supervisor Dipl.-Ing. Dr.techn. Johanna Pirker, BSc of the Institute of Interactive Systems and Data Science at Graz University of Technology. She consistently allowed this thesis to be my own work but also helped by steering me into the right direction whenever she thought I needed it.

I would also like to thank Prof. Anders Drachen, Ph.D. of DC Labs at the University of York and Assistant Prof. Günter Wallner of the Department of Industrial Design at Eindhoven University of Technology. Without their tremendous help and input, the articles written over the course of this thesis would not have been possible.

Furthermore, I would like to extend my appreciation to my study colleagues, Philipp Hafner, Michael Holly, and Lukas Schabler who accompanied me along the entirety of my studies, helped with team projects and gave valuable input for writing this thesis.

Finally, I must express my very profound gratitude to my parents and to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Author

Michael Schiller

Contents

Abstract **iii**

1. Introduction **1**
1.1. Motivation 1
1.2. Goals and Objectives 2
1.3. Methodology and Structure 2

2. Background & Related Work **5**
2.1. Retention, Performance, and Experience 6
2.2. Understanding Player Behavior 7
2.2.1. Psychology and Personality Types 7
2.2.2. From Psychology to Player Behavior 14
2.2.3. Motivation of Play 15
2.2.4. Psychological Analysis and Motivation Research in Games 20
2.2.5. Comparing Analyses 23
2.3. The Social Side of Gaming 25
2.3.1. Graph Theory 25
2.3.2. Social Network Analysis (SNA) 29
2.3.3. Applications of Social Network Analysis in Games . . 31
2.3.4. Comparing Social Analyses 37
2.4. Behavioral Analysis and Clustering 39
2.4.1. *k*-means Clustering 39
2.4.2. Archetypal Analysis (AA) 41
2.4.3. Video Games and Behavioral Profiling 45
2.4.4. Comparing Algorithms 52
2.5. Summary 54

3. Destiny - Gameplay **56**
3.1. Setting 57
3.2. Class System 58
3.3. Expansions 59
3.4. Game Modes 60
3.4.1. Player-versus-Environment (PvE) 60
3.4.2. Player-versus-Player (PvP) 62
3.5. Summary 64

4. Datasets **65**
4.1. Dataset "the100.io" 65
4.1.1. Overview 66
4.1.2. Feature Description 66
4.2. Dataset "Destiny" 69
4.2.1. Overview 70
4.2.2. Feature Description 71
4.3. Summary 73

5. Data Preparation and Processing **74**
5.1. Overview 74
5.2. Processing the Destiny Dataset 76
5.2.1. Phase 1: Transform 76
5.2.2. Phase 2: Filter 77
5.2.3. Phase 3: Enhance 80
5.2.4. Resulting Data Model 83
5.3. Processing the the100.io Dataset 85
5.3.1. Data Model 85
5.3.2. Phase 2: Filter 86
5.3.3. Phase 3: Enhance 87
5.4. Summary 90

6. Analysis and Results **92**
6.1. General Analysis 92
6.1.1. Overview 92
6.1.2. Players 100
6.1.3. Groups 112

6.2. Archetypes 118
6.2.1. Players 119
6.2.2. Groups 125
6.3. Discussion 132
6.4. Limitations 135
6.5. Summary 136

7. Conclusion and Outlook **139**
7.1. Conclusion 139
7.2. Future Work 140

Appendix A. Bartle Test **148**

Bibliography **153**

Ludography **164**

1. Introduction

With video games becoming more and more popular, they have also become an interesting field of study – both for academic and financial reasons as they can help developers in designing games that exhibit high levels of player retention. This thesis aims at helping to understand player activity while taking different metrics and social structures into account. Specifically, the impact of various measures on activity is examined.

1.1. Motivation

It has been shown that player experience and engagement in multiplayer games are impacted by social interactions with other players (Yee, 2006; Gajadhar, de Kort, & IJsselsteijn, 2008). In this regard, social relationships are key factors for retention and monetization in games (Alsén, Runge, Drachen, & Klapper, 2016; Rattinger, Wallner, Drachen, Pirker, & Sifa, 2016). Being able to find, team up with or play against similarly skilled players has been identified as an important factor for keeping players engaged, thus, facilitating a healthy player base within multiplayer-online, and massively multiplayer online games (Ducheneaut, Yee, Nickell, & Moore, 2006; Seif El-Nasr, Drachen, & Canossa, 2013; Alsén et al., 2016; Rattinger et al., 2016). While some games offer these means of matchmaking and group building within the game itself, others rely on the distribution platform, for example, *Steam* (Valve Corporation, 2003), or on other third parties entirely to allow players to form groups besides ad-hoc matchmaking. Thus, understanding player behavior forms an important aspect in building a good long-term game experience for players, as well as for creating a financially viable game. Analyzing players' behavior has been shown to be useful, for example, when battling toxicity (Maher, 2016) or when identifying illicit behavior (Keegan, Ahmed, Williams, Srivastava, & Contractor, 2010). Social networks formed by players can be analyzed using the techniques from social network analysis (SNA) as shown by Ducheneaut and Moore (2004), Stafford, Luong, Gauch, Gauch, and Eno (2012), Seif El-Nasr et al. (2013), Iosup, Van De Bovenkamp, Shen, Jia, and Kuipers (2014), Jia et al. (2015), Rattinger et al. (2016). In this application, social networks are used to model competitive and cooperative in-game interactions. In this regard, Pirker, Rattinger, Drachen, and Sifa (2018) analyzed social networks built from implicit social connections, for example, match co-occurrences, that is, players that played with one another were connected, as well as from explicit social connections, such as, clan membership. From the applications so far, we can see that in most cases in-game measures were analyzed.

1.2. Goals and Objectives

This thesis, aims at combining in-game measures with demographic data as is available from `the100.io` to build a more comprehensive understanding of player behavior in general and to identify the influencing factors for activity, specifically. In this regard, the main research questions can be stated as follows.

RQ1: Are there noticeable differences in behavior between players/groups?
RQ2: Which metrics impact players' activity, that is, player retention?
RQ3: Can a model for predicting activity be derived from the data?

1.3. Methodology and Structure

In order to answer these questions, SNA is used on a friendship network of the looking for group (LFG) website, `the100.io` which is an external tool helping players find other players with whom to play. Besides in-game features, the dataset contains demographic information, as well as activity measures calculated by `the100.io`. Furthermore, archetypal analysis (AA) is applied to both players and groups in order to find clusters of similar players and groups, respectively. Additionally, correlations between social measures, in-game measures, and activity are investigated.

The structure of this thesis is shown in Figure 1.1 with the different topics covered in the various chapters shown as surrounding hexagons. Chapter 2 investigates the theoretical foundation of the thesis, including the definitions for *player retention*, *player performance* and *player experience*, as well as different frameworks for understanding behavior, both in the field of psychology and with a focus on video games. As a next step, important concepts in the field of SNA are introduced. Furthermore, this chapter also presents clustering algorithms as they are used to find patterns of player behavior. Finally, this chapter contains details for state-of-the-art research and introduces other studies performed on video games. Next, Chapter 3 tries to give a brief overview of the game analyzed, *Destiny*[1]. In this regard, features of the game, its setting and available game modes are discussed. In the following chapter, Chapter 4, the two datasets used, one extracted from the Bungie.NET application programming interface (API)[2] and the other extracted from `the100.io`[3], are described with the features they contain and descriptive statistics. In Chapter 5 the steps necessary to work with the datasets as well as augmentation performed on the data is described in more detail. Chapter 6 is the penultimate chapter and describes the analyses performed and discusses the results. Here, it has to be noted that some of the analyses presented in this thesis have already been published in Schiller et al. (2018) and Wallner, Schinnerl, Schiller, Pirker, and Drachen (2019). Additionally, this chapter contains a discussion on limitations that have to be taken into consideration when evaluating the results. Finally, Chapter 7 shows conclusions which can be drawn from the results and discussed them. Furthermore, this chapter also gives examples for future research.

[1] Bungie, 2014. https://www.destinythegame.com/d1.
[2] https://bungie-net.github.io/multi/index.html
[3] https://the100.io

Figure 1.1.: Thesis Structure

2. Background & Related Work

Today, more than 2.5 billion people play video games all around the world. The global games market is generating over $150 billion in annual revenue and is still growing (Wijman, 2019). This amounts to the video games industry generating more than twice the revenue of the film industry, $38.6bn, (*Theatrical Market Statistics 2016*, 2017) and the music industry,$19.1bn, (*Global Music Report 2019*, 2019), combined. With video games being such an important global phenomenon, they are also of academic interest and a field of research worth studying. At the time of writing the five most popular PC games are (Newzoo, 2019):

1. *League of Legends*[1]
2. *Minecraft*[2]
3. *Hearthstone: Heroes of Warcraft*[3]
4. *Counter-Strike: Global Offensive*[4]
5. *Fortnite*[5]

with four out of these five being focused heavily on the player-versus-player (PvP) game mode – having a pronounced focus on competitive gameplay. The remaining game, *Minecraft*, which emphasizes player-versus-environment (PvE) gameplay also offers a collaborative multiplayer game mode. From this list alone, it is already obvious that the social side of gaming is an important avenue of research. In general, it is important for game developers to understand how players play games. Understand players'

[1] Riot Games, 2009. https://www.leagueoflegends.com.
[2] Mojang, 2011. https://www.minecraft.net.
[3] Blizzard Entertainment, 2014. https://playhearthstone.com.
[4] Valve Corporation, 2012. https://store.steampowered.com/app/730/CounterStrike_Global_Offensive.
[5] Epic Games, 2017. https://www.epicgames.com/fortnite.

needs and wants can help with increasing *player retention*, *player performance* and improving *player experience* which will be described shortly.

2.1. Retention, Performance, and Experience

Player retention describes how long a player engages with a game over the course of the player's entire gaming history. Therefore, it can be thought of as a measure of involvement with and commitment to the game. In a time where free-to-play (F2P) games become more and more relevant, retention is also an important aspect for monetization – to understand and improve *player retention*, to give the right incentives at the right time to keep players engaged (Weber, Mateas, & Jhala, 2011; Drachen, Lunquist, et al., 2016).

Player performance is linked to a player's enjoyment of a game. Video games as a medium are interactive, hence players will try to engage with the game to overcome challenges when presented with adversity. These confrontations and their resolution have a psychological impact on players' emotional state. Successfully surmounting obstacles will lead to a positive, euphoric state of mind while being overwhelmed can lead to negative feelings, such as anger or frustration. In this regard, self-determination, that is being in control of your own success is an important aspect in competitive gameplay (Vorderer, Hartmann, & Klimmt, 2003; Ryan, Rigby, & Przybylski, 2006; Klimmt, Blake, Hefner, Vorderer, & Roth, 2009).

Player experience has been described as the aggregate of interactions between players and the game they are playing. Player experience encompasses concepts, such as "Flow", immersion, fun, excitement, challenge and boredom. *Immersion*, for example, may be partitioned further into sensory, imaginative and challenge-based immersion (Ermi & Mäyrä, 2005). Sensory immersion then describes the audio-visual aspects of videos games, while imaginative immersion corresponds to atmosphere, involvement in the narrative and identification with characters. Challenge-based immersion is closely connected to *player performance* as it describes the gameplay parts that correspond to physical or mental challenges. Immersion can be seen as a necessary precondition to "Flow" as described by Csikszentmihalyi (2008). This state of mind describes complete involvement in the game and entails a loss of sense of time and context. Csikszentmihalyi also describes people in this state as *autotelic*, self–sufficient. They are driven by intrinsic motivation and feel rewarded by the activity itself (Nacke & Lindley, 2008; Nacke et al., 2009).

2.2. Understanding Player Behavior

When looking at the wide variety of games offered and the diversity in video game genres it is obvious that not all players are the same. They play various games for various reasons, are driven by differing motivations and find a wide range of games enjoyable. However, it can be conducive to understand these preferences, differences and varying motivations not at an individual level but rather try to find similarities between individual players. The different ways of describing similarities and various groupings are discussed in the following section.

2.2.1. Psychology and Personality Types

As an intuitive first step towards understanding player behavior it is helpful to consider concepts from psychology that are used to describe personality and character traits. This can help in identifying commonalities in players allowing them to be grouped and in trying to derive in-game mechanics that cater to the needs and wants the players of a game might have. A wide variety of theories describing personality types have been devised, some of which are described in this next section.

Myers-Briggs Type Indicator (MBTI)

In an attempt to formally describe personality traits, Briggs Myers and Briggs Myers (1995) introduced the concept of the Myers-Briggs type indicator (MBTI) which describes preferences forming peoples' personalities and, thus, emphasize the differences between Sensing (S) and Intuition (N), between Thinking (T) and Feeling (F), between Introversion (I) and Extraversion (E), as well as between Judgment (J) and Perception (P) and order these preferences as *(E/I)(S/N)(T/F)(J/P)* to form the 16 types, identified by their respective four-letter abbreviations, as seen in Figure 2.1. The authors described the four different dimensions they considered relevant for personality, in the following way (Briggs Myers & Briggs Myers, 1995):

- **EI Preference**: *Introversion (I)* describes personality types which focus on the "inner world", that is, ideas and concepts, while *Extraversion (E)* emphasizes the importance of the "outer world", that is, people and things. Introverts are directed by inner drives and do not rely on external encouragement while extraverts' motivation often stems from their surroundings. Moreover, extraverts are more articulate and, according to the authors, outnumber introverts approximately three to one. Introverts often fail to see the outer world because they are consumed by their ideas, while extraverts do not tend to understand abstract principles but rather need to see concrete examples to comprehend.
- **Two Ways of Perceiving – SN Preference**: *Sensing (S)* describes the perception based on the five senses, while *Intuition (N)* focuses on unconscious processes for perceiving a person's surroundings.

  > *"Anyone preferring sensing to intuition is interested primarily in actualities; anyone preferring intuition to sensing is mainly interested in possibilities."*
  >
  > (Briggs Myers and Briggs Myers, 1995, chap. 5)

  Intuitive people tend to be inventors and innovators and seek out situations where they find inspiration. They tend to score higher on IQ (intelligence quotient) tests than people of the sensing type. This may be caused by the fact that they quickly answer the given questions on a "hunch" or using "a woman's intuition" while sensing people are likely to read the question multiple times and try to gather all the facts and details before arriving at a conclusion. Sensing types do not trust answers that suddenly appear and are more careful than intuitive people.
- **Two Ways of Judging – TF Preference**: *Thinking (T)* identifies the process of judging using facts and logic, while *Feeling (F)* describes reaching judgments by applying subjective values to arrive at a conclusion.

The two ways of judging, *thinking* and *feeling* are the two mechanisms used in decision-making. The perspectives can be thought of evaluating “true – false” (thinking) and “agreeable – disagreeable” (feeling), respectively. The *thinking* position is impersonal and therefore works well as long as no people are involved, while the *feeling* angle is sympathetic and works well when personal values are concerned. Briggs Myers and Briggs Myers note that this dimension also shows a notable sex difference: the percentage of *feeling* types is considerably higher in the female population and feeling people tend to be more social and tactful than their thinking counterparts.

- **JP Preference**: *Judgment (J)* describes a personality preferring judgment (cf. TF Preference) – a person who wants to live an orderly life – over *Perception (P)* (cf. SN Preference) – a person who lives in the moment. The JP preference is described as an axis that has to be balanced between the two extremes *judgment* and *perception*. Judgmental personalities try to arrive at final conclusions, while perceptive types are averse to conclusions. Perceptive types aspire to considering any and all aspects of a situation, being convinced that a problem can be solved by understanding it better. Perception without judgment leads to drifting aimlessly without ever finding meaning and purpose, whereas judgment without perception leads to a rigid and narrow worldview without the possibility of considering others’ points of view. The prime characteristic is then *prejudice* – a quick judgment without proper perception of and consideration for all factors involved.

In this section, the psychological concept of the MBTI was introduced and its four dimensions, Introversion – Extraversion, Sensing – Intuition, Thinking – Feeling, and Judgment – Perception were described. In the next section, a related concept, Keirsey Temperaments, will be introduced.

Keirsey Temperaments

Keirsey (1998) further developed the works of Briggs Myers and Briggs Myers to derive four temperaments. The author chose to analyze personality along two very different axes than were analyzed in the MBTI prior. The relationship between and close coupling of MBTI and Keirsey temperaments can be seen in Figure 2.1.

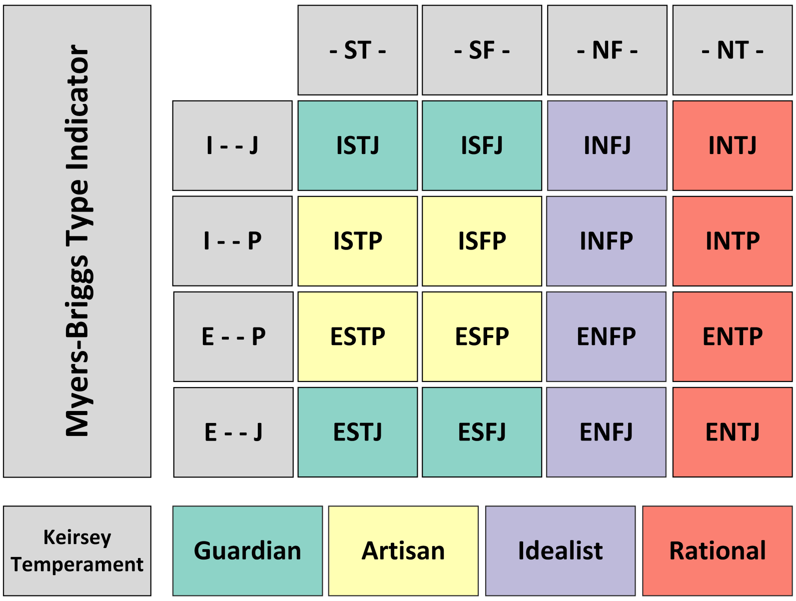


Figure 2.1.: Congruities between Myers-Briggs Type Indicators (adapted from Briggs Myers and Briggs Myers, 1995, chap. 3) and Keirsey Temperaments

Keirsey described these two dimensions as being vital in separating man from animal: *words* and *tools*. The author visualized them as a matrix as seen in Figure 2.2. The dimension of words is being divided into *abstract* and *concrete* word usage, with the former relating to the use of symbols, fiction, analogies and so forth, and the latter describing detailed speech, specific terms and facts. The second dimension, *tools*, relates to humans, as more than "word-using animals". Humans developed to shape the world according to their needs in order to survive in it. This dimension is categorized into *cooperative* tool usage, that is limiting the usage to agreed-upon rules in order to maintain relationships to others, while *utilitarians* do not tend to restrain their usage of tools to societal norms, rules or morals when trying to achieve their goals – following the motto "the end justifies the means".

The four personality types following from this categorization are given (with their corresponding MBTI preferences in parentheses) as (Keirsey, 1998):

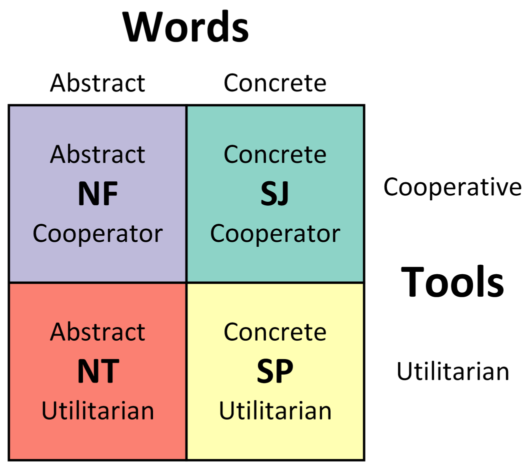


Figure 2.2.: Keirsey Temperament Matrix (adapted from Keirsey, 1998, chap. 2)

Artisans (SP) are *concrete utilitarians*. They use concrete words, as well as gestures to convey their messages, and do not feel bound by protocols or norms when pursuing their goals. Noteworthy representatives of this temperament are Lord Byron, Winston Churchill and Ronald Reagan.

Guardians (SJ) are *concrete cooperators*. People belonging to this temperamental type, like artisans, use concrete information to communicate, but other than that they differ in being cooperative instead of utilitarian. They feel obliged to follow societal norms and morals and strive for responsibility of command – such positions as, military officers and CEOs. Famous examples of the SJ temperamental type include George Washington and Harry S. Truman.

Idealists (NF) are *abstract cooperators*. Rather than speaking about the world surrounding them, idealists tend to speak about feelings, tend to exaggerate and use metaphors. Like guardians, idealists tend to cooperate with others but other than the SJ temperament, idealists do not only want to comply with society but rather work on building consensus. Examples of this temperament can be found in the personalities of Mahatma Ghandi and the writer Emily Dickinson.

Rationals (NT) are *abstract utilitarians*. They speak about concepts rather than observable facts and tend to build an unambiguous and technical vocabulary. Like artisans, they are utilitarian in striving for their goals but differ from those in striving for efficiency rather than effectiveness. NT temperamental types do not aim for societal status, degrees or prestige and are more goal-oriented. Notable representatives of this temperament include Napoleon, Nikola Tesla and Howard Hughes.

After inspecting the MBTI and its derivative, Keirsey temperaments, the final classification for personality traits discussed here, will be the *Big Five personality traits* as it is a commonly used framework with a wide range of studies supporting its usefulness.

Big Five (OCEAN)

The *Big Five* personality traits, or OCEAN (following from the combination of the first letter of each dimension), are the result of decades of researching personality traits and are derived from lexical analysis of adjectives describing personality attributes. Then a factor analysis was applied yielding five distinct clusters and – depending on the personality lexicon used for analysis – varying numbers of sub-factors (Goldberg, 1990). It is important to note that the *Big Five* are dimensional as opposed to categorical (cf. Keirsey Temperaments). In studies around the world this five-factor model has been shown to be highly consistent and reliable to retesting. While there are different measures for the Big Five with the dimension of *Neuroticism* sometimes being inverted to reflect *Emotional Stability*, we will focus on the personality dimensions and their facets, that is, unique aspects of their related traits as described in the Revised NEO Personality Inventory (NEO PI-R) as described by Costa and McCrae (1992):

Openness to experience contains the facets *Aesthetics, Values, Feelings, Actions, Ideas* and *Fantasy*. The trait *openness* has been found to relate to creativity in both artistic and scientific fields, as well as to a tendency to favor social and political liberalism over conservatism.

Conscientiousness encompasses the facets *Order, Dutifulness, Striving for Achievement, Deliberation, Competence* and *Self-Discipline*. This trait dimension can be seen as describing reliability and (work) ethic. Taken to the lower end of the scale, *conscientiousness* predicts criminal and anti-social behavior, while on the other extreme, it predicts perfectionism and compulsive behavior, such as seen in "workaholics". *Conscientiousness* combined with lower scores in *openness* is a predictor for political conservatism.

Extraversion consists of the facets *Activity, Excitement Seeking, Positive Emotion, Warmth, Assertiveness* and *Gregariousness*. Similarl to the MBTI (Briggs Myers & Briggs Myers, 1995), extraversion in the framework of the *Big Five* describes the character attribute of seeking reward outside oneself. Extraverts more likely enjoy social gatherings and tend to prefer time spent around others over time spent alone.

Agreeableness involves the facets *Straightforwardness, Altruism, Compliance, Tender-mindedness, Trust* and *Modesty*. This trait dimension can be seen as describing the tendency of pursuing social harmony and cooperation over favoring selfish behavior. This trait also relates to trust in and honesty in communicating with others, as well as response to conflict: deferral and cooperation versus vindictiveness.

Neuroticism consists of the facets *Hostility, Anxiety, Self-Consciousness, Vulnerability to Stress, Impulsiveness* and *Depression*. This dimension can described as emotional instability, lack of self-control or ineptitude of handling psychological stress.

Due to the five-factor model being grounded in a lexical analysis of descriptions of personality, it has since been argued that the *Big Five* personality traits can account for and model further trait adjectives in some way:

> *"[I]t now seems reasonable to conclude that analyses of any reasonably large sample of English trait adjectives in either self- or peer descriptions will elicit a variant of the Big-Five factor structure, and therefore that virtually all such terms can be represented within this model. In other words, trait adjectives can be viewed as blends of five major features, features that relate in a gross way to Power, Love, Work, Affect, and Intellect."* (Goldberg, 1990)

Several measures of the *Big Five* personality traits exist, including the aforementioned NEO PI-R (Costa & McCrae, 1992), the International Personality Item Pool (IPIP) (Goldberg, 1999; Goldberg et al., 2006) the short-hand rating techniques of the *Five Item Personality Inventory* and the *Ten-Item Personality Inventory (TIPI)* (Gosling, Rentfrow, & Swann, 2003) and relative-scored *Big*

Five measures (Hirsh & Peterson, 2008) that have been shown to be useful for suppressing deceptive responses.

2.2.2. From Psychology to Player Behavior

After the more general introduction to character types as described in psychology, it is useful to understand how players displaying different character traits interact with games in different ways. These groups of players are driven by various motivations and expect and want different aspects and gameplay elements from games. As a first step, we will take a look at applications of the *Big Five* personality traits as they relate to video games. In a longitudinal study, Witt, Massman, and Jackson (2011) analyzed the impact of the *Big Five*, as well as socioeconomic variables on computer usage overall, and on video game consumption in particular. The authors analyzed 11 to 16-year olds over a three-year time span and found that *openness* is positively related to video game usage, *openness* and higher self-esteem positively predict general computer usage, and the trait *extraversion* positively predicts the usage of communication technology. Finally, in their longitudinal analysis, the authors further noted that overall computer usage tends to decrease as the participants got older. Braun, Stopfer, Müller, Beutel, and Egloff (2016) investigated the relationship between the *Big Five*, computer usage and video game use up to an unhealthy level, corresponding to internet gaming disorder (IGD) – as described in the DSM-V (American Psychiatric Association [APA], 2013). Braun et al. first divided their sample into three groups: non-gamers, gamers and gaming addicts. The authors observed that video game use is positively correlated with trait *neuroticism* and negatively correlated with *extraversion* and *conscientiousness*. When comparing group differences, the authors observed that gaming addicts scored significantly lower on trait *extraversion* than both gamers and non-gamers and *conscientiousness* is highest in non-gamers, lower in gamers, and lower still in gaming addicts. Lastly, the authors observed that players preferring action games scored highest in *extraversion*, participants playing role-playing games (RPGs) than other genres scored higher in *openness* and, finally, a preference for simulation games predicts a high score inthe trait *conscientiousness*. In another study investigating the *Big Five* and their impact on genre preferences, Peever, Johnson, and Gardner (2012) used the aforementioned TIPI to determine the five factors and combined the results with a questionnaire about video game genres. The authors found the following correlations:

- *Openness* significantly positively correlates with action adventure and platformer games.
- *Conscientiousness* is significantly positively correlated with sport, racing, simulation and fighting games, while there seems to be no relationship with puzzle or educational games.
- *Extraversion* significantly positively correlates with party, casual and music games.
- *Conscientiousness* is slightly negatively correlated with RPGs of all sorts (massively multiplayer online role-playing game (MMORPG), action role-playing, etc.), as well as with turn-based and real-time strategy games.

These discoveries – at least for the trait *conscientiousness* – are in line with the findings of Braun et al. (2016).

2.2.3. Motivation of Play

One of the first researchers to analyze and classify different motivations of play was Bartle (1996). According to the author, players' interests in a multi-user dungeon – or, more general, in a multiplayer game – may be analyzed along two axes: one axis represents the focus on *players* on one side of the spectrum, and the focus on the surrounding *world* on the other, while the other axis reflects the importance of *acting* on one side and the importance of *interacting* on the other. This relationship can also be seen in Figure 2.3. Bartle (1996) conjectured that while players' interests lie on a spectrum, there are four extremes which, associated with suits of cards, are as follows:

- **Achievers** (♦) are *"always seeking treasure"* and consequently they are Diamonds. Achievers prefer to *act* on the *game world* itself, score points and level up their characters. Besides comparing their scores achievers do not take fellow players into consideration much.

- **Explorers** (♠) *"dig around for information"* and thus they are associated with Spades. Explorers *interact* with the *game world*, want to test how a game works and try to go beyond the limits of what the programmers thought would be possible. Apart from talking to other explorers about new interactions with the game that they could try, they also do not socialize much.
- **Socialisers** (♥) are empathic towards other players. Hence, they are Hearts. Socialisers like to *interact* with other *players*. They chat with them and build lasting relationships while the game itself serves as a conversation starter or backdrop.
- **Killers** (♣) are Clubs as *"they hit people with them"*. Killers *act* on other *players*. They force themselves onto other players and wish to cause mayhem by killing unsuspecting players. They usually only ever socialize with other killers to discuss new methods of killing people or with their victims to mock them.

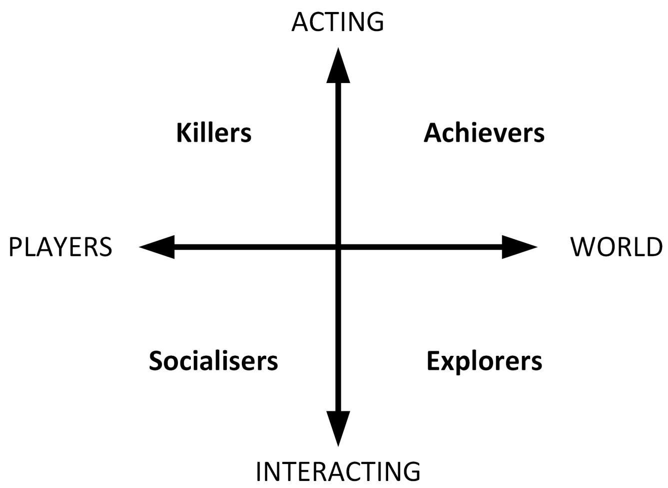


Figure 2.3.: Interest Graph (adapted from Bartle, 1996)

Building on the works of Bartle (1996), a test for players' personalities was devised by Andreasen and Downey (2001, 1). The test, now known as the "Bartle Test", is composed of an equal number of questions for each combination of play styles. The test contains a set of binary questions, each of which tests one play style against another. For example, the styles *achiever* and *socialiser* may be tested against each other by asking[6]:

> *Would you rather be:*
>
> - *Popular*
> - *Wealthy*

After finishing the test, a quotient is calculated that describes the strength of the association with each player type. While the quotient for each play style may not exceed 100 %, the total sum of all calculation quotients may or may not exactly add up to 200%[7].

Kennerly (2004) built on the results obtained through the process of "Bartle testing" described above. The author asked whether the play styles identified by these tests were at all related to the psychological concepts introduced earlier. Kennerly examined the possible connection between the MBTI (Briggs Myers & Briggs Myers, 1995) and the "Keirsey temperaments" (Keirsey, 1998). Analyzing correspondences between MBTI and *Bartle's Player Types*

MBTI	**Keirsey**	**Total**	♦	♣	♠	♥
NT	Rational	45%	47%	51%	50%	35%
NF	Idealist	34%	27%	24%	31%	46%
SJ	Guardian	13%	17%	15%	12%	12%
SP	Artisan	8%	9%	10	7%	8%

Table 2.1.: MBTI - *Bartle's Player Types* Overlap, ♦... Achiever, ♣... Killer, ♠... Explorer, ♥... Socialiser (adapted from Kennerly, 2004)

(seen in Table 2.1), the author finds isomorphism between the two models as seen in Table 2.2 with the exception of the temperament type *Rational* which translates to both player types *Killer* and *Explorer*. Finally, the author lists possible explanations for the NT temperament type mismatch, such as an

[6]Question taken from the original question set found at http://www.andreasen.org/bartle. For the full set of questions also see Appendix A.

[7]For further results, see Andreasen and Downey (2001, 1), https://web.archive.org/web/20000818064001/http://www.andreasen.org/bartle/stats.cgi, Archived version from August, 18th 2000

unexpectedly high fraction of *Rational*s and the possibility of those players trying to exercise power over other players or players showing different character traits in the offline and online world.

MBTI	Keirsey	Bartle
NT	Rational	Killer (1.12), Explorer (1.09)
NF	Idealist	Socialiser (1.36)
SJ	Guardian	Achiever (1.31)
SP	Artisan	Killer (1.29)

Table 2.2.: MBTI - *Bartle's Player Types* Correspondences (adapted from Kennerly, 2004)

In this section until now, player behavior was mostly described as being separable into distinct and distinguishable, that is, exclusive quadrants. However, more recent research points to this not being the case, and indicates that overall player behavior is best described as three overlapping clusters with players belonging to each cluster to some extent. One possible explanation for the divergence between MBTIs, Keirsey temperaments and player behavior, mentioned earlier, was given by Yee (2006). The author tried to empirically test and validate *Bartle's Player Types* using a questionnaire based on the *Bartle Test* (Andreasen & Downey, 2001, 1) with roughly 3,000 respondents across multiple different MMORPGs. To achieve this, a two-stage principal component analysis (PCA) was conducted yielding ten and three principal components in a first and second round, respectively. The components that emerged during the first step were grouped together, matching the groupings in the second step, and can be explained using the Gaming Motivation Scale (GMS) as seen in Table 2.3. Furthermore, the author inspected correlations between demographic variables and cluster membership and found that gender can explain association with the achievement component, that is, male players are more likely to have a higher *achievement* association. While both males and females have a similar association with the sub-component *socializing*, female players are more likely to also be associated with the sub-component *relationship*, while male players are more likely to associate with the *teamwork* sub-component. Lastly, the author notes that problematic usage of video games is best predicted by the sub-component *escapism* and the feature *hours played per week*.

Achievement	Social	Immersion
Advancement	Socializing	Discovery
Mechanics	Relationship	Role-Playing
Competition	Teamwork	Customization
		Escapism

Table 2.3.: Bartle's Player Types - Factor Analysis: Components (adapted from Yee, 2006)

Another framework used for helping to try to understand player motivation is the *player experience of need satisfaction (PENS)*, a framework based on self-determination theory (Ryan et al., 2006; Rigby & Ryan, 2007). Self-determination theory was introduced by Zuckerman, Porac, Lathin, and Deci (1978) and aids understanding human behavior by investigating the extent to which motivation is driven by internal and external factors. Besides the three factors identified by self-determination theory, *competence*, *autonomy* and *relatedness*, PENS further adds *presence* as a scale to the factors playing a role in player motivation. *Competence* can be defined as the need to be able to control the outcomes by improving oneself and honing one's skills. *Autonomy* describes the need to be able to choose and influence one's own journey while *relatedness* can be seen as the desire to interact with others, care for and share experiences with them. Lastly, *presence* as introduced by PENS expresses the need to feel immersed in the game by means of, for example, compelling stories, authentic characters and environments that feel real. In this regard, Johnson and Gardner (2010) explored connections between personality – as expressed by the *Big Five* personality traits – and player motivation, as measured by using the PENS. The authors compared scores between those two measures and found significant positive correlation between *agreeableness* and *competence*, a significant negative correlation between *emotional stability* and *presence*, and, furthermore, a positive correlation between *openness* and *autonomy*.

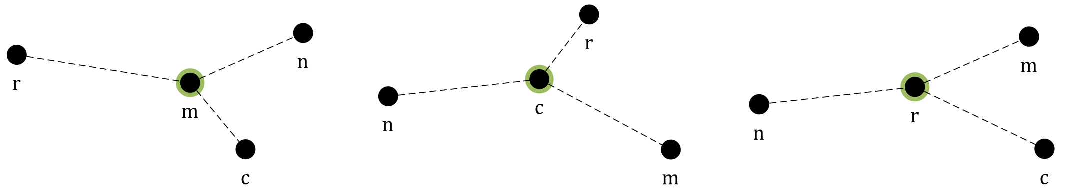
2.2.4. Psychological Analysis and Motivation Research in Games

In this next section, studies using the theories and frameworks introduced in this chapter will be discussed. Owing to the focus of this thesis, the works described here will revolve around research into games and related studies.

Analysis in "The ICE"

Further corroboration for the GMS (Yee, 2006) was given by Thawonmas and Iizuka (2008) who used *KeyGraph*, an approach originally designed for automatic keyword extraction and introduced by Ohsawa, Benson, and Yachida (1998). In their setup, Thawonmas and Iizuka used a rich set of player logs of a PvE game called *The ICE* to extract series of single actions, such as, *chat (with players)*, *trade*, *attack*, *talk (to non-player characters (NPCs)*. The authors noted, that due to the placement of NPCs and monsters relevant for quests, the action *walk* was removed from consideration as it is necessarily part of the log of all play types. This is due to the fact that progressing through the game requires players to walk around in order to take on missions, deliver items and so forth. Some major missions that players have to complete in this game, include delivering items from one NPC to another, hunting monsters for an NPC and trading with NPCs to increase an amount of currency granted by another NPC. In order to extract the necessary co-occurrences, the authors preprocessed lines in the log files, each line representing an action, to generate time-series from consecutive actions. The authors then applied *KeyGraph* to extract the keywords, that is, the most important actions when distinguishing play styles. The results of this clustering are shown in Figure 2.4.

Players in the three clusters under consideration, that is, achievement, social and immersion (see Figure 2.4a, Figure 2.4b, Figure 2.4c, respectively) differ in their key actions. The key actions for achievers, socialisers and explorers are *interaction with a mission master* (m), *chat* (c) and *interaction with a remote object* (r), respectively.



(a) Achievers (b) Socialisers (c) Explorers

Figure 2.4.: Action-based KeyGraphs for different player types (adapted from Thawonmas and Iizuka, 2008). Actions: interaction with a (m)ission master; (c)hat; interaction with a (r)emote object; interaction with a (n)earby object.

Analysis in "EverQuest II"

In an analysis around the game *EverQuest II*[8], Williams, Yee, and Caplan (2008) combined GMS, demographic, and health measures, as well as a questionnaire regarding media consumption to gain insights into correlations and emerging patterns. In their study, the authors observed that in their sample male participants outnumbered female participants roughly four to one but female players played slightly more hours per week. Furthermore, players have better-than-average physical but slightly worse-than-average mental health. Lastly, Williams et al. (2008) noted that players of the MMORPG *Everquest II* rated the motivational dimension *Achievement* higher than *Immersion*, which in turn was more important than *Sociability*.

Analysis in "World of Warcraft"

In their work, Graham and Gosling (2013) chose to analyze the MMORPG World of Warcraft (WoW)[9] applying the *Big Five* personality traits and GMS (Yee, 2006) in order to relate these two measures and find correlations. During the evaluation of the composite questionnaire, the authors discovered the following correlations:

[8] Sony Online Entertainment, 2004. https://www.everquest2.com.
[9] Blizzard Entertainment, 2003. https://worldofwarcraft.com.

- *Sociability* is positively correlated with the trait *extraversion* and to a lesser extent with *agreeableness*, *neuroticism*, and *openness*. *Sociability* is negatively correlated with *conscientiousness*.
- The motivational dimension *Achievement* is negatively correlated with the trait *conscientiousness*.
- Players looking for *Immersion* also score relatively high on *openness*.

Lastly, the authors noted that differences in personality traits manifest in different ways when comparing online and offline worlds – pronounced differences where discovered for *conscientiousness*.

In another study utilizing GMS, the *Big Five* personality traits as well as measures associated with social phobia, mental and physical health, Lehenbauer-Baum et al. (2015) investigated the relation between personality, engagement and addiction as it relates to IGD (APA, 2013) in the game *WoW*. The authors found that addicted players scored significantly higher on social phobia, while scoring significantly lower on physical and mental health, as well as on environmental scores. Furthermore, players in the *addicted* category showed significantly lower scores in *agreeableness*, *conscientiousness* while scoring significantly higher in the trait *neuroticism*. Lastly, addicted players showed significantly higher scores on the GMSs for the clusters *Achievement* and *Immersion*.

Analysis in "Battlefield 3"

Conducting a study on the correlation between the *Big Five* and player performance, Tekofsky, Van Den Herik, Spronck, and Plaat (2013) used questions from an 100-question IPIP, as well as in-game telemetry from the game *Battlefield 3*[10]. The authors observed that the analyzed players overall scored high on *openness*. Furthermore, the trait *conscientiousness* to some extent predicts the play style by being negatively correlated with *actions per minute* and *deaths per second*. This implies a cautious play style resulting in fewer overall kills but also lower numbers of deaths. *Score per second* is negatively correlated with the traits *conscientiousness* and *extraversion*. Lastly,

[10] EA DICE, 2011. https://www.battlefield.com/games/battlefield-3.

Tekofsky et al. note that items related to work ethic are negatively correlated with in-game performance.

2.2.5. Comparing Analyses

In this section, the analyses discussed will be compared. The games analyzed, techniques used and use cases fulfilled in each publication are briefly contrasted.

From Table 2.4 we can already see that psychological theory has not yet been widely applied in the domain of game analysis. This is especially true for the analysis of players within the confines of individual games. From the research done so far, we can see that correlations exist between psychological frameworks, such as, the *Big Five* personality traits and the Gaming Motivation Scale (GMS) developed for analyzing player motivations. Furthermore, it has been shown that different player types, that is, players driven by differing motivations can be distinguished by their behavioral patterns.

In the following section another aspect of gaming will be described, which relates, specifically, to the dimension of *Sociability* in the GMS. This facet becomes more and more relevant as games increasingly offer multiplayer modes and player interaction.

	Game	Method	Use Case
Thawonmas and Iizuka (2008)	The ICE	KeyGraph on action sequences.	Analyze players of the three-cluster solution according to their in-game actions.
Williams, Yee, and Caplan (2008)	EverQuest 2	Demographics, GMS, health measures.	Theory forming not based on stereotypes but measurements.
Graham and Gosling (2013)	WoW	Big Five, GMS.	Relate motivations of play to personality traits.
Lehenbauer-Baum et al. (2015)	WoW	Big Five, IGD questionnaire, in-game data.	Explore personality differences between addicted and non-addicted players.
Tekofsky, Van Den Herik, Spronck, and Plaat (2013)	Battlefield 3	Big Five, in-game measures.	Correlate personality with in-game (performance) measures.

Table 2.4.: Comparison: Psychology in Games

2.3. The Social Side of Gaming

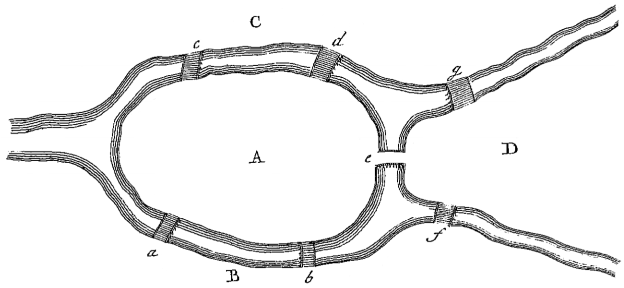
As soon as players play together in multiplayer games, it is not only necessary to understand how they interact with the game – the aspects we have focused on until now – but it is also important to understand how players interact with *each other*. One possible type of analysis that offers itself to such analyses is social network analysis (SNA). The foundation for SNA has been laid by Tichy, Tushman, and Fombrun (1979) who introduced the *Social Network Framework*, a framework for modelling social actors in a graph connected by their interactions. In their work, Tichy et al. found that organizations - as social networks - can exhibit emergent clusters, that is, communities of actors that interact with each other across prescribed boundaries, such as departments in a company.

Since SNA is based on and applies concepts of *graph theory* to social systems, it is in order to shortly introduce *graph theory* and define measures and properties insofar as they relate to SNA and this thesis.

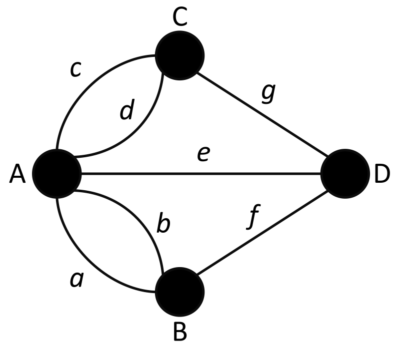
2.3.1. Graph Theory

The first known application of graph theory is believed to be the negative proof of "The Seven Bridges of Königsberg" by Euler (1741). In this paper, the city of Königsberg (modern-day Kaliningrad) is observed – its islands "Kneiphof" (A) and "Lomse" (D), and the bridges (a–g) connecting islands and shores (see Figure 2.5a). The author tries to answer the following question: *"Is it possible to plan a walk that crosses each of the bridges once and only once?"* – Is there an *Eulerian path*?

To find a solution to this problem, we observe that (1) no more than two regions may have an odd number of connecting bridges, (2) if there are two regions with an odd number of approaching bridges, a solving walk has to originate in of these regions, (3) if all of the regions have an even number of connecting bridges, a solution may start in any region, and (4) if there are two bridges connecting the same regions, we may remove them as they do not affect the solution. From Figure 2.5a we see that each of the four regions



(a) Geographic layout (adapted from Euler, 1741)



(b) Graph representation.

Figure 2.5.: Seven Bridges of Königsberg

A, B, C and D has an odd number of connecting bridges – there is no path satisfying the condition, hence no *Eulerian path*.

According to Biggs, Lloyd, and Wilson (1999, p. 9), *"A graph consists of a finite set of vertices, a finite set of edges, and a rule which tells us which edges join which pairs of vertices."* We may graphically represent the regions and bridges of Königsberg as vertices and edges, respectively, as seen in Figure 2.5b.

Consequently, a graph G may be written as a tuple $G = (V, E)$ with V as a set of vertices, and E being a set of edges, that is, ordered pairs (i, j) where i and j denote vertices connected by the edge.

Definitions and Properties of Graphs

After introducing the concept of a graph, we may now define properties and measurements that are either used directly in graph theory or form the basis of further metric in social network analysis.

Path "A path is a sequence of vertices and edges,

$$v_0,\ e_1,\ v_1,\ e_2,\ v_2,\ \ldots,\ v_{r-1},\ e_r,\ v_r$$

in which each edge e_i joins the vertices v_{i-1} and v_i $(1 \leq i \leq r)$" (Biggs et al., 1999, p. 9).

Connectedness We define that a graph is **connected**, if for every pair of vertices (i, j) there exists (at least) one path that starts in i and ends in j (Biggs et al., 1999).

*"A **disconnected** graph, that is, one which is not connected, splits up into connected parts, called its **components**"* (Biggs et al., 1999, p. 9).

Largest connected component (LCC) The largest connected component is the component (see definition above) containing the highest amount of connected vertices.

Degree (Valency) The degree of a vertex v, $\deg(v)$ is the number of incident edges, that is, the number of edges that are connected to v (Biggs et al., 1999). Note that, when summing the degrees of all vertices of a graph, twice the number of edges in the graph, that is, $2 \cdot |E|$ is obtained, due to the fact that each edge is counted twice – once for each "end" (cf. Euler, 1741, §16).

Density The density (or edge density) of a graph is defined as the proportion of existing edges to the number of possible edges (Diestel, 2005, p. 164). For undirected simple graphs, this can be written as

$$D = \frac{|E|}{|V|\,(|V| - 1)}$$

with $|E|$ in the numerator being the number of actual edges and the denominator following from the fact that each of the vertices may be (at most) connected to every other vertex. This definition disregards the possibility of multiple edges between vertices and loops, that is, an edge starting and ending in the same vertex (cf. Diestel, 2005, p. 28).

Clique A clique is a fully connected sub-graph, that is a set of vertices where each vertex is connected via an edge to every other vertex in the subgraph (Luce & Perry, 1949).

Community A community is a part of a network, that is, a set of vertices, with a relatively higher number of edges between them than edges that connect to other outside vertices – the inter-community density of edges is higher than the intra-community density (Girvan & Newman, 2002). An example of this property of community structure can be seen in Figure 2.6.

From this graph we see that vertices have more edges connecting to vertices inside their community (black) than to vertices outside their respective community (gray).

Clustering coefficient The *local* clustering coefficient of a vertex is defined according to its neighbors. If a vertex has n neighbors than the sub-graph containing those neighbors can have at most $\frac{n(n-1)}{2}$ edges. The clustering coefficient then represents the fraction of this possible maximum that are part of the graph. The *global* clustering coefficient is defined as the fraction

$$\frac{\text{number of actual triplets}}{\text{number of possible triplets}}$$

with triplets being permutations of the vertices and *actual triplets* being triangles in the graph, that is, a set of three vertices where each vertex is connected to both other vertices (Watts & Strogatz, 1998).

Small-World Property A network exhibiting the small-world phenomenon is characterized by short average path lengths and a high clustering coefficient (Milgram, 1967; Watts & Strogatz, 1998).

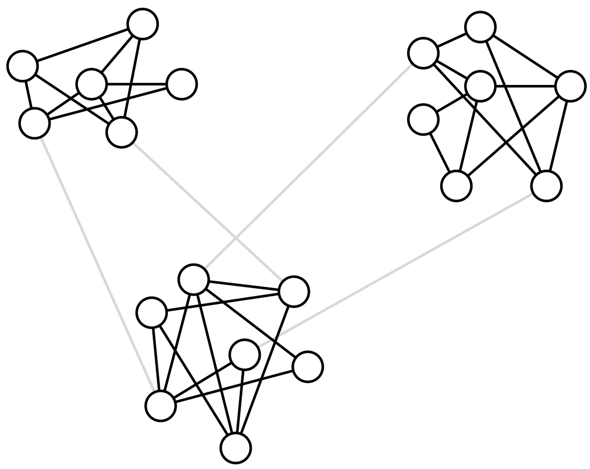


Figure 2.6.: A Network Exhibiting the *Community Structure* Property

After this short introduction to the fundamentals of graph theory, the concepts of SNA will be discussed in the next section, its applications in general and specifically the analyses applied to datasets related to video games.

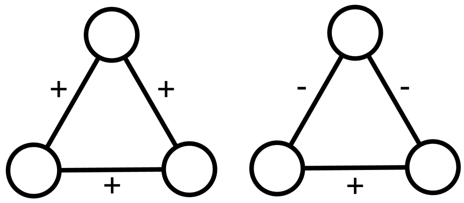
2.3.2. Social Network Analysis (SNA)

Based on graph theory mentioned earlier, methods of analyzing relationships between actors and their interactions with each other have been derived. Since its inception, SNA has been used to analyze a wide range of social networks. SNA has been proven to be useful when analyzing social structures in dolphin populations in New Zealand (Lusseau & Newman, 2004) and when studying the transmission of disease across possum populations (Krause, Croft, & James, 2007). Naturally, SNA has also been used to investigate social media and network sites, such as *Facebook*[11] (Nazir, Raza, & Chuah, 2008) and *Twitter*[12] (Ediger et al., 2010; Stepanyan, Borau, & Ullrich, 2010). It has been applied to companies' networks (Tichy et al., 1979; Lin et al., 2012) and also been used to model and understand benign structures, such as, open-source communities (Xu, Christley, & Madey, 2006). By contrast, it has also been used to study illicit organizations, such as drug trafficking rings (Bright, Hughes, & Chalmers, 2012), money launderers (Dreżewski, Sepielak, & Filipkowski, 2015) and terrorist cells (Koschade, 2006). Furthermore, SNA has been applied in the context of economic geography (Ter Wal & Boschma, 2009) and to aid stakeholder analysis (Prell, Hubacek, & Reed, 2009).

From this wide variety of applications it can be already seen that SNA can be a viable strategy for modelling player interaction and study patterns of group formation. In the context of cooperative and competitive play as a social system, SNA helps with understanding how players work together and against each other. As this thesis relies on data from `the100.io`[13], a site that places players into groups according to their preferences, it is worth mentioning how the initial group size of 100 has grounding in theory. Dunbar (1993) suggested an upper limit of 150 stable relationships which a modern-day human could maintain. This number is now known as *Dunbar's Number*. To arrive at this conclusion, the author used measures related to the size of the neocortex of primates and hunter-gatherer tribes and extrapolated

[11] https://facebook.com
[12] https://twitter.com
[13] https://the100.io. For further information about `the100.io`, see section 4.1.

these numbers to the neocortical size of modern humans. Furthermore, the author analyzed sizes of typical villages through the ages, as well as troop sizes of "companies" from the times of the Roman empire up until the 20th century and observed the same approximate limit. Moreover, Dunbar predicts that maintaining a social circle of this size would, on average, take up as much as 42% of the total time budget – more than twice the time spent by other primates, such as, baboons, macaques and chimpanzees (between 15 and 20%). Lastly, Dunbar noted that language may be seen as a shortcut used to reduce the time spent on social grooming by way of creating a more efficient method of communication and, in that fashion, allowing for more efficient bonding. This cognitive limit hypothesis was further substantiated by Gonçalves, Perra, and Vespignani (2011) in a study that analyzed Twitter networks. The authors analyzed a sample collected over the duration of six months containing over 380 million tweets written by over 1.7 million individual accounts, calculated a time-sensitive metric for interactions and found that social relationships with other Twitter users peak at around 150 and 200 contacts. Although it is possible to interact with more than 200 other users, the data clearly shows that after this peak the frequency of interactions plummets. Another important aspect of social networks, introduced by Heider (1958, chap. 7), is called *balance* and has since been formalized in *Social Balance Theory*. It describes states of stability and instability using the concepts of *dyads*, that is, a group of two people and *triads*, which describes a group of three people. Heider defined harmony for a *dyad* to exist, when its relationships are coherent, that is, either all interactions are positive or all are negative. In simpler terms, this could be described as either *friend* or *foe*. Similarly, for a *triad*, the author defined balance as either all positive relations, that is, a *a group of friends*, or containing two negative and one positive relation, which could be considered a *common enemy*. These structures, considered balanced, as described by Heider can be seen in Figure 2.7. The author further noted that disharmonious groups tended to evolve into harmonious, balanced groups over time thus reducing cognitive dissonance and minimizing mental load.



(a) All positive interactions.

(b) One positive, two negative interactions.

Figure 2.7.: Balanced Triads

2.3.3. Applications of Social Network Analysis in Games

Collaboration and competition have always been important aspects of playing games. Be it *Dungeons & Dragons*[14], bridge or chess, players have always been able to find ways to organize, form interest groups and, as a result, player communities far predate modern-day multiplayer online games (MOGs) (Pearce, 2009, p. 3). Identifying such *communities of play*, investigating their formation, analyzing their structure and interactions in and between groups can be a vital strategy in improving player experience and aid game development. For this reason, examples of SNA applied to games are listed and described here.

Social Analysis on "World of Warcraft (WoW)"

One phenomenon which has been observed in WoW is being *"alone together"*, where players are grouped in a party but do not really interact with other party members (Ducheneaut et al., 2006). This may be due to players in groups having to wait for one another, which results in players who never join groups levelling their avatars notably faster. Besides that, the authors noted that being part of a guild puts "social pressure" on their members. Guild members also play more and than non-members. Furthermore, guild

[14] Gygax, Gary and Arneson, David Lance, 1974. https://dnd.wizards.com.

membership ensures higher player retention as seen by higher-level avatars where fluctuation tends to be higher for players not partaking in a guild than in their guild counterparts. Ducheneaut, Yee, Nickell, and Moore (2007) have studied the formation, existence and decline of guilds in *WoW*, as well as factors contributing to the persistence of such. The authors identify that guild size is negatively correlated with its longevity while other elements, such as, inter-connectedness of players, "together ratio", that is, a measure for how much play schedules overlap and a wide variety of different character levels of players inside the guild positively predict the life-time of guilds. Furthermore, the Ducheneaut et al. identified that most guilds are considerably smaller than the limit predicted by Dunbar – consisting of roughly 35 members. In another study done by Poor (2015), the author found that guild membership in *WoW* does not necessarily help with progressing through the game faster. In this publication, the author used a dataset collected over a three-year period containing information about roughly 500 guilds and roughly 90,000 characters to measure the effect of being part of a guild on the process of leveling a character. This result is in line with the observations made by Ducheneaut et al. (2006) – ungrouped players level faster.

Social Analysis on "Ultima Online"

Analyzing *Ultima Online*[15], Allen (2004) found further corroborating data for Dunbar's Number as an upper limit. He argued that groups of sizes close to Dunbar's Number would emerge and persist only if social cohesion is strongly incentivized. To substantiate this claim, Allen analyzed group sizes of the online game and showed that most groups fall into categories of sizes of about 60. Furthermore, the author argued that – restricting Dunbar's hypothesis further – optimal group sizes fall into the range of 25–80, specifically into the range 45 – 50, where best performance may be observed.

[15] Origin Systems, 1993. https://uo.com.

Social Analysis on "Defense of the Ancients (DotA)"

One important question when building a social network is the selection of features to model as edges, as well as the removal of any observations deemed noise. In this regard, Van De Bovenkamp, Shen, Iosup, and Kuipers (2013) showed the effects of mappings and thresholds when translating observations into unweighted graphs by comparing common network measures in the game *Defense of the Ancients (DotA)*[16]. To achieve this, the authors used two datasets from popular matchmaking websites, *DotAlicious* and *Dota-League*. *DotA* is the first known example of a multiplayer online battle arena (MOBA) game. MOBA games are highly competitive and usually played in 3v3 or 5v5 modes, where players take control of a single hero and try to penetrate the other teams' defenses. The authors selected edges to represent aspects such as *matches played together*, *matches in the same team*, *matches in opposing teams*, appllied different thresholds, that is, cut-off points which define a required minimum of individual observations before an edge is added to the graph and subsequently analyzed network metrics such as component sizes, link density and network diameter. Furthermore, a component-based matchmaking algorithm was proposed.

In an attempt to challenge a form of undesirable behavior in online games, namely, *toxicity*, Märtens, Shen, Iosup, and Kuipers (2015) studied another dataset obtained from the game *DotA*. In the context of online gaming, *toxicity* describes perceived animosity towards other players and can negatively impact player experience. The authors built and trained a classifier that was able to automatically label toxic in-game messages sent during gameplay. The authors discovered that levels of toxicity, in this case only considering messages transferred within teams, for winning/losing teams fell/rose as the matches progressed and victory/defeat became more apparent. From this observation, Märtens et al. tried to predict game outcomes from "bad" words, *praise* and *slang* used and found that toxicity is not a good predictor for games' outcomes. Lastly, the authors concluded that their results, while based on a highly competitive MOBA game, could be generalized to wider areas and genres of games and that classifying and labelling such messages could help create a more positive experience for gamers in general.

[16] Eul, Steve Feak, IceFrog, 2003.

Social Analysis on "Pardus"

Empirically testing *social balance theory*, mentioned earlier, Szell, Lambiotte, and Thurner (2010) discovered corroborating data for the theory. The authors used a dataset composed of approximately 300,000 players from the game *Pardus*[17] with positive interactions, such as, friendships, private messages, and trades, on one hand, as well as negative interactions, such as, enemies, attacks, and bounties, on the other. Szell et al. showed that balanced triads, that is, those of type + + + and + - -, are overrepresented, while unbalanced triads (+ + -) are significantly underrepresented when compared to a random graph. This lends credence to *social balance theory* and indicates usefulness in regards to MOGs.

In further analyses of data obtained from *Pardus*, Szell and Thurner (2010) tested for preferential attachment, that is, the theory that "the rich get richer". In the context of social networks, this would be observable by newer players connecting to established high-degree nodes. The authors observede a positive trend between the probability of connecting to a node and this node's in-degree. In another step, Szell and Thurner discovered that the most common messages are exchanged between positively related characters meaning that friends much more often exchange messages than foes or characters with contradictory markings (A marked B as a friend, B marked A an enemy). The authors further collected data confirming the *weak ties hypothesis* ("The Strength of Weak Ties", cf. Granovetter, 1973) positing that casual links act as local bridges connecting remote communities in essential ways. Furthermore, the authors discovered data substantiating another aspect of *social balance theory*: triadic closure (Rapoport, 1953; Granovetter, 1973). Triadic closure describes the evolution of a not fully connected triad into a fully connected, *balanced* triad (see Figure 2.7), thereby reducing cognitive dissonance. Lastly, it is worth noting that Szell and Thurner also found confirmation for Dunbar's Number as the out-degrees of nodes in the constructed networks show an approximate upper bound of $k_{out} \approx 150$.

[17] Bayer&Szell OG, 2004. https://www.pardus.at.

Social Analysis on "Halo: Reach"

When analyzing *Halo: Reach*[18], Mason and Clauset (2013) discovered that both online and offline friendships with play partners have a measurable impact on in-game performance. For their analysis the authors combined demographic data from a survey with in-game data collected through the use of an application programming interface (API). While the authors found a noticeable bias in their survey's responses, such as significantly higher play-times and higher win ratios when compared to random samples, the results indicated the existence of "social pressure" leading to players with friends in their team to commit fewer betrayals, that is, killing teammates and aggregating more assists, that is helping a teammate score a point. Furthermore, playing with friends was shown to be beneficial to both individual and team performance. Lastly, the authors proposed a simple model for inferring friendships from observed behavior by using *length of the longest series of games played together* as a heuristic.

Social Analysis on "League of Legends"

Investigating social interactions, Kokkinakis, Lin, Pavlas, and Wade (2016) analyzed users' nicknames and inferred age to predict in-game behavior in the popular MOBA game *League of Legends*[19]. The game allows players to send positive feedback ('honor') or negative feedback ('reports') after a match has concluded. Working with this data about this feedback, as well as user names labelled as either "social" or "anti-social", the authors discovered that users with "anti-social" nicknames had significantly higher rates of both *reports sent* and *reports received*. The authors further noted that the number of interactions increased with age, while the rate of negative interactions – reports sent/received – decreased with age.

[18] Bungie, 2010. https://www.halowaypoint.com/en-us/games/halo-reach.
[19] Riot Games, 2009. https://www.leagueoflegends.com.

Social Analysis on "EverQuest II"

Another application of SNA has been shown by Keegan et al. (2010). In this work the authors work with a labelled set of accounts banned from the massively multiplayer online game *EverQuest II*[20] to construct a network of accounts involved in the illicit practices of "gold farming" and "real money trading". The authors observed that gold farmers and their affiliates display behavior that differs from the general population: farmers tend to have low connectivity to the overall network and lower-than-average transaction frequency, while their affiliates have a higher-than-average connectivity and transaction frequency. In addition, the farming network at large also revealed significantly higher clustering than the general population. Just like the real-world drug trafficking network *Caviar*, the farming network showed disassortative mixing, as opposed to the associative mixing in the non-affiliate network. This property makes the gold farming network more resilient against attacks, since a larger fraction of important nodes have to be removed in order to break apart the network and have a relatively low fraction of nodes in the LCC.

Comparative Analysis on "DotA", "StarCraft II", and "World of Tanks"

Jia et al. (2015) analyzed four community datasets of three different games: *DotA*, *World of Tanks*[21] and *StarCraft II*[22]. Again, the authors constructed networks describing various relationships, such as *matches played together*, *matches in the same team*, *matches in opposing teams*, *won together* and *lost together*. They then applied different threshold values n, for example, "for an edge to form players must have lost *at least five* matches together". Additionally, the authors considered a temporal filtering meaning that the n threshold value had to be met in a defined time span, such as, *one day* or *one week*. For all different network types the authors observed the small-world property and a power-law distribution for the nodes' degrees. When removing high-degree nodes, networks of all types quickly broke down with only

[20] Sony Online Entertainment, 2004. https://www.everquest2.com.
[21] Wargaming Minsk, 2010. https://worldoftanks.com.
[22] Blizzard Entertainment, 2010. https://starcraft2.com.

a small fraction of nodes remaining in the LCC. Furthermore, the authors noted that most graphs are socially balanced, meaning balanced triads are overrepresented in real-world signed graphs when compared to randomly signed graphs, while unbalanced triads are underrepresented, thus, further pointing towards the veracity of *social balance theory*. Lastly, based on this observation, the authors proposed a socially-aware matchmaking algorithm.

Social Analysis on "PlanetSide 2"

Hitherto, the focus of analyses discussed was on the player level. Another important level of analysis is on groups or clans to model and understand the dynamics in and between short- and long-lived groupings. Poor (2014) analyzed the online first-person shooter (FPS) *PlanetSide 2*[23], its different layers of player organization and their impact on mostly small persistent groups. For one there are so-called *outfits*, the equivalent to the more commonly named *guilds*, long-term groupings. Besides, there are also shorter-term groups named *squad*s, which consist of two to twelve players and *platoons* which are composed of two to four squads. Despite the game's focus on group battles, many of the outfits observed consisted of two or fewer players with a non-trivial amount of single-player outfits. Both small and large outfits were found to be beneficial in connecting the network of groups, as super connectors can be found in both categories. Lastly, the author also found connections, that is, explicit friendships as expressed through a friend list across faction borders, even though *PlanetSide 2* – a pure PvP game – focuses heavily on the conflict between the three factions.

2.3.4. Comparing Social Analyses

After discussing individual analyses at great length, this section is dedicated to giving an overview of some of the analyses discussed, the games they have been applied to, as well as the use case the methodologies fulfilled.

[23] Sony Online Entertainment, 2012. https://www.planetside2.com.

	Game	Focus	Method	Use Case
Ducheneaut, Yee, Nickell, and Moore (2006)	WoW	Players in groups and guilds.	Time series, SNA.	Fostering communities, structuring incentives.
Ducheneaut, Yee, Nickell, and Moore (2007)	WoW	Guilds.	Time series, SNA.	Fostering communities, ensuring group longevity & player retention.
Szell and Thurner (2010)	Pardus	Players, triads, interaction networks.	Time series, SNA, social balance.	Validating social balance theory, testing Dunbar's Number, evolution of triads.
Mason and Clauset (2013)	Halo: Reach	Players.	Questionnaire, time series of games.	Finding impact of (online/offline) friendships on retention, study play style.
Poor (2014)	PlanetSide 2	Groups, squads, outfits.	SNA.	Identifying super-connectors, investigating impact of small groups.

Table 2.5.: Social Analyses – A Comparison

From Table 2.5 it can be easily seen that so far group-level analysis has been neglected in favor of player-level analyses or general SNA. To date, player-focused analysis has been used for a wide range of applications, such as, verifying social balance theory, identifying nodes vital for connecting remote parts of the overall network, identifying illicit behavior, such as, "gold farming" and "real money trading" or undesirable behavior, such as, toxicity. In the same time, group level analysis has mostly been overlooked and – if done – relied heavily on capturing guild members' in-game data in realtime and hoping for minimal sampling issues.

2.4. Behavioral Analysis and Clustering

In recent years, game developers have introduced telemetry tracking and game analytics into their development and maintenance cycles (El-Nasr, Drachen, & Canossa, 2013, part I). Two types of metrics are of academic interest for this thesis: *gameplay metrics* and *community metrics*, where the former relates to the actual in-game behavior of the user and the latter relates to player interactions. In this section we will take a closer look at different approaches, tools, methods and analyses that have been carried out on various games.

2.4.1. k-means Clustering

k-means clustering is a form of clustering first introduced by Steinhaus (1957). Given n data points, k-means aims to partition the feature space into k regions, such that the variance of points inside a region, or cluster, is minimized. This variance is expressed as the sum of squared differences between the cluster center, that is, the *centroid*, and the points associated with it. The cluster method aims at minimizing the term (Pollard, 1981)

$$W_n = \frac{1}{n} \sum_{l=1}^{n} \min_{1 \leq j \leq k} \left\| x_l - c_j \right\|^2. \tag{2.1}$$

This definition uses the Euclidean norm, denoted as $\|\cdot\|$ as a quality measure. The intuitive algorithm for finding clusters, that is, sets of data points $\mathbf{S} = \{S_1, \ldots S_k\}$ and their centroids $c_1, \ldots, c_k$ can be written as shown in algorithm 1.

Algorithm 1: *k*-means Algorithm

Result: clusters $S_1, \ldots, S_k$; centroids $c_1, \ldots, c_k$
Initialize $c_1^{(0)}, \ldots c_k^{(0)}$ to e.g. random data points, $t = 0$;
repeat

```
//Step 1: Assignment
```

$S_i^{(t)} = \left\{ x_l : \left\| x_l - c_i^{(t)} \right\|^2 \leq \left\| x_l - c_j^{(t)} \right\|^2 \forall j : 1 \leq j \leq k \right\};$

```
//Step 2: Update
```

$c_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_l \in S_i^{(t)}} x_l;$

$t = t + 1;$

until *clustering converges*;

Each point in the data set is assigned to exactly one cluster during the *Assignment* step – this amounts to hard cluster memberships. In the *Update* step, the centroid is re-calculated as the average of all data points in the cluster. This is, in most cases, no actual point in the data set. Convergence has been reached, when no point is assigned to a different cluster in the *Update* step.

k-means has been criticized for its dependence upon selecting a correct value for k – the number of clusters. The quality of the resulting clustering relies heavily on this choice (Sugar & James, 2003). Furthermore, *k*-means finds a local optimum, rendering it sensitive to the initial conditions selected. Consequently, algorithms for selecting the correct number of clusters (Pelleg & Moore, 2000), as well as "good" starting conditions have been proposed (Bradley & Fayyad, 1998).

k-Medoids Clustering

As mentioned earlier, k-means does produce cluster representatives as centroids, which often do not lie on a specific data point and therefore can lead to issues in interpretability of the results. As a consequence, Kaufman and Rousseeuw (1987) introduced a clustering algorithm based on *medoids*. A *medoid* is defined as the point with smallest average distance to points in a dataset, and introduced the notion of a *dissimilarity measure* $d(x_i, x_j)$ between points x_i, x_j for this application. k-medoids now aims at minimizing the dissimilarity inside a cluster as (Kaufman & Rousseeuw, 1987)

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} d(x_i, x_j)\ z_{i,j} \tag{2.2}$$

with $z_{i,j} \in \{0, 1\}$ indicating cluster associations.

This approach is more versatile than k-means as it allows for distance measures $d(\cdot)$ other than the Euclidean norm and it is more robust to outliers. Furthermore, the resulting cluster centers, *medoids*, are easier interpretable as they are actual data points.

2.4.2. Archetypal Analysis (AA)

Archetypal analysis (AA) is an approach to unsupervised learning first introduced by Cutler and Breiman (1994). The main idea of this algorithm is to find extreme points, so called "pure types", or "archetypes" that can be used for describing the data. Data points are then described as convex combinations of the archetypes.

These p archetypes are found by minimizing the residual sum of squares for n data points in an m-dimensional feature space as given by

$$RSS = \min_{\alpha_{ik}} \sum_{i=1}^{n} \left\| x_i - \sum_{k=1}^{p} \alpha_{ik} z_k \right\|^2 = \sum_{i=1}^{n} \left\| x_i - \sum_{k=1}^{p} \alpha_{ik} \sum_{j=1}^{n} \beta_{kj} x_j \right\|^2 \tag{2.3}$$

with z_k being the archetypes, as shown by Cutler and Breiman (1994, p. 15). Equation 2.3 can be rewritten as

$$RSS = \|\mathbf{X} - \mathbf{Z}\mathbf{A}\|_F = \|\mathbf{X} - \mathbf{X}\mathbf{B}\mathbf{A}\|_F \tag{2.4}$$

with $\mathbf{Z} \in \mathbb{R}^{m \times k}$, $\mathbf{A} \in \mathbb{R}^{k \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times k}$ (Sifa, Bauckhage, & Drachen, 2014a). The matrices $\mathbf{A}$ and $\mathbf{B}$ are column stochastic, that is, subject to the constraints

$$\begin{aligned} a_{jk} &\geq 0, \\ \|\mathbf{a}_j\|_1 &= 1 \qquad \forall j \in [1, 2, \ldots, n], \\ b_{ki} &\geq 0, \\ \|\mathbf{b}_i\|_1 &= 1 \qquad \forall i \in [1, 2, \ldots, k]. \end{aligned} \tag{2.5}$$

One important aspect of AA is the soft clustering, that is, membership to an archetype is not expressed in a binary fashion but rather as coefficients denominating relative membership to each archetype following from the fact, that data points are being approximated as convex combinations of the archetypes as

$$\mathbf{x}_j \approx \mathbf{Z}\mathbf{a}_j = \sum_{i=1}^{k} \mathbf{z}_i a_{ij}. \tag{2.6}$$

Furthermore, Cutler and Breiman express archetypes as convex combinations of data points as

$$\mathbf{z}_i = \mathbf{X}\mathbf{b}_i = \sum_{j=1}^{n} \mathbf{x}_j b_{ji}. \tag{2.7}$$

To find a solution, the authors proposed an alternating algorithm that iteratively solves convex least squares problems to find matrices $\mathbf{A}$ and $\mathbf{B}$ sufficiently minimizing the error. More general, the search for optimal matrices $\mathbf{A}, \mathbf{B}$ subject to the constraints given by Equation 2.5 is a non-negative matrix factorization (NMF). An optimization of NMF is simplex volume maximization (SIVM) as introduced by Thurau, Kersting, and Bauckhage (2010) which can be used on high-dimensional data and – by way of speeding calculations up – permits the analysis of very large data sets. This speed-up over the commonly used *convex* NMF is achieved by optimizing the approximation of $\mathbf{Z} = \mathbf{XB}$ – cf. Equation 2.4. The authors' proposed approach builds on distance geometry. For n data points and k basis vectors, that is, archetypes it offers a computational complexity of $\mathcal{O}(kn)$, with $k \ll n$. Lastly, the authors applied the algorithm on an image database consisting of 80 million images being able to identify the fundamentals of the color space used, namely sinusoid encodings with differing phases, orientations and frequencies, as well as constructing archetypal basis vectors from tweets of Barack Obama containing, for example the rather long tweet

> *"Thinking we can cut oil consumption by 2.5 million barrels of oil per day and take 50 million cars' worth of pollution off the road by 2020"*[24]

but also the rather short example of

> *"Yes we can."*[25]

One variation of NMF is *convex-hull* NMF as shown by Thurau, Kersting, and Bauckhage (2009). The authors restrict the problem further and only consider points on the data convex hull as candidate archetypes. This approach helps speed up finding archetypes - as the expected size of the convex hull is much smaller than the size of the dataset - and increases interpretability of archetypes, as the identified archetypes are close to actual data points. While it is generally not possible to compute a closed form solution for (Equation 2.3), Mørup and Hansen (2012) suggested a fast algorithm for fitting the AA model and compared the performance of AA to other clustering and dimensionality reduction algorithms, such as, singular value decomposition, PCA, independent component analysis, NMF and k-means clustering. Furthermore, the authors observed that the construction of "pure" archetypes, that is, archetypes as convex combinations of data points – cf. Equation 2.7 – may not always be possible as seen in Figure 2.8. Hence, Mørup and Hansen proposed to relax the constraints for archetypes, such that they do not have to lie on the convex hull but still allowing for data points to be expressed as convex combinations of archetypes (cf.

[24]For the original Tweet see https://web.archive.org/web/20081025203013/https://twitter.com/BarackObama/status/55928192, Archived version from October, 25th 2008

[25]For the original Tweet see https://web.archive.org/web/20100325200322/https://twitter.com/BarackObama/status/10852146480, Archived version from March, 25th 2010

Equation 2.6). Lastly, the authors showed that AA yields similar or better results than widely used methods for a wide variety of domains, such as, computer vision, neuro-imaging, chemistry, text mining and collaborative filtering.

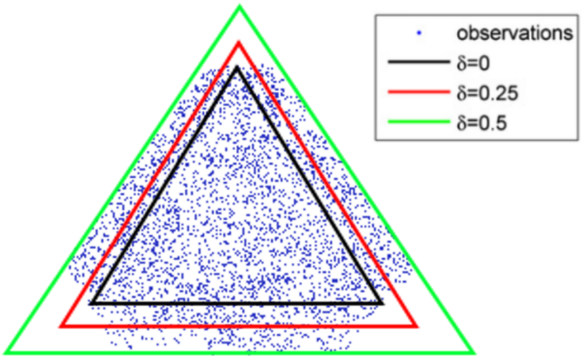


Figure 2.8.: Illustration of relaxed PCH/AA-δ (adapted from Mørup and Hansen, 2012)

Bauckhage and Sifa (2015) further built upon the idea of clustering using points on the data convex hull as archetypes and devised the method of *k-maxoids clustering*. The authors introduce the notion of an extreme point in a data set $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{n} \subset \mathbb{R}^m$, a *maxoid*, as

$$\mathbf{m} = \arg\max_{\mathbf{x}_j} \frac{1}{n} \sum_{i=1}^{n} \left\| \mathbf{x}_j - \mathbf{x}_i \right\|^2, \tag{2.8}$$

which is the point of highest average distance to all other points (in a dataset), the point furthest away from the mean. The authors proposed a clustering algorithm that not only uses maxoids but also maximizes the distance between maxoids. One important aspect of the authors' approach is the interpretability of clusters or archetypes. This was shown on the example of a image database of faces where k-means lead to blurry images as they are averages of their associated faces, k-medoids resulted in archetypal images that were not noticeably distinct – the respective medoids were relatively close together in the feature space and k-maxoids culminated in clear images that are easily distinguished by the human eye.

2.4.3. Video Games and Behavioral Profiling

Over the course of the last few years, AA has seen a wide range of applications in the field of games research. These use cases include recommender systems and clustering in-game behavior, such as, weapon and vehicle usage, as well as differing overall play styles.

Behavioral Analysis and "Steam"

One notable application of AA in relation to video games has been shown by Sifa et al. (2014a). The authors designed a recommender system for digital games on the online distribution platform *Steam*[26]. The authors proposed two recommendation models based on decomposition of the playtime matrix $\mathbf{T}$ as

$$\mathbf{T} \approx \mathbf{G}^T \mathbf{P}, \tag{2.9}$$

with $\mathbf{G}$ being the *game factor matrix* and $\mathbf{P}$ being the *player factor matrix*. The first model – a *factor-oriented model* (AAF) – uses the decomposition in Equation 2.9 to estimate the strength of an association between a player i and a game j as

$$\hat{t}_{ji} = \mathbf{g}_j \mathbf{p}_i = \sum_{u=1}^{k} g_{uj} p_{ui}. \tag{2.10}$$

The second model – a *neighborhood-oriented model* (AANeP) – uses the player matrix $\mathbf{P}$ obtained from Equation 2.9, a similarity measure $\mathrm{sim}(\mathbf{v}_1, \mathbf{v}_2)$ and a player's neighborhood $\mathcal{U}$ containing the closest players to estimate the player-game association for a player i and a game j as

$$\hat{t}_{ji} = \frac{\sum_{u \in \mathcal{U}} \mathrm{sim}\left(\mathbf{p}_i, \mathbf{p}_u\right) \mathbf{t}_{ju}}{\sum_{u \in \mathcal{U}} \mathrm{sim}\left(\mathbf{p}_i, \mathbf{p}_u\right)} \tag{2.11}$$

These two models were compared to baseline algorithms, such as, a *random* recommender, a recommender based on global popularity considering only installations or purchases, a recommender based on total global playtime and a nearest-neighbor recommender. To test the performance of these two

[26]Valve Corporation (2003). https://store.steampowered.com

models, the authors blinded games, that is, they removed known play times and tried to predict the player-game association and compared the *recall* (cf. Equation 2.12, defined by Kent, Berry, Luehrs, and Perry, 1955) of the Top-L highest ranked recommendations returned by each model.

$$\text{Recall} = \frac{\text{selected items}}{\text{relevant items}} \tag{2.12}$$

The authors concluded that the proposed AA-based recommenders consistently performed better than the baseline recommender systems and were able to reach recall rates as high as 97%, with AANeP having higher rates of recall for smaller L and being overtaken by AAF for larger and larger numbers of top-L results. In another study, Sifa, Bauckhage, and Drachen (2014b) have demonstrated the usefulness of AA for exploring playtime patterns for different genres of digital games on *Steam*. Based on the playtime distribution, the authors were able to distinguish between four archetypal game types. For example, one such type was composed of short-lived games, that exhibit a rapid decline in play time after only a few hours, while another could be classified as large-scale production, AAA titles that have users playing the game for 25 hours or more. In this regard, AAA refers to games with the backing of a large publisher, and are consequently characterized by higher budgets for development and marketing. Using this pattern recognition, the authors discovered similarities in time series between various FPS games as well as a different set of time series distribution for F2P games. Finally, the authors concluded that the different genres of games show variations in their respective tail distributions. In another related study, Sifa, Drachen, and Bauckhage (2015) applied k-means clustering to play-time data in order to find patterns hinting at prototypical players. The authors observed that with their 11-cluster solution almost half of the players were dedicated to one of four of *Valve*'s major titles (*Dota 2*[27], *Team Fortress 2*[28], *Counter-Strike*[29], and *Counter-Strike: Source*[30] with players in most other clusters also clearly dedicating a majority of their play

[27] Valve Corporation, 2013. https://www.dota2.com.

[28] Valve Corporation, 2007. https://store.steampowered.com/app/440/Team_Fortress_2.

[29] Valve L.L.C, 2000. https://store.steampowered.com/app/10/CounterStrike.

[30] Valve Corporation, 2004. https://store.steampowered.com/app/240/CounterStrike_Source.

time to a single game or game series. Furthermore, Sifa et al. re-ran the AA (cf. Sifa et al., 2014b) on a preprocessed and enhanced data set to try and identify games that offer long retention times. The authors discovered similar clusters and were able to break the genres of each cluster down further. As mentioned earlier, the longest player retention is achieved by big-production AAA titles, with the quickest fall-off being observable on old titles, such as re-releases. The second-most slow drop in retention was observed for older AAA titles and in the game genres *adventure*, *action* and *point & click*, specifically. Additionally, Sifa et al. mined rules for recommending games to players based on their past play times using Frequent Item Set Mining (Agrawal & Srikant, 1994) and Association Rule Mining (Gow, Colton, Cairns, & Miller, 2012). The resulting rules are heavily skewed by the prevalence of *Team Fortress 2*. Lastly, the authors tried to find correlations between ratings on *Steam* and on the review aggregation site *Metacritic*[31] and concluded that there was little to no correlation between sales, play time and ratings on either *Steam* or *Metacritic*.

Behavioral Analysis in "Star Wars: Galaxies"

In one of the earlier attempts to understand players' in-game behavior and players' interactions with each other, Ducheneaut and Moore (2004) analyzed *Star Wars Galaxies*[32], a MMORPG launched in 2003. The game was designed around a number of different inter-dependent classes. Players need to interact with players of other classes. For example, a bounty hunter needs a doctor for healing and buffs, that is, a temporary increase of character stats. In exchange, the doctor receives in-game currency. Besides trading services, players may choose to use gestures, such as, greeting, smiling, cheering, winking, and waving in order to garner attention by other players. The authors analyzed players' behavior by creating interaction profiles contrasting *gestures received* with *gestures sent* and comparing relative frequencies of these profiles at two distinct locations, namely the *cantina* on one hand and the *starport* on the other. The authors observed different player profiles, that is, players in the *cantina* emoted more often as players use it to initiate trades

[31] https://www.metacritic.com
[32] Sony Online Entertainment, 2003.

and exchange currency for services and buffs, while players in the *starport* did not use gestures as frequently but rather chatted while waiting for the next interplanetary transport. Lastly, the authors noted that the macro system was frowned upon by players and may be seen as detrimental to the game. The system was seen as a nuisance by players due to outdated exchange ratios and results in lowered player engagement by automating instrumental exchanges, such as *healing*, while effectively eliminating any reason to chat with other players.

Behavioral Analysis in "Tomb Raider: Underworld"

Drachen, Canossa, and Yannakakis (2009) used self-organizing maps and *k*-means to analyze and cluster player behavior in the game *Tomb Raider: Underworld*[33]. The authors selected features, such as *causes of death*, *total number of deaths* and *completion time* and subsequently trained an artificial neural network. Drachen et al. then noted that *k*-means already indicated a three- or four-cluster solution with the self-organizing map pointing towards a four-cluster solution. The clusters of players were then identified as *Veterans*, who seldomly died and finished the game in a short amount of time, *Solvers*, who carefully progressed through the game, rarely dying and taking a long time to finish the game, *Pacifists*, who most often died to active opponents, and *Runners*, who finished the game fast while dying a lot in the process.

Behavioral Analysis in "World of Warcraft (WoW)"

Thurau et al. (2009) analyzed guilds in *WoW* by applying convex-hull NMF on player activity time series. The authors found eight different archetypes of guild activity profiles. Those included *seldom active*, *high activity until second update*, *formed early then faded away* and *formed early then slowly disbanded after first update*. Furthermore, Thurau et al. noted that the largest portion of guilds in the data set were close to the archetype *seldom active*, while none

[33] Crystal Dynamics, 2008. https://store.steampowered.com/app/8140/Tomb_Raider_Underworld.

of the most successful guilds exhibited a high belongingness coefficient for this archetype.

Comparative Analysis: "Tera", and "Battlefield: Bad Company 2"

Analyzing the in-game behavior of players in the games *Tera*[34] and *Battlefield: Bad Company 2 (BF:BC2)*[35], Drachen, Sifa, Bauckhage, and Thurau (2012) combined and compared both, *k*-means and SIVM. To approach this clustering, the authors selected features, such as, *number of monster kills*, *number of looted items*, *character level* and *accuracy*, *kill/death ratio*, *score per minute* for *Tera* and *BF:BC2*. The authors noted that the two approaches, *k*-means and SIVM, serve different purposes in exploratory data analysis: *k*-means offers an overview of the general behavior of the player population, while SIVM helps understand extreme behaviors. Furthermore, Drachen et al. identified the usefulness of SIVM in detecting illicit behavior, such as gold farming. Using either method for the game *Tera*, the authors were able to discern two distinct groups, *elites* and *stragglers*. Similarly, both methods yielded the clusters *veterans*, *assassins* and *training dummies* for *BF:BC2*. To sum up their work, both *k*-means and SIVM have their own distinct applications and are useful tools for analyzing and exploring player behavior.

Behavioral Analysis in "Battlefield 3"

Another facet of in-game behavioral analysis that benefits from AA has been shown by Bauckhage and Sifa (2015). The authors compared the clustering methods *k*-means and *k*-maxoids for the usage of vehicles in *Battlefield 3*[36]. Bauckhage and Sifa showed that differences between player archetypes in regards to commandeering vehicles are much more pronounced. The resulting clusters are more easily explainable when using *k*-maxoids clustering over *k*-means, as seen in Figure 2.9.

---

[34] Bluehole Studio, 2011. http://tera.enmasse.com.
[35] EA DICE, 2010. https://www.battlefield.com/games/battlefield-bad-company-2.
[36] EA DICE, 2011. https://www.battlefield.com/games/battlefield-3.

(a) Clusters: k-means

(b) Clusters: k-maxoids

Figure 2.9.: Comparison: vehicle clustering k-means/ k-maxoids (adapted from Bauckhage and Sifa, 2015)

In the clustering obtained from *k*-means (seen in Figure 2.9a), we see that the first few clusters are very similar with most clusters using mainly tanks, while the clustering obtained from *k*-maxoids (seen in Figure 2.9b) results in very distinct clusters, with one cluster focusing on planes, another preferring helicopters, and so forth.

Behavioral Analysis in "Destiny"

Drachen, Green, et al. (2016) chose to analyze data from the game *Destiny*[37]. The authors divided the dataset at hand into two parts: the first containing telemetry data collected during PvE gameplay and the second over the course of PvP games. The authors then applied different clustering approaches, videlicet, *k*-means, *k*-maxoids, Gaussian mixture models (GMM), as well as AA. In order to minimize level-related side effects, Drachen, Green, et al. then restricted their avatars under consideration to only contain those that reached the maximum level. After inspecting the results obtained for a variety of numbers of clusters, the authors decided to restrict their results to only feature AA and GMM as those methods lead to highly interpretable solutions. Further, Drachen, Green, et al. selected to use four-cluster solutions for the PvP data set and when analyzing the PvE data, a five-cluster solution for GMM, as well as a four-cluster solution for AA. It is worth noting that all of the solutions found contained a cluster or even multiple clusters of players relying heavily on long-range combat. Lastly, it was shown that the difference in PvP and PvE gameplay lead to changes in play styles and different groups of players tended to approach the game differently by, for example, being more team-oriented and resurrecting players more frequently. In another analysis of in-game telemetry data of *Destiny* focused on weapon usage, Rattinger et al. (2016), used AA on features, such as, scores, kill/death ratio, medals and weapon preferences to classify different play styles and were able to identify different archetypal clusters as, for example, *newbies*, *ranged elites* and *melee* player types. Further, Rattinger et al. noted that this kind of clustering benefits from a soft membership function, as hard classification could lead to a large margin of errors, due to a large fraction of players belonging to two or even three archetypes to a non-trivial

[37] Bungie, 2014. https://www.destinythegame.com/d1.

extent, while still almost none belong to four or even all five clusters. The authors further described that the network constructed from *matches played together* indicated assortative mixing in regards to calculated archetypes. Lastly, while players seem to begin with one kind preference of weapon type, they tend to evolve into using other weapon types over time. In follow-up work, Pirker et al. (2018) then inspected *Destiny* match data for influence of social aspects, such as interactions, relationships and clan membership on individual performance, team performance and engagement. The authors split the dataset into two groups, the first being composed of *focused players*, those who play with the same group of players regularly, and the second representing *open players*, those who interact with a wider range of players more frequently. From their analysis, the authors find that *focused players* win more often, perform better individually (as expressed by the kill/death ratio) and tend to be able to control the game forcing shorter play times (upon their victory). Furthermore, clan membership seems to imply a more focused play style with better performance and engagement with more overall matches played.

2.4.4. Comparing Algorithms

In this next section, the presented learning algorithms are shortly compared and their applications are described. For the sake of this comparison, *k*-maxoids is classified as an extension of AA.

As seen in Table 2.6, AA offers soft clustering and, as *k*-medoids, offers robustness against "bad" choices of initial conditions, that is, cluster centers. AA, furthermore, aims at finding extreme points and describing all other points as convex combinations of these *archetypes*. Therefore, it can be argued that AA is a soft clustering approach. All of the described algorithms in their basic form require the number of clusters (or archetypes) to be defined a priori and given as an input.

	Algorithm		
	k-means	k-medoids	AA
Clustering	Hard.	Hard.	Soft.
Notable Points	Centroids.	Medoids.	Archetypes.
Outliers	Sensitive by design.	Robust.	Sensitive, can be made robust.
Initial Conditions	Sensitive.	Robust.	Robust.
Application	Exploring "usual" behavioral patterns.	Finding prototypes of "normal" behavior.	Detecting extreme behaviors.
Number of Clusters		Need to select.	
Shape of Data	Spherical clusters.	Spherical clusters.	Arbitrary.
Interpretability	Centroids are not intuitively explained.	Medoids can be hard to distinguish.	Archetypes are distinct and vary greatly.

Table 2.6.: Comparison of Clustering Algorithms

2.5. Summary

This chapter is meant to offer an introduction to the fundamental topics associated with this thesis and related publications. Firstly, the important concepts of *player retention*, *player performance* and *player experience* were introduced. These notions are highly connected to player behavior and it is of utmost importance to understand them. As a first step towards understanding players, psychological theories of personality were discussed, as well as frameworks developed to directly classify players. The personality assessments offered by the MBTI, and the closely related Keirsey Temperament Sorter were introduced and discussed. Afterwards, the state-of-the-art character descriptor, the *Big Five* personality traits were described. Furthermore, attempts of bridging the gap between games research and psychology were discussed. In this context, studies applying psychological theories to players were discussed. These publications tried to understand the interconnections between player personality and gaming preferences, players' characters and possible relations to addict behavior traits. Next, theories of understanding behavior specifically tailored to playing games were introduced and discussed. In this regard, first attempts at understanding player behavior using the framework proposed by Bartle (1996) were shown. Subsequently, attempts at finding links between frameworks for player classification and psychological theories were presented. However, the player classification proposed was found lacking, thus, giving way to the three-cluster model, GMS, suggested by Yee (2006). Concrete examples of the application of psychological theory and player motivation theories were discussed and contrasted. In the penultimate section, the focus was on SNA. First, the relevance of SNA in social groupings was motivated using a wide range of analyses. Consequently, this field of study is also important for understanding cooperative and competitive gaming. As SNA, heavily relies on graph theory, the latter was shortly introduced and concepts and measures relevant to this thesis were defined. Next, theories related to SNA, such as, Dunbar's Number and *social balance theory* were introduced and described shortly. Concluding this section, specific examples of SNA being applied to games and testing the concepts of *social balance theory*, and Dunbar's Number were briefly discussed. In the last section of this chapter, algorithms and approaches for clustering and their application towards finding clusters of player behavior were discussed. The first algorithm, k-means, was briefly introduced, as it has seen a wide field of applications – including games research. Next, an algorithm using a closely related concept, namely, *medoids*, was described – k-medoids clustering. Lastly, a technique used for finding extreme examples, archetypal analysis (AA), was introduced. The mathematical foundation was described, as well as the related algorithm of k-maxoids clustering, and optimizations to the algorithm, such as, convex NMF, *convex hull* AA and relaxing the convexity constraint to arrive at AA-δ. Thereafter, applications of these clustering algorithms in the field of games research were briefly discussed. Finally, key characteristics of the algorithms, and important differences between them were highlighted.

3. Destiny - Gameplay

The information about Destiny in this section was retrieved in part from Destinypedia community (2013), and Destiny Wiki community (2014).

Destiny[1] is an online multiplayer first-person shooter (FPS) game developed by *Bungie* with Jason Jones as a creative director, published by *Activision* and was released in September of 2014. Besides being a FPS game, *Destiny* also features role-playing game (RPG) elements and is set in a large shared-world environment making it an *FPS-massively multiplayer online role-playing game (MMORPG)* hybrid, effectively. Every *Guardian* has a *Super*, an ultimate ability that depends on the class and sub-class chosen. The *Super* abilities are powerful offensive or defensive abilities which may only be used after a cooldown which may be reduced by scoring kills or by picking up *Orbs of Light*. In-game progression is based on gathering *experience points* and leveling the *Guardian* and – after reaching the maximum character level – shifts to maximizing the *Light level* calculated from the players' equipment. In this way, *Destiny* also is an exemplar of the *loot shooter* sub-genre of *shooter* games. An example of gameplay can be seen in Figure 3.1.

[1] Bungie, 2014. https://www.destinythegame.com/d1.

Figure 3.1.: *Destiny* Gameplay. © Bungie, Inc. Destiny, the Destiny logo, Bungie and the Bungie logo are registered trademarks of Bungie, Inc. All rights reserved. Image courtesy of Bungie, Inc.

3.1. Setting

In *Destiny*, the player takes on the role of a *Guardian*, a member of a faction tasked with preserving the *Light* emanating from the *Traveler* and defending the solar system against the *Darkness*. *Guardians* are at war with other factions, such as the *Cabal*, *Fallen* and *Hive* who all pursue different objectives related to the *Traveler* or the *Light*. *Guardians* are accompanied by robotic drones called *Ghosts* which were created by the *Traveler* by shedding pieces of itself. The story mode revolves around missions given by non-player characters (NPCs) in the hub-worlds on Earth, its Moon, Mars and Venus. Players can also choose between three different species for the player character with regular *Humans*, *Awoken* as humanoids with pale blue-grey skin descending from *Humans* and *Exos* being conscious war machines built by humans during humanity's *Golden Age*. Besides these species, players get to select one of three classes for their *Guardian* as described in the next section.

3.2. Class System

Figure 3.2.: *Destiny* Classes. © Bungie, Inc. Destiny, the Destiny logo, Bungie and the Bungie logo are registered trademarks of Bungie, Inc. All rights reserved. Image courtesy of Bungie, Inc.

For their playable character, players are able to choose between three classes (depicted in Figure 3.2), all of which offer different inherent traits and various sub-classing choices that further allow to specialize in different playstyles. The distinction between classes is briefly discussed here:

Hunter

> *"What does it mean to be a Hunter? I say, it's all about where you belong. The Warlocks have their libraries, the Titans have their walls... But Hunters belong in the wilds."*[2] (Cayde-6)

Hunters allow players to play the well-established thief/rogue playstyle which is centered around dealing large amounts of damage quickly. Besides that, their sub-classes further allow them to specialize in very long-range or very-short range combat. Furthermore, the *Hunter*'s abilities enable it to use crowd control skills.

Titan

[2]Quote taken from in-game dialogue.

> *"What does it mean to be a Titan? As a Titan, you are a part of the City – in a way no Warlock or Hunter could understand. The dream of the City rests upon our shoulders."*[3] (Commander Zavala)

Titans focus on strength and endurance. They fulfill the role of "tanks", that is, this class offers a larger life pool than other classes. *Titans* are designed to grant their fellow players defensive support by either shielding off incoming damage or withstanding it for them. Due to their sub-class abilities, *Titans* excel in melee-range and short-ranged combat.

Warlock

> *"What does it mean to be a Warlock? Power. Only Warlocks understand true power."*[4] (Ikora Rey)

Warlocks are a mage-like or "space wizard" class with a smaller life pool and fast life regeneration. They use the power granted by the *Traveler* to harness the elements to either inflict damage or support other players by granting them buffs. Players controlling their *Guardians* with these specific classes are able to choose to play in a variety of different game modes. The most fundamental distinction that can be made is between player-versus-environment (PvE) and player-versus-player (PvP) game modes. In this next section, further classifications and game types are introduced.

3.3. Expansions

After its initial release, the game received five major expansions which furthered the story, added more missions, introduced new core game features, included a number of endgame content and raised the level cap. The consecutive additions and changes in maximum player and *Light levels* can help explain any observed plateaus in the collected data. The expansions added to the game are:

The Dark Below was released in December 2014, added three new story missions centered around the *Hive* race and offered a new raid.

[3] see [2]
[4] see [2]

House of Wolves, the second expansion, released in May 2015, furthered *Destiny*'s story by adding five new story missions focusing on the *Fallen* race and added a new PvE arena and another PvP game type.

The Taken King marks the end of *Destiny*'s "Year One". It was released in September 2015 and adds a number of story missions revolving around the *Taken* race and their king. The maximum *Light level* was increased to 320, a new raid was added and this expansion introduced new *Guardian* sub-classes. A later update to the expansion raised the maximum *Light level* again, this time to 335.

Rise of Iron, the fourth expansion, was released in September 2016 and marks the end of "Year Two" of *Destiny*. The expansion focused on the *Fallen* – adding new story missions deepening this race's lore, introduced a new PvP game mode and a new raid. The *Light level* was further raised to 400.

Age of Triumph is the last expansion to date and was released in March 2017. It added in-game tracking for player progression and raised earlier raids to higher *Light levels*.

3.4. Game Modes

As mentioned earlier, *Destiny* offers a number of game modes in both PvE and PvP. It is important to note that all kinds of activities can be played with other players, in multiplayer mode – either by their very design, as PvP, or in a fireteam in PvE modes. In this next section, first PvE and then PvP game modes are going to be described.

3.4.1. Player-versus-Environment (PvE)

Destiny offers a wide range of different PvE game modes for players to participate in. These various game modes are going to be described here in short.

- **Quests** are tasks given by NPCs in different locations. These quests often tie into the main story but can also relate to the player's chosen *Guardian* class, specific events, expansions or even *The Crucible*. *Quests* can award gear, that is, weapons an armor, experience points or Emblems.
- **Story Missions** are the way of exploring *Destiny*'s main story arc and reveal the game's lore. These missions can be played in singleplayer mode or cooperatively in fireteams composed of up to three players. Most missions are designed to be a series of objectives concluding in a boss fight. *Story missions* can be replayed indefinitely without rewards on repeated playthroughs, though. These missions are unlocked by completing quests and every day, a story mission is selected to be a *Daily Heroic Story Mission* which offers higher rewards. *Story missions* were expanded in some of the expansions following the initial release.
- **Strikes** are similar to the story missions described earlier. As such, they also feature consecutive objectives that players have to fulfill and end in a boss fight. They can also be played alone or in teams of up to three players cooperatively. *Strikes* reward different items according to the playlist played and in accordance with the players' *Light level*.
- **Patrols** are activities encountered when roaming freely through the worlds offered in *Destiny*. Players can accept *patrols* by interacting with beacons found during exploration. In the same way friends as well as random players can join them. Among others, *patrols* can be "fetch missions" where players have to collect items dropped by enemies. *Patrols* can also be "kill missions" where players have to kill a number of enemies, "assassination missions" where players have to kill a specific target or "survey missions" where players have to travel to certain locations and inspect the area.
- **Public Events**, like *patrols* are encountered when exploring worlds and can be tackled alone and cooperatively in groups. *Public events* are time-critical and the players' performances are rated after the event has concluded. These events require players to fulfill tasks, such as, killing bosses or mini-bosses, defending a specific area or to travel and kill certain enemies.
- **Bounties** are a kind of side mission and can be picked up at the Bounty Tracker. They require the player to fulfill specific task and reward experience, reputation, as well as, sometimes exotic weapons

on completion. Some *bounties* tie into the main story, while others relate to playing *strikes* or *The Crucible*. In addition, some of the *bounties* are tied to the *Guardian* class the player chose to play.

- **Raids** are the highest-level activity in the PvE game mode and alternate every week. They exhibit high difficulty and require the players to play as a six-*Guardian* fireteam. *Raids* are structured like missions, involving a sequence of goals that players have to reach, culminating in a boss fight and offer unique loot based on the type of *raid* completed.

3.4.2. Player-versus-Player (PvP)

The PvP mode in *Destiny* is called *The Crucible*. In order to level the playing field, level differences and edges gained through gearing differences are mitigated by stat normalization in most *Crucible* game modes. Playing *Crucible* matches awards reputation, gear and currency to buy gear. The game modes offered in *The Crucible* and their variations will be described here briefly.

- **Rumble** is a six-player free-for-all PvP game mode. Each player plays on their own and points are awarded for kills. As such it is similar to the well-established Deathmatch game mode.
- **Clash** is *Destiny*'s equivalent to the well-known multiplayer game mode Team Deathmatch. In this 6v6 game mode, two teams battle for points gained by kills and assists. The first team to reach the point limit, wins. *Destiny* also offers another variation of this mode, called *Inferno Clash*. In *Inferno Clash* players cannot use their radars or trackers and are only awarded points for kill – not for assists.
- **Control** is another two-team 6v6 game mode in *Destiny* and a blend of the game modes Team Deathmatch and Domination. Points are awarded for capturing or neutralizing zones, as well as for kills and assists. Again, a variation of this game mode, named *Inferno Control* is offered, with no trackers or radars, and points being awarded only for kills. Furthermore, in *Inferno Control*, capturing zones increases the points gained per kill.
- **Mayhem** can be seen as a kind of mutator for the *Clash*, *Control*, and *Rumble* game modes. It too provides the two-team 6v6 gameplay that

Clash offers. Compared to *Clash*, in *Mayhem*, cooldowns for grenades, melee and super abilities are vastly reduced.

- **Supremacy** has players competing against each other in two teams of six players each (6v6). Unlike the game modes described so far, in *Supremacy* players need to confirm their kills by picking up Crests dropped by enemies on death. Points are awarded for picking up the Crests and can be denied by picking up your teammates' Crests before your opponents do. Again, there exists a variation of this game mode called *Inferno Supremacy* with no trackers and radars.
- **Rift** has players compete in two six-player teams (6v6). Similar to the well-known multiplayer game mode Capture the Flag, players have to grab a *Spark* and take it to their enemies' Rift. Points are awarded for carrying the *Spark* and killing enemies – especially the *Spark* carrier.
- **Skirmish** is a game mode for two teams of two or three players each (2v2, 3v3). It is similar to Team Deathmatch and – unlike the game modes covered so far – allows players to revive their teammates. Besides gaining points by killing or assisting in a kill, players also gain points for resurrecting teammates. A round ends when a team eliminates the opposing team, and a game ends when a team reaches the point limit. Additionally, there is the variation, *Inferno Skirmish*, with no trackers/radars and points only being awarded for kills.
- In **Salvage** players aim at probing relics and keeping their opponents from doing the same. Players play in two teams of three (3v3) and earn points by means of kills, assists and revives but also for probe-related tasks. These include deploying a probe, neutralizing and salvaging it, assaulting a probe (kill a defending enemy) and defending a probe (kill an enemy attacking your probe). In addition, there also is an *Inferno Salvage* game mode with radars and trackers disabled and points only awarded for kills.
- **Elimination** is another game mode for six players, with players being split up into two teams (3v3). A round is won, when the entire enemy team is defeated. No respawns are allowed. This game mode can be seen as a training ground for the *The Crucible* event *Trials of Osiris*. This event mode can only be entered at specific times during weekends and stat normalization is disabled. Every week, players compete on a different map. A game of *Trials of Osiris* ends when a team won five rounds in the way described above.

As seen from these descriptions, most of the PvE game modes can and almost all of the PvP game modes have to be played in teams. As *Destiny* relies so heavily on team play, one has to wonder why no in-game mechanism exists for creating and managing lists of friends to play with or form guilds. In light of the lack of in-game friend lists, looking for group (LFG) sites have emerged to fill that gap. One such site is `the100.io`[5].

3.5. Summary

In this chapter, the popular FPS game *Destiny* and its key features were introduced and briefly described. First, the gameplay was described. From this point of view, *Destiny* is a FPS-MMORPG blend. For example, as common in RPGs, it allows the player to create and control an in-game avatar, a *Guardian*, for which a class and, furthermore, a sub-class can be chosen. The main progression in the game is leveling up a character by gathering experience and endgame progression is achieved by increasing the *Guardian*'s *Light level* by means of finding and equipping better gear. After briefly describing the game's setting, the class system offering three *Guardian* classes, *Hunter*, *Warlock*, and *Titan* was described in greater detail. Next, the expansions *Destiny* received after its initial launch, were described. On this note, it should be mentioned that the expansions are especially important as they introduced new game modes, and increased the *Light levels* players could achieve. In the following section, the game modes *Destiny* offers were described. These can be divided into PvE and PvP. After this broad classification, the specific game modes available in each of these categories were introduced. While almost all PvE activities can be played as a singleplayer game, it is worth noting that most of them can be played cooperatively and the majority of PvP game types requires fireteams consisting of at least three players. Finally, the absence of in-game matchmaking facilities in *Destiny* was discussed. Related to this lack of functionality, LFG sites, such as, `the100.io` tried to fill the gap. LFG sites allow users to form groups and schedule games.

[5]https://the100.io. For further information, see section 4.1.

4. Datasets

This chapter describes the datasets used for the thesis as well as their basic information. First, the source of each dataset and the means of extracting the data are described. Following this, key metrics about the information contained in each dataset are described. Finally, the features contained in each dataset are discussed in greater detail.

4.1. Dataset "the100.io"

As mentioned in Chapter 3, for most activities in *Destiny* teammates are required. The game itself, though, does not offer any in-game matchmaking facilities. Therefore, looking for group (LFG) sites stepped up to fill this niche to try and help players find teammates. While most such matchmaking sites focus on instant matchmaking and try to find ephemeral groups to play with right now, `the100.io`[1] uses a different approach. The main idea is to have long-lasting, guild-like groupings that allow for more consistent gameplay and an overall better game experience. After entering basic data, such as age, sex, platform, preferred playtime and timezone, the player is assigned to a group matching these criteria. Besides this automatic group association, players can also choose to enter different groups from the one they are initially assigned to.

[1] https://www.the100.io

4.1.1. Overview

The first dataset used in this thesis was extracted from this LFG site, `the100.io`, using its application programming interface (API) and a Python script. The data was collected on December 16th 2016 and contains information up to this point. Key features of the data collected can be seen in Table 4.1.

`the100.io` - **Features**	
Players	218, 214
Friendships	273, 688
Groups	2, 468
Number of Moderators	2, 522
Number of Sherpas	4, 374
Games	637, 823

Table 4.1.: Key Features of the `the100.io` Dataset

From this table, it can be seen that more than 200,000 players in roughly 2,500 groups scheduled over 630,000 games from the launch of `the100.io` up to December, 16th 2016. These groups contained more than 2,500 moderators and over 4,000 sherpas, that is, players willing to help other players. Furthermore, users of `the100.io` formed more than 270,000 friendships with other players on the platform.

4.1.2. Feature Description

Besides the overall key features seen in Table 4.1, further characteristics are available from the dataset. In this regard, each player is given the opportunity to enter demographic and basic game experience information. This information includes the following:

- **Age** – The player's self-reported age.
- **Gender** – The player's self-reported gender.
- **Timezone** – The player's self reported timezone.
- **Profanity** – The player's attitude towards profanity. It takes on one of these values: "OK with profanity", "some profanity", "no profanity".
- **Preferred Platform** – One of Xbox 360, Xbox One, PlayStation 3 (PS3), PlayStation 4 (PS4), and PC.
- **Play Style** – The player's approach to the game. A player's play style is either *serious* ("Getting things done"), or *casual* ("Having fun with the game").
- **Preferred Playtime** – The player's preferred game schedule. The options offered to the players are "weekday mornings and weekends", "weekday afternoons and weekends", "weekday evenings and weekends", and "weekday late-nights and weekends".

Besides these self-reported features and information about players' preferences, players in the dataset also bear these in-game characteristics:

- **Level** – The highest level of any of the player's characters since a player may have multiple characters in *Destiny*.
- **Light Level** – The player's characters' *Light level* is a trait derived from the gear equipped. Higher *Light levels* increase a character's damage output and mitigate incoming damage to the player.

Furthermore, the dataset also offers insights into platform-related information, that is, metrics associated to the player's activities on `the100.io` itself. These are:

- **Karma** – Players can reward other players with *karma*. This is usually done after playing with them and rating the experience pleasant or the player particularly helpful or friendly. Then, `the100.io` calculates a score from the *karma* rewarded.[2]
- **Sherpa Score** – The *sherpa score* is another inherent measure available on `the100.io`. In the context of *Destiny*, a *sherpa* is a player who actively seeks to help other players on `the100.io`, as well as in-game. The *sherpa*

[2]For further information see "What is Karma and How do I give it?": https://web.archive.org/web/20190714151050/https://the100io.zendesk.com/hc/en-us/articles/208927887-What-is-Karma-and-How-do-I-give-it-, Archived version from July, 14th 2019.

score is calculated by `the100.io` using the number of activities a player participated in and helped other players.[3]
- **Friends** – Users of `the100.io` may connect to other players by way of sending friend requests. Some information is already visible after sending a request – even for unconfirmed friendships.
- **Number of Groups** – The number of groups the player is a member of.
- **User Active Game Count** – The sum of recent games, as well as planned impending games as they are listed on the player's profile page. In contrast to the *user activity score*, which measures overall activity, this number indicates activity at the time of gathering the sample.
- **User Activity Score** – A figure denoting a player's overall activity on `the100.io`. This can be increased by creating game sessions, joining them, as well as by inviting players to join `the100.io`.[4]

Lastly, the dataset also contains group-level information that relates to a group's actions on the `the100.io` site itself. These measures are:

- **Group Size** – The number of players that are members of a group.
- **Play Style** – A group, like a user, may be classified as either *serious* or *casual* - this labelling was extracted from `the100.io` itself and not done during the analysis presented in this thesis or the published papers. For more information, see user-related *Play Style* above.
- **Number of Moderators** – The number of moderators a group contains. Players can only become moderators by reaching a defined level of activity, that is, a threshold of *activity score* which increases as the group matures. After that, they can manage the group by creating games, posting news and kicking players that have been reported. It is worth noting that – while a group usually may only have three

[3]For further information see "Game Session Types", specifically "Sherpa Requested", and "Beginners Welcome": https://web.archive.org/web/20190713171749/https://the100io.zendesk.com/hc/en-us/articles/215007698-Game-Session-Types, Archived version from July, 13th 2019.

[4]For further information see "Activity Score and how to increase it": https://web.archive.org/web/20190713165716/https://the100io.zendesk.com/hc/en-us/articles/208292158-Activity-score-and-how-to-increase-it, Archived version from July, 13th 2019.

moderators – supporters, that is, paying members of `the100.io`, may choose to promote others to the rank of *moderator* after becoming moderators themselves.[5]

- **Number of Sherpas** – The number of *sherpas* that are members of the group. For further information about *sherpas*, see user-related measure "Sherpa Score" above.
- **Group Activity Score** – The *activity score* of a group, similarly to the *user activity score* reflects a group's activity. In contrast though, the *group actity score* is calculated based on the most recent 14-day period, thus decays over time. It can be increased by a group's members creating or joining games and by inviting players to `the100.io`.[6]
- **Group Active Game Count** – Similar to the *user active game count* described above, this is a figure representing the number of recently completed, as well as planned upcoming games. Again, as this measure depends on sessions happening during a limited time frame, it is sensitive to the point in time of sampling.

4.2. Dataset "Destiny"

The next dataset was extracted from the Bungie.NET API[7]. The original structure of the data is a single file containing the game data, each line in the file representing a single match as a `JSON` object. One aspect worth mentioning is the amount of superfluous and redundant information also contained in this initial representation of the data.

[5]For further information see "How do I become a mod?": https://web.archive.org/web/20190713170407/https://the100io.zendesk.com/hc/en-us/articles/208291098-How-do-I-become-a-mod-, Archived version from July, 13th 2019.

[6]For further information see "Group Activity Score": https://web.archive.org/web/20190713165501/https://the100io.zendesk.com/hc/en-us/articles/360020044351-Group-Activity-Score, Archived version from July, 13th 2019.

[7]https://bungie-net.github.io/multi/index.html

4.2.1. Overview

The data contained in this dataset encompasses full information of over 900,000 matches played in *Destiny* in a time span between September, 8th, 2014 and January, 12th 2016.

Destiny In-Game - Features	
Matches	930,721
Observations	10,387,020
Players	3,450,622
Characters	4,487,458
Kills	116,123,023
Clans	320,278
Game Modes	13
Points Scored	24,615,436,442
Time Played	over 185 years

Table 4.2.: Key Characteristics of the Destiny Dataset

From Table 4.2 it can be concluded that *Destiny* is a popular game with players spending an aggregate play time of more than 185 years in a time span of less than one and a half years. The matches captured contained over 10 million unique *Guardian* in-game performances, included 13 different game modes (as described in subsection 3.4.2) played by over 3,400,000 unique players in over 320,000 clans. During their matches, the players accumulated over 100 million kills and scored, in aggregate, over 24.5 billion points.

4.2.2. Feature Description

One level of analysis supported by the data is the *match level*, that is, a single match in *Destiny*'s player-versus-player (PvP) mode, *The Crucible*. In order to support this level of analysis, the dataset contains the following information:

- **Game Mode** – The game mode of the match played. This is one of the game modes described in section 3.4
- **Period** – The point in time at which the match started.
- **All Participants Count** – The number of players that played in the match. This also includes players that left early and those who only joined while the match was already in progress.
- **All Participants Time Played** – This is the aggregate time that all the participants combined spent in the match.
- **All Participants Score** – The aggregate number of points scored during the match.

As all the matches gathered in the dataset are games in *The Crucible*, that is, PvP, the game modes mostly focus on team play. As such, the dataset also contains data at the team-level, comprised of the following features:

- **Standing** – The result of the match, for example, "Victory", "Tie", "Defeat".
- **Team Name** – The name of the team. This is in most cases *Alpha* or *Bravo*.
- **Team Score** – The overall score the team managed to accumulate during the match.

Furthermore, the dataset contains information about every player's performance within the confines of a single game. These are too numerous to list here in their entirety, so the following shortened list will have to suffice:

- **Activity Duration Seconds** – The number of seconds the player spent in this match.
- **Score** – The amount of points the player was able to gather during the match.
- **Character Level** – The level of the player's *Guardian* at the time of the match.

- **Light Level** – The *Light level* of player's *Guardian* at the time of this match.
- **Kills** – The number of kills the player achieved.
- **Deaths** – The number of deaths the player experienced during the game
- **Assists** – The number of "assists" a player performed. An *assist* in this context, is lowering an enemies health with a teammate ultimately scoring the kill.
- **Kill-Death Ratio** – The kill-death ratio (KD) is a measure of in-game performance and depends on the total amount of kills and the total amount of deaths. It is then defined as $KD = \frac{\text{kills}}{\text{deaths}}$.
- **Kill-Deaths Assists** – Similar to the KD, the Kill-Death-Assists (KDA) ratio is a metric describing in-game performance. In contrast, this measure also takes assists into account. It is defined as $KDA = \frac{\text{kills + assists}}{\text{deaths}}$.
- **Longest Kill Spree** – The longest consecutive series of kills the player managed to perform without dying.
- **Longest Single Life (Seconds)** – The longest time span the player was able to survive in the match.
- **Kills Of Player {Hunter, Titan, Warlock}** – The player's kills of opponents – grouped by the opponents' classes.
- **Precision Kills Of Player {Hunter, Titan, Warlock}** – The number of precision kills, that is, head shots or critical hits, the player was able to gather during the match – grouped by the opponents' classes.
- **Deaths From Player {Hunter, Titan, Warlock}** – The player's deaths caused by opponents – grouped by the opponents' classes.
- **Weapon Kills {Auto Rifle, Fusion Rifle, Grenade, Hand Cannon, Machinegun, Melee, Side Arm, Rocket Launcher, Pulse Rifle, Scout Rifle, Sniper, Shotgun, Super, Sword, Relic}** – The number of kills the player managed to gather during a match – grouped by weapon type and ability, in the case of *Super* abilities.
- **Weapon Kills Precision Kills {Auto Rifle, Fusion Rifle, Hand Cannon, Machinegun, Pulse Rifle, Rocket Launcher, Scout Rifle, Shotgun, Side Arm, Sniper}** – The number of precision kills, that is, head shots or critical hits the player landed over the course of the match. Note that not all weapons that can be used for kills also allow for precision kills.

- **Medals** – The number of medals awarded for the player's performance during the game. Since the medals are too numerous – 104 different kinds of medals are contained in the dataset – and highly situational as they are also closely linked to the game type played in a match, they are not described in detail here.

Albeit, this dataset is already rich in information, it has been further extended and augmented, as will be described in the next chapter.

4.3. Summary

In this chapter, the two datasets used for this thesis were introduced. The first dataset discussed was extracted from `the100.io` and contains data of over 200,000 players that used this LFG platform. These players formed over 270,000 friendships, are members in roughly 2,500 groups. Furthermore, over 4,000 players, or roughly 2% of players were willing to help other players in-game, as well as on `the100.io` by taking on the role of *sherpas*. Besides these metrics, the dataset contains player-level information, such as demographic information, group level details, such as, group sizes, number of sherpas and moderators, as well as game-related data that are available from `the100.io` which were discussed and their meaning and calculation were described. The second data set described in this chapter was scraped directly from Bungie's API and contains full information about over 900,000 matches from *The Crucible*, *Destiny*'s PvP game mode. These matches consist of over 10 million observations of in-game performances of roughly 3.5 million players in approximately 300,000 clans. During these matches, players scored over 100 million kills and an aggregate of more than 24 billion points. This second dataset strictly contains in-game data and a wide range of information about players' class selections and *Guardian*s, as well as information about matches, that is, game sessions, their results and team results if applicable. Lastly, also player-level performance indicators, such as, medals, kills, assists and deaths, but also objective-related metrics are contained in the dataset and were described in this chapter.

5. Data Preparation and Processing

In this chapter, preliminary data preparation steps necessary for further analyses will be presented. The datasets used as an input for the preprocessing described here were discussed at great length in Chapter 4. In short, the two datasets processed were extracted from the looking for group (LFG) website `the100.io`[1] and from Bungie itself, using its application programming interface (API). The latter was further augmented using engagement data from *DestinyTracker*[2].

5.1. Overview

Figure 5.1 depicts the steps taken during the preprocessing phase that were necessary to handle and analyze the data efficiently and effectively. The highlighted parts of the processing pipeline are in the scope of either this thesis or the analyses already published by Schiller et al. (2018), Wallner et al. (2019). The left part of the graphic shows the processing steps performed on the *Destiny* dataset, while the right part shows the processing done on the `the100.io` dataset. It is worth mentioning that the two datasets different preparation steps but were ultimately stored as a `.hd5` file in the hierarchical data format (HDF). The benefits of doing this will be described briefly in the following section.

[1]https://the100.io
[2]https://destinytracker.com/d1

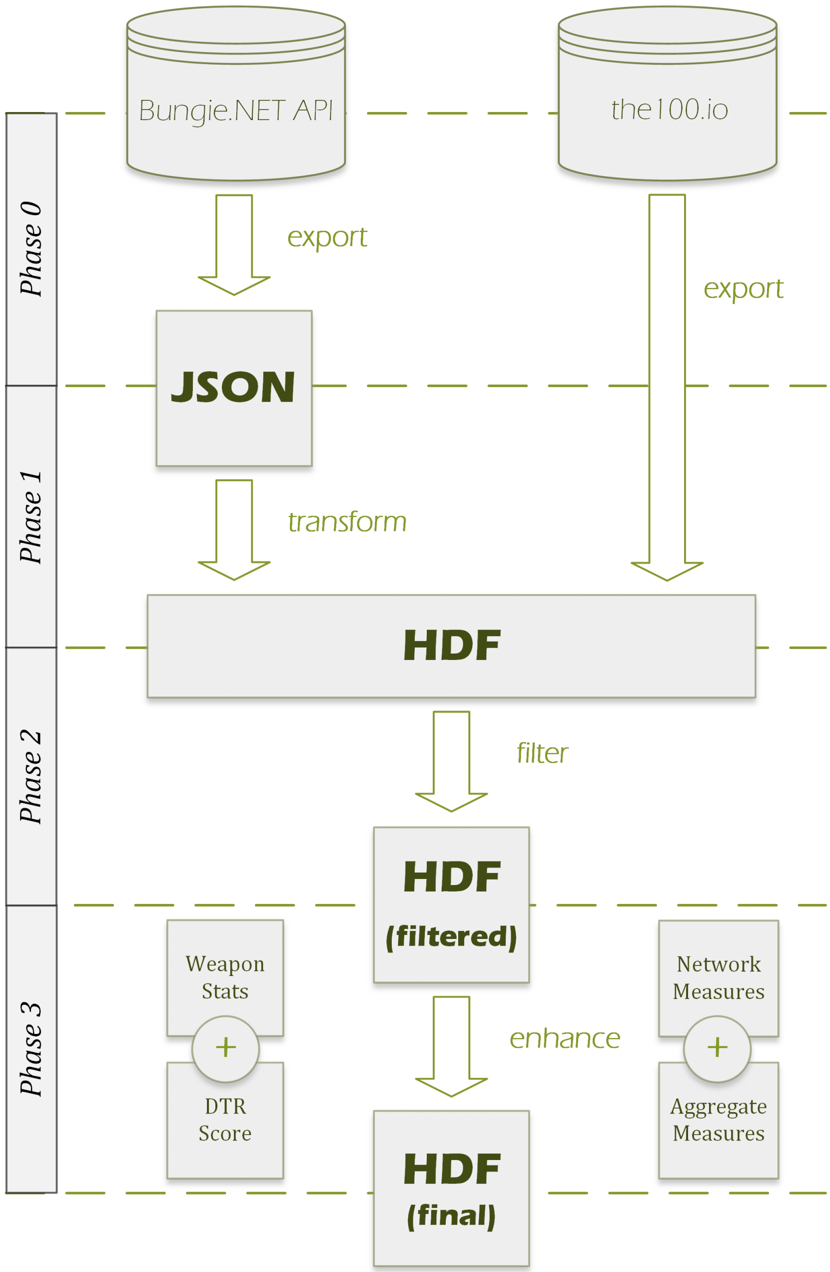


Figure 5.1.: Data Preprocessing Procedure

The HDF Format

The HDF data format is maintained and developed by *The HDF Group*. As its name suggests, it is a hierarchical file format built from groups and datasets. While groups may contain datasets and other groups themselves, datasets are best understood as multidimensional arrays. From this structuring follow a number of benefits, including the handling of very large datasets while maintaining fast access times. Furthermore, this file format is supported in a wide range of programming languages, including C++, Python and R.[3]

5.2. Processing the Destiny Dataset

In this section, the processing of the *Destiny* dataset extracted using the Bungie.NET API will be described. The file contains information about 930,721 matches, each one as a single line – amounting to a 54.1GB large, unparsable `JSON` file. First, the steps necessary for the *Filter & Transform* phase will be discussed and then the resulting HDF file and its structure are described. Lastly, the process of collecting an engagement metric, its meaning and the resulting file structure are described.

5.2.1. Phase 1: Transform

As mentioned previously, at the start, the dataset was stored as a very large text file which was too unwieldy to work with. Initial tests showed that processing the file, that is reading and parsing, once in this state would take between six and eight hours on modern computer hardware with sufficient main memory. This was unacceptable. For this reason, ways to reduce processing times were derived. As a first step, common sections in the data were extracted and the dataset was normalized, meaning, redundancy was reduced and integrity was increased. Following this procedure, the

[3]For further information see "What is HDF5?": https://web.archive.org/web/20170223010222/https://support.hdfgroup.org/HDF5/whatishdf5.html, Archived version from February, 23rd 2017.

JSON-based data container was transformed into the aforementioned HDF format.

5.2.2. Phase 2: Filter

As the file is fundamentally a dump of responses received from the API[4] of Bungie.NET, it contains status codes and error messages, as well as throttle notifications that only relate to the usage of the API and not the transmitted data itself. After confirming that none of the entries were malformed or incomplete – all entries contained `"ErrorCode": 1, "ErrorStatus": "Success", "Message": "Ok"` – these parts were stripped from the data. Furthermore, the dataset contained a variety of different object hashes that could not be reconstructed using the API and, thus, were removed. These included JSON objects, such as shown in Listing 5.1.


```
{
  "dyes": [
    {
      "dyeHash": 2797525833,
      "channelHash": 662199250
    },
    {
      "dyeHash": 1441129912,
      "channelHash": 1367384683
    },
    {
      "dyeHash": 4116939015,
      "channelHash": 218592586
    }
  ],
  "itemHash": 1054763959
}
```


Listing 5.1.: Hash-only Object

Additionally, some of the entries in the dataset only contained file paths to assets, for example for background images. For lack of knowledge how such

[4]https://bungie-net.github.io/multi/index.html

data could be interpreted, these structures were removed as well. Next, the measure *fireTeamId* was removed due to it being either empty or `INT_MIN` ($-2, 147, 483, 648$), thus, conveying no obvious meaning. Additionally, the entry *membershipType* was removed from the *Bungie.NET User Information*, as each and every player had a value of 254. Moreover, entries for medals earned by players, as seen in Listing 5.2, contained the amount of medals earned of the specific type, as well as the amount of points earned for this number of medals. This data was checked for consistency and weights were extracted and stored separately.

```
"medalsKillSpree1": {
  "weighted": {
    "displayValue": "200",
    "value": 200.0
  },
  "basic": {
    "displayValue": "2",
    "value": 2.0
  }
}
```

Listing 5.2.: Medal Entry with Weights

Handling Player Data

As a next step, issues regarding the redundancy in player information were tackled. An example of redundancy and how it was reduced can be seen in Listing 5.3a. From this example, it can be observed that each contained measure has both a numeric `value` and a `displayValue`. After inspecting the data, for each measure either one type of value or the other was chosen, resulting in objects similar to the one shown in Listing 5.3b, vastly reducing redundant information and, therefore, file size.

The numeric `value` was chosen for trivial metrics, such as, *Score*, *Kills*, *Average Score Per Life*, *Deaths* and *Activity Duration Seconds*, while the string representation, `displayValue`, was used for non-trivial measures, for example, *Team* and *Completion Reason*.

```
{
  "activitiesPerChar": [
    "activityId": {
      "values": {
        "STAT1" : {
          "statId" : "STAT1",
            "basic": {
              "displayValue": ...,
              "value": ...
            }
        },
        "STAT2" : {
          "statId" : "STAT2",
          "basic": {
            "displayValue": ...,
            "value": ...
          }
        }, ...
      }
    }
  ]
}
```

(a) Full JSON Format

```
{
  "activitiesPerChar": [
    "activityId": {
      "values": {
        "STAT1" : [value],
        "STAT2" : [displayValue],
        ...
      }
    }
  ]
}
```

(b) Reduced JSON Format

Listing 5.3.: User Data

Handling Match Data

Similarly, the match-level data was processed. An example of the original `teams` array can be seen in Listing 5.4a. Again, the prevalence of redundant information is noticeable. After processing, the resulting `JSON` array is given by Listing 5.4b. The string representation, `displayValue`, was used to encode non-trivial data, such as, *Standing* and *Team Name*, while it was dropped in favor of the numeric *value* for the measures *Team ID* and *Score*.

As most other measures also contained the format shown in Listing 5.5, either one or the other of these representations was chosen to further denote the measure.

5.2.3. Phase 3: Enhance

Following the, *Filter & Transform* stage described so far, the dataset was enhanced by reconstructing data that was lost due to only being encoded by hash values and IDs, as well as by adding an engagement measure from the site *Destiny Tracker* to the player information.

Reconstructing Information

As a next step, information about the used weapon type was reconstructed from the dataset. An example of how the weapon statistics looked after the preprocessing can be seen in Listing 5.6. Note that there is no way of knowing the type of weapon this relates to, as the weapon is only identified by the *referenceId*. This issue, again, is linked to items encoded as IDs and hashes which, on their own, convey little to no meaning and contain no further information.

Accordingly, a statistical approach for reconstructing this missing information was chosen. By analyzing *kill* and *precision kill* statistics and cross-referencing awarded medals, it was possible to infer and uniquely identify the weapon type of each weapon described by its *referenceId* contained in the dataset. Furthermore, this information was extracted and stored in the newly created HDF dataset *Weapon Stats*.

```
"teams": [
  {
    "standing": {
      "basic": {
        "displayValue": "Defeat",
        "value": 1.0
      }
    },
    "teamId": 16,
    "score": {
      "basic": {
        "displayValue": "19075",
        "value": 19075.0
      }
    },
    "teamName": "Alpha"
  },
  {
    "standing": {
      "basic": {
        "displayValue": "Victory",
        "value": 0.0
      }
    },
    "teamId": 17,
    "score": {
      "basic": {
        "displayValue": "20195",
        "value": 20195.0
      }
    },
    "teamName": "Bravo"
  }
],
```


(a) Full JSON Format


```
"teams": [
  {
    "standing": "Defeat",
    "teamId": 16,
    "score": 19075.0
    "teamName": "Alpha"
  },
  {
    "standing": "Victory",
    "teamId": 17,
    "score": 20195.0
    "teamName": "Bravo"
  }
],
```


(b) Reduced JSON Format

Listing 5.4.: Match Data

```
{
  "basic":
  {
    "value": ...,
    "displayValue": ...
  }
}
```

Listing 5.5.: Original Variable Encoding

```
{
  "referenceId": 4091910568,
  "values": {
    "uniqueWeaponKillsPrecisionKills": 0.5,
    "uniqueWeaponPrecisionKills": 2.0,
    "uniqueWeaponKills": 4.0,
  }
}
```

Listing 5.6.: Weapon Statistics

Augmentation with the Destiny Tracker (DTR) score

As a next step, the dataset was enriched with a more generalized performance measure, DTR score. This score is calculated from in-game measures such as kills, deaths, assists and medals. Furthermore, DTR takes into account statistics from all in-game characters, *Guardians*, a player may have. Thus, it combines both, performance and retention, into one measure. Since this measure had to be retrieved for all 3,450,839 *Destiny* users, it had to be retrieved by leveraging high levels of parallelism. Also, it had to be taken into account that there is no publicly available API for retrieving the DTR score, forcing the use of the query functionality on the site and scraping values from the page if it was found. Furthermore, also related to the lack of a public API, it was necessary to handle timeouts as there was no way to reasonably enforce rate limiting to the querying of the pages. For these reasons, it was necessary to use a MapReduce architecture – running in multiple processes and threads, storing partial results and restarting worker threads when they were stuck in a timed out state. After finishing this task, re-checking irretrievable DTR scores multiple times, a total of 2,720,785 valid DTR scores could be recovered – out of 3,450,839. These were then stored in their corresponding *Destiny User* data structures.

5.2.4. Resulting Data Model

After processing the data as described in this chapter so far, applying the steps in the phases *Transform*, *Filter* and *Enhance*, the resulting data was stored back into a HDF container to a number of distinct HDF datasets. At this point, it is worth noting, that the HDF format does not enforce key rules as relational database management systems do. For this reason, primary keys and foreign keys shown in Figure 5.2 are only to indicate how rows in each dataset can be uniquely identified and how entries are interconnected, respectively. From the visualization, it can already be observed that information initially stored at a player in-game observation level, such as, *Destiny User* and *Bungie.NET User* data, was extracted and stored separately, thus reducing redundancy and file size. Note, that the *Destiny User* dataset also contains the DTR scores scraped during the *Augmentation* step. Furthermore, match-level information was extracted to a dedicated *Match Detail* dataset containing information such as *Number of Participants* and *Game Mode* played. Additionally, and since most matches contained in the dataset are game modes that are played in teams, a dedicated *Team Results* dataset was extracted from the *Game Detail* dataset. The *Medal Weights* dataset contains the score associated with earning a medal for each type of medal contained in the dataset. Lastly, in the dataset *Game Details*, each row represents the performance of a single player during a player-versus-player (PvP) match in *The Crucible*. Besides the usual in-game metrics, such as, *kills*, *deaths* and *score*, it also contains information about the number of medals awarded of each type of medal.

Taking into account all the steps described in this section, the dataset's size on disk could be reduced from 54.1GB to 2.8GB – that is a reduction of roughly 95%, consequently also speeding up processing times significantly.

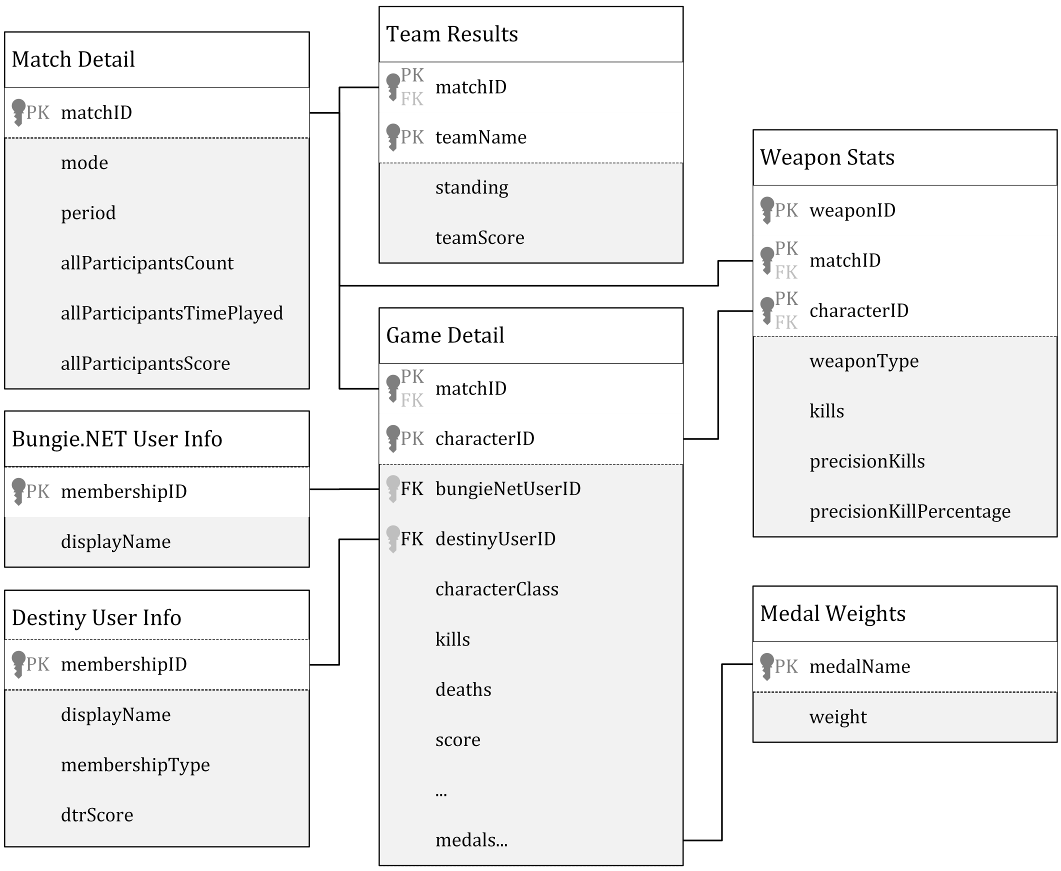


Figure 5.2.: *Destiny* Dataset: Data Model

5.3. Processing the `the100.io` Dataset

Besides the *Destiny* dataset described so far, the `the100.io` dataset was preprocessed in order to be useful for analysis as well. In this section, the preprocessing steps performed on this dataset will be described. It is worth noting that this dataset was already stored in the HDF format, rendering *Phase 1: Transform* unnecessary.

5.3.1. Data Model

As mentioned above, the `the100.io` dataset already was stored as a `.hd5` HDF file to begin with. This fact made the steps *Phase 0 - Export* and *Phase 1 - Transform* unnecessary. The overall structure of the dataset is shown in Figure 5.3. The two main containers are *User Info* and *Group Info* with a number of dataset tables connecting them with each other and to additional data. The *User Info* dataset contains information about self-reported characteristics, such as, *age* and *gender*, platform-related measures, such as *activity score* and *sherpa score* and, finally, in-game information, such as the *gamer tag*, *level* and *Light level*. The *Group Info* dataset contains information, such as, its *activity score*, the style of play – serious or casual as its *group type*, as well as the platform on which the group is playing. In regards to the auxiliary datasets, the *Friend Info* contains a row of two user IDs if at least one of the parties initiated a friend request. Furthermore, a `pending` flag indicates whether the receiving party has accepted the friend request. The *Group-User* dataset expresses group membership of `the100.io` users, the *Group-Moderator Info* table connects players to the groups which they are moderators of. Similarly, *Group-Sherpa Info*, connects players to the groups which listed them as *sherpas*. *User-Game* and *Group-Game Info* tables connect games listed at the profile page to their listing users and groups, respectively. Lastly, the *Group-Title* info links groups with the game titles they are playing according to their profile, for example, Destiny[5], or Titanfall 2[6]. As with the data model seen in Figure 5.2, it is important to note that neither primary nor foreign key relationships and conditions are

[5] Bungie, 2014. https://www.destinythegame.com/d1.
[6] Respawn Enterntainment, 2016. https://www.ea.com/games/titanfall/titanfall-2.

enforced by HDF, thus, they are only marked to show the connections between the datasets and to indicate uniqueness of rows.

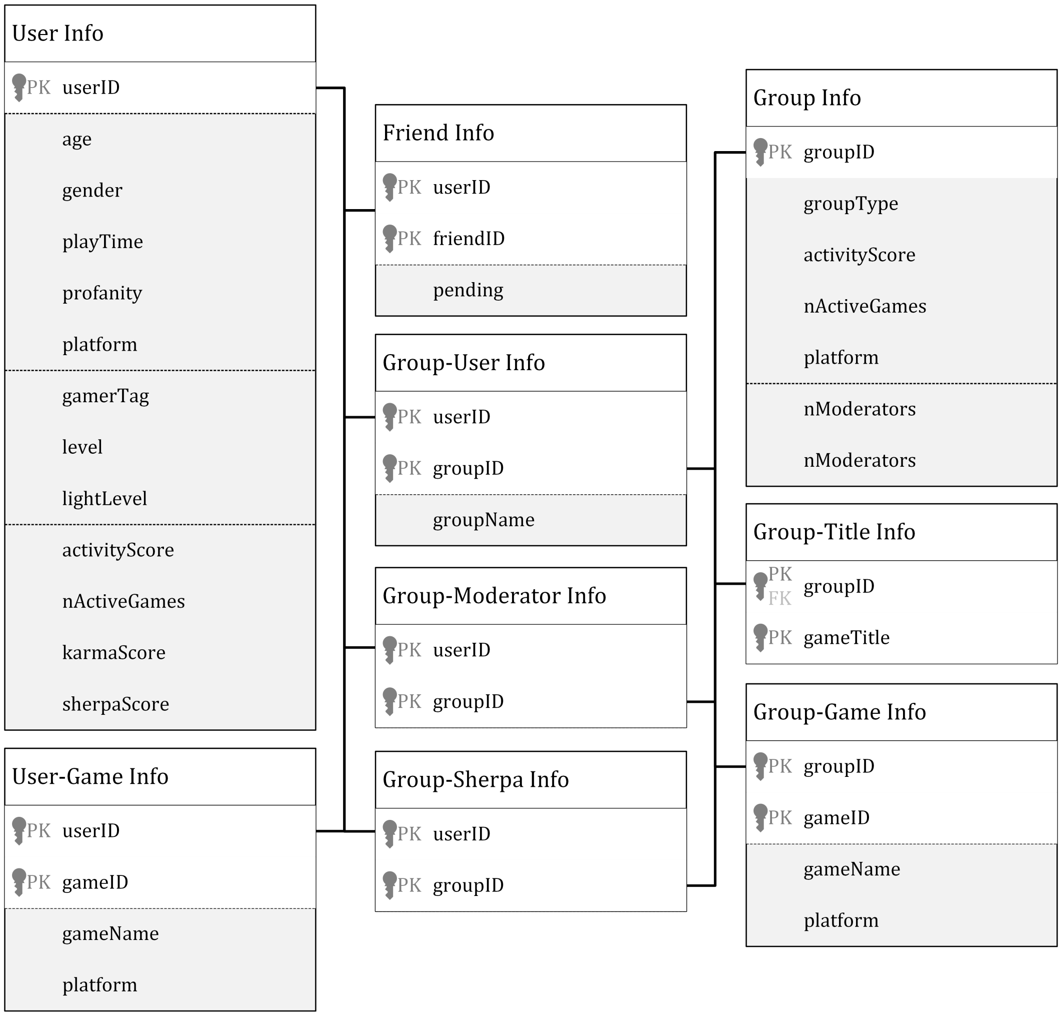


Figure 5.3.: `the100.io` Dataset: Data Model

5.3.2. Phase 2: Filter

The first step in preprocessing the `the100.io` dataset was *Filter*. In this step, mainly numerical values used for correlation were cleaned. In this regard, missing *Sherpa scores*, denoted as -1 were replaced by 0 to be usable for correlation. Furthermore, entries with invalid values, that is, players with levels or *Light levels* higher than the maximum possible or lower than the minimum were removed. These values may arise from issues in either storing or retrieving the data. Furthermore, players with invalid values for *karma* and *activity scores* were removed. Next, entries with implausible values for age, such as, lower than 0 or exactly 99 – the maximum possible, were removed. As a next step, numerical entries that could lead to spurious correlations such as IDs and the `supporter` flag indicating a paid membership with `the100.io`, were removed, therefore, also not taken into consideration.

5.3.3. Phase 3: Enhance

After the *Filter* step described earlier, the dataset was augmented by constructing additional measures. These additional measures, which will be described in the following part, encompass network measures, as well as group-level measures calculated from the members of a group.

Network Measures

As a first step, a graph was constructed from friendships as expressed on `the100.io`. Here, it is worth noting that friendships on `the100.io` can be either *pending* or not – with the latter being the final state after a player has accepted another player's friendship request. Since some of the information between players is exchanged already upon sending a friendship request, it is not entirely clear whether there is enough incentive to accept a friendship request, thus, marking it *not pending* or *accepted*. In either case, information about friendships being either *pending* or *accepted* was conserved for potential later use. On the resulting *player* graph a number of network measures were computed for each of the nodes representing a player. These network measures included the following measures which are defined on the graph's vertices in V:

- **Degree Centrality** defines a node in the graph to be more central, the higher it degree is. Thus, the degree centrality $C_D(v)$ of a node v can be expressed as
$$C_D(v) = \deg(v). \tag{5.1}$$
- **Closeness Centrality** $C_C(v)$ measures the average length of shortest paths of a node to all other nodes. Nodes with higher values for their closeness centrality are defined to be more central (Bavelas, 1950). The closeness centrality of a node v is then defined as
$$C_C(v) = \frac{1}{\sum\limits_{u \in V \setminus \{v\}} d(v,u)} \tag{5.2}$$
with $d(v_1, v_2)$ denoting the path length of the shortest path between the two vertices v_1, v_2.
- **Farness Centrality** $F(v)$ is the reciprocal of the closeness centrality. Thus, for a node v it can be written as
$$F(v) = \sum_{u \in V \setminus \{v\}} d(v,u). \tag{5.3}$$
- The **Eccentricity** $\epsilon(v)$ of a node v is defined as the maximum distance between the node an all other nodes in the graph and can be written as (Harary, 1999)
$$\epsilon(v) = \max_{u \in V} d(v,u). \tag{5.4}$$
- **Betweenness Centrality** β was first established by Freeman (1977). It can be understood as a measure of the importance of a node v in efficiently connecting other nodes s, t in the network. To formally describe the betweenness centrality, we first define the number of shortest paths between the nodes s and t as σ_{st} and the number of shortest paths between s and t *passing through* v as $\sigma_{st}(v)$. The betweenness centrality of v is then defined as
$$\beta(v) = \sum_{\substack{s \in V:\ s \neq v \\ t \in V:\ t \neq v}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{5.5}$$

- For **Eigenvector Centrality**, the main idea is that a node's centrality can not accurately be described by only looking at the node itself, but by also taking its neighbors into account (Bonacich, 1972). Using the adjacency matrix $\mathbf{A}$, the centralities for all nodes can be calculated as follows:
  1. Compute the principal, that is, largest eigenvalue λ^* of $\mathbf{A}$.
  2. Compute the eigenvector corresponding to the eigenvalue λ^*:

  $$\lambda^* \mathbf{x} = \mathbf{A}\mathbf{x} \tag{5.6}$$

  3. The eigenvector centrality of the i-th node, v_i (corresponding to the i-th column and row in $\mathbf{A}$), can then be read off the i-th component of the eigenvector $\mathbf{x}$.

Group-Level Measures

As mentioned earlier, also group-level information was constructed from per-user data or the network measures described in detail above. These measures included

- **Average Age** The average age of `the100.io` users in a group was computed.
- **Average Level** The average level of players' *Guardians* in the group was computed.
- **Average Light Level** Similarly, also the average *Light level* of the players' *Guardians* was computed.
- **Median {Age, Level, Light Level}** To mitigate the influence of outliers, also median ages, levels and *Light levels* were calculated on the group level.
- **Group Density** The density of the subgraph representing all members of the group.
- **Clustering Coefficient** The *global* clustering coefficient of the group subgraph.
- **Closeness Centrality** For both, *moderators* and *sherpas*, the average closeness centralities were computed from the group subgraphs.

- **Connectivity** For nodes of the types *moderator* and *sherpa*, the average connectivity in a group of n players for m *sherpas*, or moderators, respectively, was computed as

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\deg(v_i)}{n-1}. \tag{5.7}$$

5.4. Summary

In this chapter, the preprocessing done on the two datasets, the first directly extracted from the Bungie.NET API and the second extracted from `the100.io`, were described in great detail. A unified process for describing the preprocessing was introduced and discussed. This procedure ends with the data being stored in the HDF file format. The HDF format was briefly introduced and the benefits of this file format, scilicet, a small file size on disk and fast access and processing times were established. Next, the preprocessing procedure – as it relates to each dataset – was described. In this regard, the transformation of the *Destiny* dataset from `JSON` into HDF was described, with the subsequent filtering step. In this *Filter* step, actions were described that were taken to help reduce file size, while still retaining information. Furthermore, the importance of checking for consistency during normalization was discussed. After describing the *Filter* phase of the *Destiny* dataset, the *Enhance* phase consisting of the reconstruction of weapon usage statistics and scraping players' DTR scores, was explained in more detail. Lastly, the resulting data model of the *Destiny* dataset was presented and briefly outlined. After describing the process for the *Destiny* dataset, a similar approach was taken for the `the100.io` dataset. Since this dataset was already stored in the HDF format, no transformation was needed. The second phase, *Filter*, on the other hand, was still necessary to alleviate issues related to spurious correlations. In this respect, entries with implausible or missing values were removed from the dataset. This dataset, again, was augmented in an *Enhance* phase. On one hand, this was achieved by computing group-level information constituting aggregate player information, such as, average age, level and *Light levels*. On the other hand, a social network was constructed from the game information and network measures, such as, centralities and clustering coefficients were computed for each user of `the100.io`.

6. Analysis and Results

Parts of the analyses and results presented in this chapter have already been published in Schiller et al. (2018) and Wallner et al. (2019). This chapter will begin with general descriptive statistics of both the *Destiny* dataset and the `the100.io` dataset which was also analyzed in Schiller et al. (2018), Wallner et al. (2019). The analyses of the `the100.io` dataset presented are separated into player-level and group-level analyses. Finally, the analyses, results of applying archetypal analysis (AA) to users and groups on `the100.io` will be shown and discussed.

6.1. General Analysis

In this section, general statistics and metrics calculated from the datasets will be presented briefly. For this reason, overviews of both datasets will be presented as well as high-level analyses will be discussed here.

6.1.1. Overview

The *Destiny* dataset contains information about 3,450,839 distinct users who played 930,721 matches in *Destiny*[1]'s player-versus-player (PvP) mode, *The Crucible*, between September 8th, 2014, and January 1st 2016. The distribution of matches across the 13 different game modes captures is shown in Figure 6.1.

[1] Bungie, 2014. https://www.destinythegame.com/d1.

Distribution across Game Modes

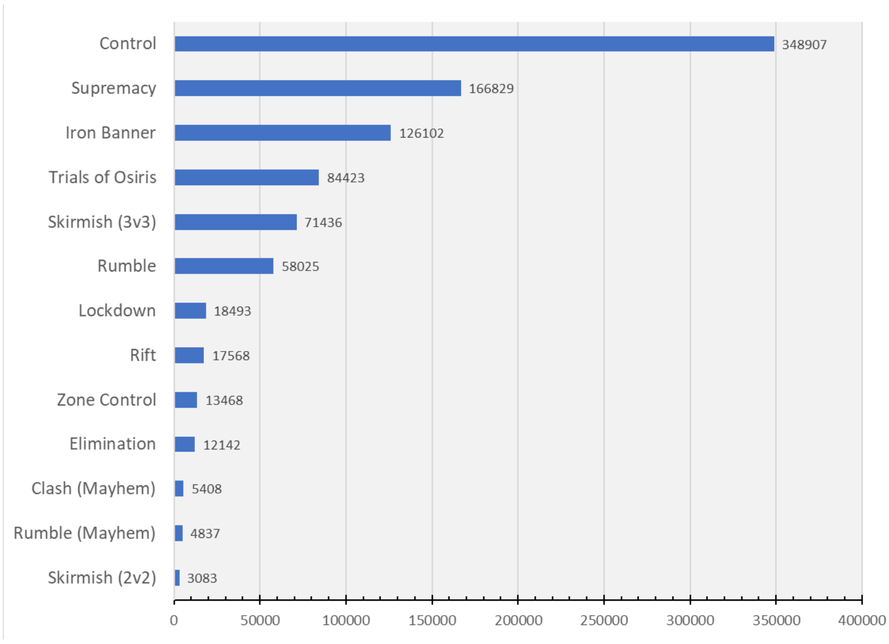


Figure 6.1.: Distribution of Games Across Different Game Modes

From this distribution, it can already be seen that *Control* is clearly the most popular game mode with over a third. Here, it should be noted that *Lockdown* was a variation of *Control* which only was available for a limited time only from the *Crucible Labs* and removed afterwards. For this reason, it was not included in subsection 3.4.2. Of the users participating in the games, 2,508,151 were also registered on Bungie.NET, which is Bungie's web presence. In this sample, 10,387,020 individual player performances, that is statistics about participating in a single game session, were collected. Furthermore, 19,759,060 single-game per-weapon performance statistics were gathered. During their gameplay, the players in this dataset achieved between 0 and 1,211 kills per session, while accumulating between 0 and 665 deaths per match. These statistics amount to players' performance measures *K/D* and *K/DA* both ranging from 0 to 49. Furthermore, players earned between 0 and 642 medals per session with 100 different types of medals captured. The aggregate duration of all captured sessions combined amounts to 5,862,897,914 seconds of playtime which is approximately 185,8 years of combined playtime.

Class Distribution

When considering all 4,482,863 *Guardians* contained in the *Destiny* dataset, 1,693,731 are of the type *Hunter* class, 1,447,568 are *Warlocks* and the remaining 1,341,567 Guardians are *Titans*. This way, the distribution of in-game characters, Guardians, amounts to the diagram shown in Figure 6.2. From this, we see that 38% of Guardians are *Hunters*, while 32% are *Warlocks*. Lastly, *Titans* represent a minority with only 30% of player characters, *Guardians*, being of this type of class.

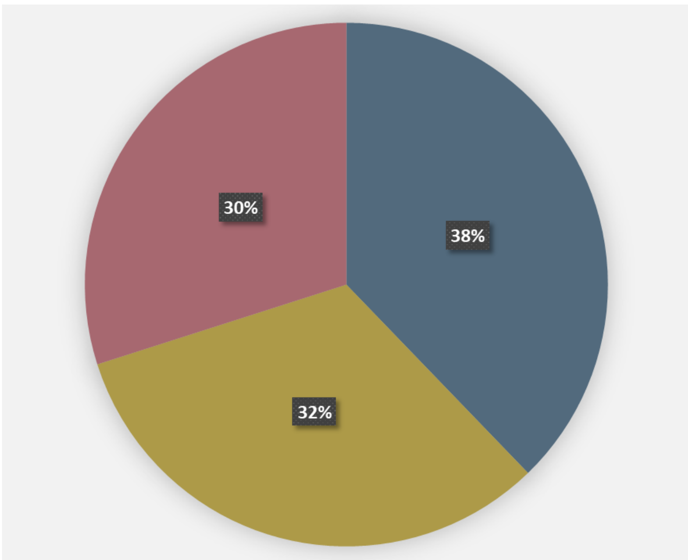


Figure 6.2.: Class Distribution of *Guardians* (■ = Hunter, ■ = Warlock, ■ = Titan)

Outliers & Peculiarities

Some surprising observations in the *Destiny* dataset include the highest numbers of kills and deaths *within one game* with the values of 1,211, and

665, respectively. Furthermore, a K/D ratio of 49 seems extraordinarily high as it indicates that a user was able to garner 49 kills before being killed once, over the course of a game, *on average*. Additionally, the number of 642 medals being awarded after a match concluded seems unusually high. The longest match recorded in this dataset spanned 581,640 seconds which corresponds to a duration of 6 days 17:34:00 [HH:MM:SS]. Lastly, the longest single life (without being killed) of a player indicates a value of 581,348 seconds which amounts to 6 days 17:29:08. After giving a short overview of the *Destiny* dataset and presenting some high-level analyses, in the following paragraphs, characteristics of the `the100.io` dataset will be discussed and some general analyses will be shown. `the100.io` is a looking for group (LFG) website facilitating matchmaking within online video games. While it offers matchmaking for a range of other games, in the scope of this thesis, *Destiny* was the focus of analysis. Considering the entirety of `the100.io`, the dataset encompasses 218,214 users in 2,468 groups. Users created 273,688 friendships with some of them being confirmed while others were still pending. Additionally, the dataset contains 60,812 relationships between users and groups, that is, group memberships, and includes and includes both, users which are not members of any groups, and users which are members of multiple groups. The maximum number of groups a user was a member of was eight. Considering a player's role on `the100.io`, the dataset indicates 2,522 group moderators which are users who can create and manage game sessions and govern group members. Furthermore, 4,374 users are labelled as *sherpas* which offer guidance and help new users both, in-game and on the platform, `the100.io`. Focusing on the main purpose of `the100.io`, it can be observed that 24,030 game sessions were listed on the players' profiles while 2,161 scheduled game sessions were recorded on group profiles.

Games Played on `the100.io`

From Figure 6.3, it can be seen that *Destiny* is the most popular game on the `the100.io` with 1,775 out of 2,468 groups (72%) listing it as a game they are interested in playing. The two next most popular games are *The*

Division[2] and *Overwatch*[3] with approximately 550 and 400 groups playing them, respectively. Furthermore, it can be observed that `the100.io` offers presets for 16 games, while also allowing for scheduling arbitrary games (see "Other"). In this way, 72 groups listed other games on their profiles. As this

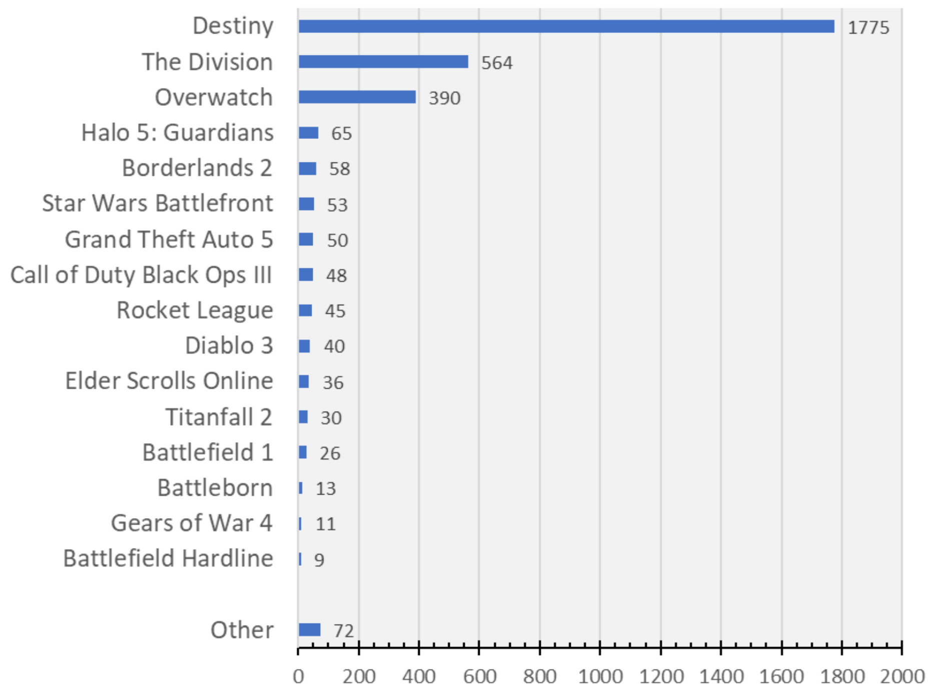


Figure 6.3.: Distribution of Video Games Supported by `the100.io`

thesis focuses on the first-person shooter (FPS) game *Destiny*, finally game-specific measures will be described here. These measures are the Guardian's character level and the *Light level*, with the former showing values in the range [1–40] and the latter ranging from 10 to 400.

[2] Massive Entertainment, 2016. https://tomclancy-thedivision.ubisoft.com.
[3] Blizzard Entertainment, 2016. https://playoverwatch.com.

Platform Preferences

As one of the first player-level analyses on the `the100.io` data, the user-reported platform preference was investigated. Of the 218,214 users of `the100.io`, 28 entries did not contain valid information for the "preferred platform". After removing these, the diagram shown in Figure 6.4 was constructed. We can see a majority of users preferring the PlayStation ecosystem, consisting of the two consoles, PlayStation 4 (PS4) (103,920; 48%) and PlayStation 3 (PS3) (8,441; 4%). While the Xbox ecosystem is not as popular in general, the Xbox One still was the preferred platform for 81,770 players, or 37% of them - resulting in Xbox One being the second most popular platform. Also placed under the umbrella term of "Xbox", the Xbox 360 was the preferred platform for 11,086 or 5% of players. Furthermore, we can see the overall trend towards current-generation consoles, namely, the PS4 and the Xbox One, which together were the preferred platforms for approximately 85% of users of `the100.io`. Peculiarly, 6% of users of `the100.io` entered PC as their preferred platform. It is worth noting, that *Destiny* was never released on PC.

When analyzing the reported timezones, it is worth noting that players selected a wide range of different time zones – 146 distinct values were chosen to be exact. However, when analyzing all of the approximately 218,000 players, 70% reported to be in a US and Canada-related timezone – Eastern Time (34%), Central Time (18%), Pacific Time (14%) and Mountain Time (4%), But also UK-related time zones have relatively high shares of players with London (6%) and Edinburgh (5%).

Preferred Play Times

Table 6.1 shows the distribution of players' choice for their preferred playtime. Here, it has to be noted that a user of `the100.io` can select any of the options or may also choose none of these. It is not possible to select multiple options. As shown, the vast majority of users of `the100.io` report to being available for play on both weekdays and weekends. Only 3 players reported to prefer to only play on weekends. Furthermore, almost half of all players (45%) prefer to play weekdays late-night, as well as on weekends.

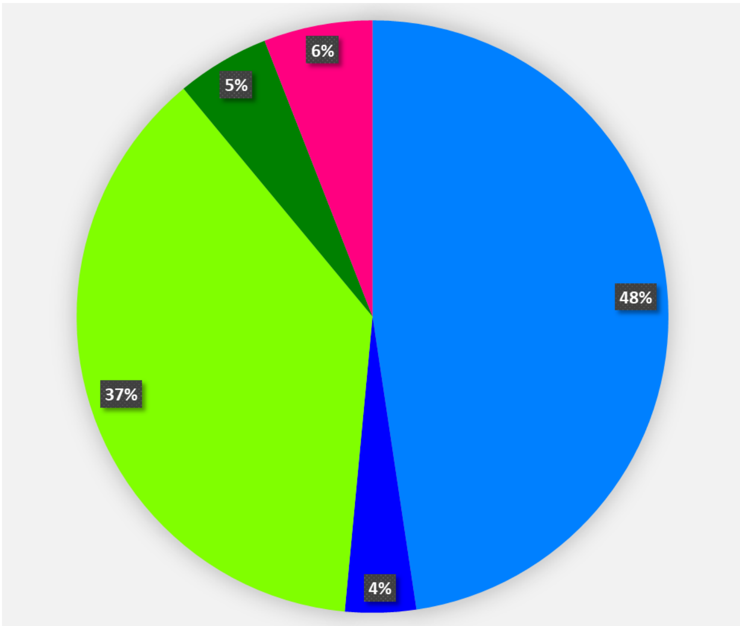


Figure 6.4.: Preferred Platform. User-reported. (■ = PS4, ■ = PS3, ■ = Xbox One, ■ = Xbox 360, ■ = PC)

Also, there is almost an even split between weekday evenings and weekday mornings with 24% and 20%, respectively. Lastly, almost every tenth (11%) user of `the100.io` reported to being available weekday afternoons and weekends. Note, that these numbers do not match the numbers presented by Wallner et al. (2019) due to the fact, that here the entirety of data was taken into account, while in their work the authors chose to limit their descriptive statistics to nodes that are part of the largest connected component (LCC).

	Number of Players
WD Late-night & WE	97,616
WD Evenings & WE	53,579
WD Mornings & WE	42,773
WD Afternoons & WE	24,140
WE Only	3

WD = Weekday, WE = Weekends

Table 6.1.: Distribution of player-reported preferred playtime on `the100.io`. Missing entries were removed. Multiple answers were not possible.

Related to users' playtime preference, Figure 6.5 represents the distribution of games scheduled on `the100.io`. Shown are 1,493,599 games scheduled between January 1, 2011 and December 31, 2017. At this point, it is worth noting that games on `the100.io` can and probably usually are scheduled in advance. Hence, the dataset contains games that were scheduled in the future relative to the time of collection. Furthermore, a large fraction of games features dates in the years 2015 (883,695) and 2016 (587,247).

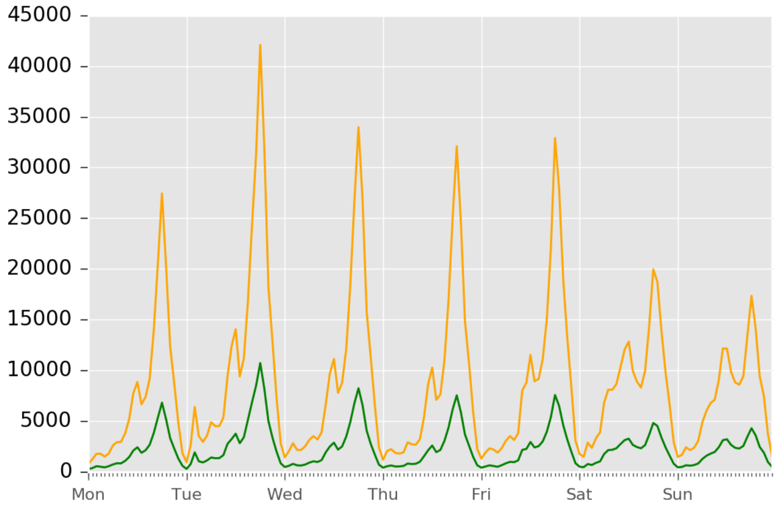
Distribution of Activity Over the Course of a Week

Figure 6.5a shows the aggregate number of scheduled games and players registered to these games projected onto the period of one week. The distribution of each day approximately follows this trend: activity rises up until midday, then falls a bit in the early afternoons. Afterwards, it rises again and peaks on late afternoons – with the highest peak at 6 PM – and falls again until midnight. This holds true for all days of the week, while the highest activity of any day is observed on Tuesdays. From plotting both games and users signed up for games, we can see that there are no noticeable trends for scheduling "larger" games, that is, games requiring more players on a team, on specific days throughout the week. In addition, Figure 6.5b shows the average number of games scheduled at any point in time during the week – categorized by group type, that is, for casual and serious groups. Here we see that both serious and casual groups follow the same overall trends as seen in Figure 6.5a. Furthermore, weekends see the lowest number of games scheduled, which may be due to players not requiring `the100.io`'s help finding others to play with during this time, or due to players not playing as much on weekends. These conjectures about the reasons for lowered activity on weekends would have to be tested in follow-up research, though. In the next sections, analyses performed specifically on the `the100.io` dataset will be presented. First, player-level analysis and community structure will be described. Subsequently, the results of group-level analysis are shown and briefly discussed.

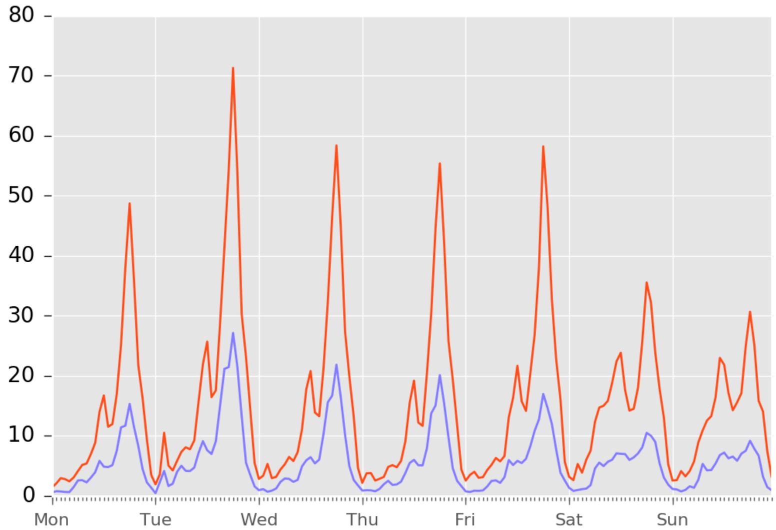
6.1.2. Players

In this section, player-based analyses will be discussed which, in part, have already been published by Wallner et al. (2019).

As a first step, descriptive statistics of player-level information available in the `the100.io` dataset were created. The overview of statistics of measures is shown in Table 6.2. From this overview, it can be seen that none of the players is a member of more than eight groups and no player was able to accumulate a *sherpa* score of over five. A broad distribution can be observed for the measures of *karma*, ranging from 0 to 572 and, particularly, for the number of friendships, ranging from 0 to 631, and activity score, being on a spectrum of 0 to 7,419. For non-numeric features (not listed), that is, gender, time zone, preferred platform, preferred playtime, play style, and profanity other values were possible. For these value ranges no sensible values for mean and standard deviation can be reported. The metrics for which this holds true will be described here briefly.
For each of the players, the metric gender had one of the following values:



(a) Number of scheduled games (■) and number of players (■) signed up for these games.



(b) Average number of scheduled games across casual (■) and serious groups (■).

Figure 6.5.: Distribution of Weekly Activity (Schiller et al., 2018).

male, female, other, or private.
Reported values for the time zone measure include large counts of Eastern Time (US & Canada), Central Time (US & Canada), but also European time zones (London, Berlin, Paris), as well as significant numbers of users reporting time zones related to Australia (Sydney, Brisbane, Melbourne). Furthermore, a small number of users indicated their time zones as Kalinigrad, Kathmandu and Fiji. Overall, 146 different time zones were reported by users of `the100.io`.
The user-reported preferred platform takes on one of the values Xbox 360, Xbox One, PlayStation 3, PlayStation 4, and PC.
A user's preferred playtime can take on at most one of these values: Weekday Mornings and Weekends, Weekday Afternoons and Weekends, Weekday Evenings and Weekends, Weekday Late-night and Weekends and Weekends only. For further information, please see Table 6.1.
Users of `the100.io` can choose between two play styles according to their tastes; either casual ("Having fun is most important"[4]) or serious ("Getting it done is most important"[5]).
Finally, for the metric profanity the following values are present in the dataset: no profanity, some profanity, and profanity OK.

Descriptive Statistics of Player-Level Information

For each of the user-level variables, the number of missing entries is reported in Table 6.3. From this overview it can be seen that the measure "play style" was missing the highest number of entries (22,323), with non-trivial amounts of missing entries for the measures "profanity" (15,248) and "Light level" (12,533). Furthermore, the "age" measure was missing from 5,078 of players' entries. While none of the entries for the "gender" metric is missing, it has to be noted that the majority of users have a value of "private". This may be due to users not wanting to share this information but also due to `the100.io` not requiring "gender" data and defaulting to "private".

For each of the numeric variables used for further analysis, that is, the *Guardian*'s level and *Light level*, as well as platform-related measures con-

[4]Description taken directly from `the100.io`: https://www.the100.io.
[5]See [4].

	Value Range	Mean	Std. Dev
Age	[13 – 99]	30.82	9.54
Level	[1 – 40]	37.24	5.37
Light Level	[10 – 400]	341.26	56.53
Sherpa Score	[0 – 5]	0.066	0.357
Friends	[0 – 631]	5.83	11.68
Number of Groups	[0 – 8]	0.49	0.77
Activity Score	[0 – 7,419]	59.04	108.66
Active Games	[0 – 23]	0.47	1.56
Karma	[0 – 572]	15.50	31.78

Table 6.2.: Descriptive statistics of user-level variables. Adapted from (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

Metric	Missing	Metric	Missing
Age	5,078	**Gender**	0
Time Zone	1	**Preferred Platform**	0
Level	35	**Light Level**	12,533
Sherpa Score	0	**Friends**	0
Number of Groups	0	**Activity Score**	129
Active Games	0	**Karma**	0
Preferred Playtime	1	**Play Style**	22,323
Profanity	15,248		

Table 6.3.: Number of Missing Values from User-Level Data

nected to `the100.io`, that is, *karma*, *sherpa* score, friends, number of groups, activity score, and active games, values were investigated. The distribution of values was plotted in the form of histograms, as shown in Figure 6.6. Here, Figures 6.6a,b are in-game statistics related to the game *Destiny* itself, while Figures 6.6c-f show measures relating to the LFG site `the100.io`. Here we can already see that a number of cluster points exist for the measures of *level* and *Light level*. In this regard, the majority of players on `the100.io` already reached the maximum level, but there are also peaks around the levels 1 and 10. Furthermore, a substantial number of players is in the level range of [30–35]. Similarly, for the *Light level*, a large number of players has a very high *Light level*, in the [300–400] range, with 400 being the maximum achievable in *Destiny*, but there is also a peak with a non-trivial fraction of players exhibiting a *Light level* of around 100.

Community Structure of the Friendship Network on `the100.io`

In another analysis, Wallner et al. (2019) examined the social network structure found within the `the100.io` dataset. The network that is at the foundation of these analyses is the friendship network constructed as described in section 5.3.3. In order to find communities, an algorithm for detecting communities in large networks, namely the Clauset-Newman-Moore community detection algorithm (Clauset, Newman, & Moore, 2004, 6), was applied to the network. This algorithm detects clusters as densely connected sets of nodes with relatively few connections to nodes on the outside. It has to be noted, that the communities detected in this way, generally, do not coincide with groups formed on the `the100.io` website. The resulting graph, depicted in Figure 6.7, shows a node-link visualization of the communities in the LCC of the friendship network of users on `the100.io`. Communities of size greater than 50 are enclosed by borders and explicitly annotated, with 36 communities matching this criterion. The size of a node encodes the number of friendship the corresponding player has on `the100.io`. It can be observed that there is a clear divide between the two ecosystems, PlayStation (shown in blue), and Xbox (shown in green). This split can be explained by *Destiny* not supporting cross-platform play. For each of the current-generation platforms, mainly one large community was formed, with #0 for PS4 (12,984 members), and #1 for Xbox One (10,993 members).

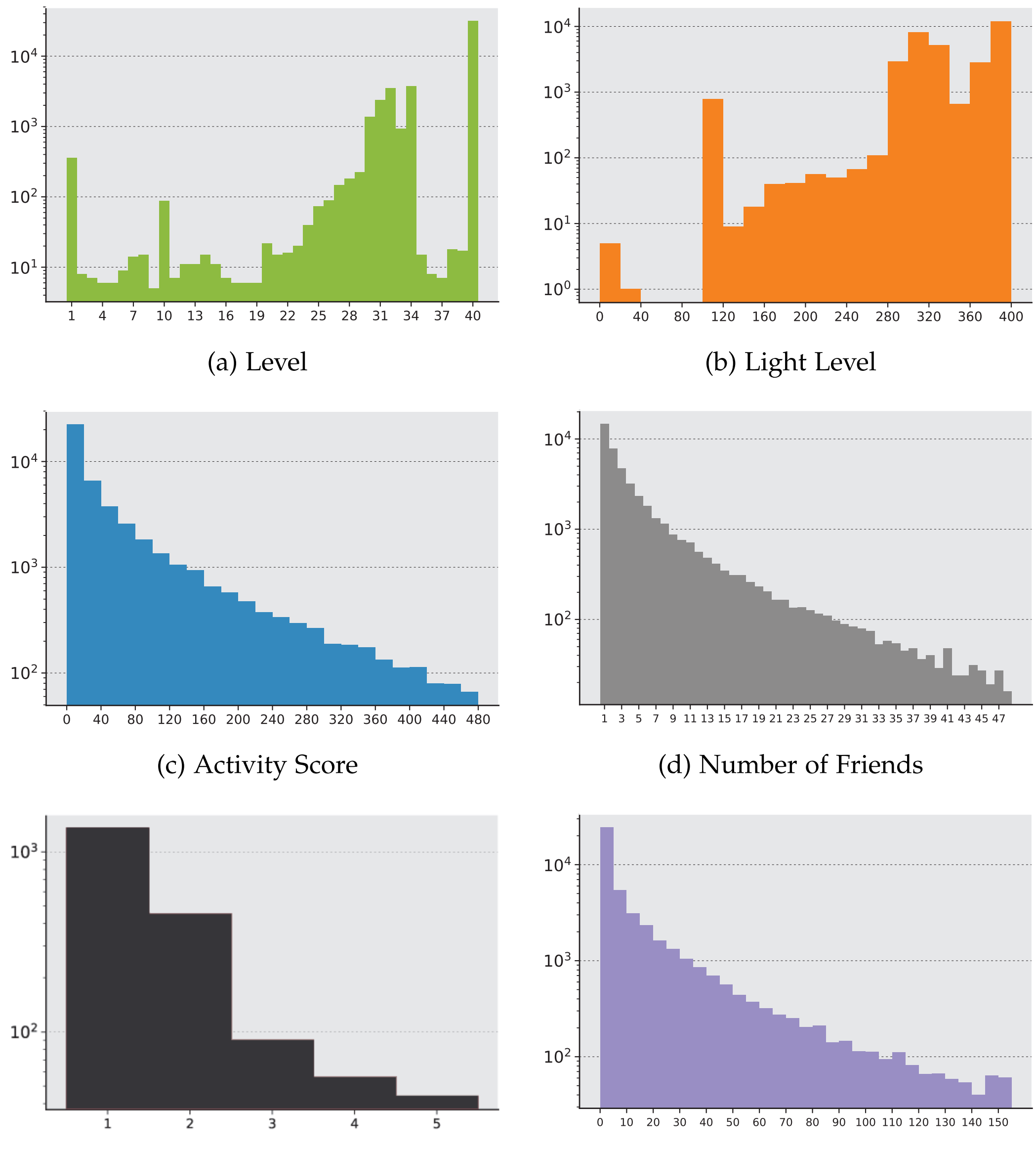


Figure 6.6.: Distribution of *Destiny* (a, b) and `the100.io` (c, d, e, f) related measures (y-axes indicate the number of players and are log-scaled). Adapted from Wallner, Schinnerl, Schiller, Pirker, and Drachen (2019).

Henceforth, these two communities, which together represent over 50% of players in the LCC, will be referred to as the principal communities. Similarly, for the previous-generation platforms, that is, Xbox 360 and PS3, two communities with high fractions of players preferring the respective platform have formed (#2, and #3, respectively). This accumulation of players preferring previous-generation platforms, Xbox 360 and PS3, within the confines of distinct communities can be explained by the fact that the users preferring current-generation platforms by far outnumber those preferring previous-generation platforms – with a factor of over 10 : 1. Therefore, it seems likely that players still playing on the older platforms tend to connect to other players on these platforms, thus, forming communities around previous-generation platforms. Furthermore, it can be observed that smaller communities mostly connect to the principal communities, while links connecting smaller communities seem rare. Figure 6.7 also shows a number of players that are not part of any community. These players only have a small number of links, connecting them predominantly to the two principal communities. These players may have signed up for `the100.io` and tried to use it and subsequently did not want to use the matchmaking service further. Lastly, it can be noted that two communities (#22 and #26) have formed which are entirely comprised of PC players. These two communities are relatively isolated and only have few and weak links to other communities. In regards to *Destiny*, this may be due to the fact that the game was never released on the PC platform. For players inside these communities, this fact possibly makes it harder to find others to play with, as users of `the100.io` cannot tell on which platform players in these communities play *Destiny*.

Table 6.4 shows a quantitative analysis of all communities in Figure 6.7 containing more than 50 members (ordered by community size). While users in larger communities tend to exhibit a higher average degree, the overall density within a community is higher for small communities. However, larger communities tend to foster higher absolute numbers of friendships – as expressed by the average degree, which indicates a higher number of incident edges, that is, friendships. Concerning the average level within communities, most communities are similar with high average levels, greater than 30, with the notable exceptions of community #12, which has an average level of about 25 and the PC-only community #22 with approximately 20. In terms of preferred playtime, all communities cover all preferences, that

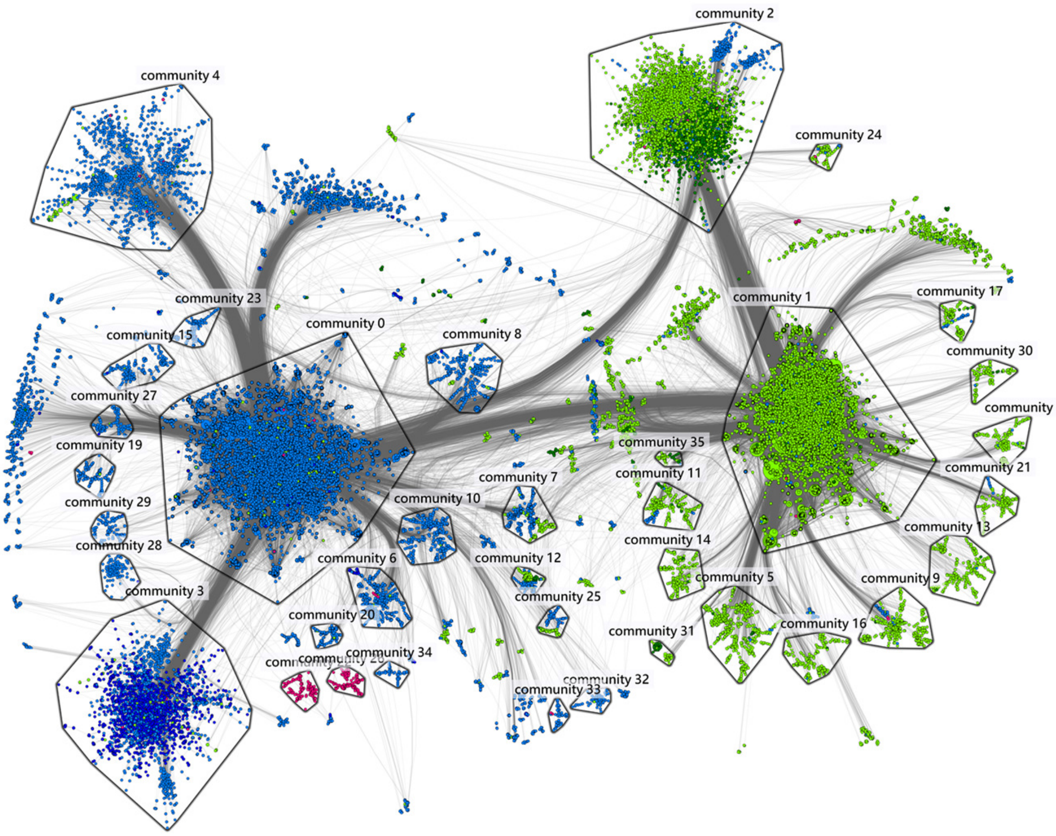


Figure 6.7.: Visualization of the community structure of the largest connected component ($|N| = 45,221$, $|E| = 135,747$) of the friendship network extracted from the `the100.io` website. Communities with more than 50 members are enclosed by borders. The coloring of the nodes show the platform on which the individual users prefer to play. (■ = Xbox 360, ■ = Xbox One, ■ = PS3, ■ = PS4, ■ = PC). The size of a node is proportional to the number of friends. Edge bundling with alpha blending was used to accentuate the flows between communities (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

	Size	Density	Degree	Platform	Act. Score	Level	Pref. Time
#0	12984	0.000506	6.57±0.11		78.53±118.14	38.05±5.02	
#1	10993	0.000744	8.18±0.15		77.62±120.61	38.56±4.11	
#2	4715	0.000928	4.37±0.09		61.01±153.99	36.30±5.25	
#3	2533	0.001563	3.96±0.11		40.85±75.37	36.49±4.99	
#4	1399	0.002424	3.39±0.12		47.28±88.51	35.58±5.17	
#5	516	0.004990	2.57±0.12		33.12±49.65	35.88±4.18	
#6	497	0.005452	2.70±0.12		34.67±60.81	35.26±5.74	
#7	446	0.008787	3.91±0.23		28.61±55.46	38.13±4.64	
#8	422	0.006878	2.90±0.18		25.94±38.42	35.99±4.52	
#9	392	0.007190	2.81±0.22		22.48±41.78	30.57±13.04	
#10	367	0.009038	3.31±0.19		42.52±60.08	36.70±5.14	
#11	247	0.010434	2.57±0.16		41.47±58.79	36.94±4.59	
#12	242	0.012585	3.03±0.30		1.34±8.21	24.93±11.41	
#13	242	0.010356	2.50±0.19		26.59±54.52	34.93±5.10	
#14	223	0.011393	2.53±0.18		34.63±60.14	35.51±5.12	
#15	207	0.017541	3.61±0.37		82.70±167.83	35.43±5.04	
#16	188	0.014621	2.73±0.19		26.40±37.58	38.22±3.37	
#17	163	0.023101	3.74±0.53		22.13±33.80	38.13±4.50	
#18	159	0.020221	3.19±0.33		33.73±53.59	36.58±5.37	
#19	155	0.016590	2.55±0.16		37.22±62.90	37.03±3.97	
#20	151	0.018190	2.73±0.30		23.26±31.61	36.03±6.18	
#21	134	0.025025	3.33±0.31		58.96±78.31	36.76±4.92	
#22	134	0.018292	2.43±0.19		10.58±21.91	20.33±13.31	
#23	131	0.027011	3.51±0.47		22.60±33.26	39.36±2.26	
#24	124	0.022292	2.74±0.27		35.87±58.00	35.39±4.37	
#25	123	0.027856	3.40±0.29		74.42±108.57	37.13±4.11	
#26	120	0.018908	2.25±0.26		13.96±28.15	38.65±5.87	
#27	119	0.027631	3.26±0.38		30.47±42.65	37.41±4.53	
#28	109	0.041454	4.48±0.63		75.42±88.41	38.40±4.59	
#29	108	0.023018	2.46±0.17		28.25±36.63	35.43±4.09	
#30	106	0.024618	2.58±0.22		32.31±42.65	37.16±5.12	
#31	67	0.064677	4.27±0.68		12.10±25.58	34.76±5.25	
#32	60	0.050847	3.00±0.44		56.57±72.03	39.98±5.24	
#33	56	0.043506	2.39±0.35		15.61±19.76	34.77±11.36	
#34	52	0.041478	2.12±0.25		22.17±35.62	32.42±8.01	
#35	52	0.042986	2.19±0.49		20.63±28.42	37.29±5.01	

Table 6.4.: Descriptive characteristics of communities with more than 50 members. Compound player-level data, that is, degree, activity score, and level are expressed as mean ± standard deviation. (= Xbox 360, = Xbox One, = PS3, = PS4, = PC, = mornings & weekends, = afternoons & weekends, = evenings & weekends, = late-night & weekends). Adapted from Wallner, Schinnerl, Schiller, Pirker, and Drachen (2019).

is, every possible option for *preferred playtime* is present in each community. However, most of the communities show a clear focus on the option late-night & weekends, with the notable exceptions of communities #10 and #18, which predominantly prefer to play on weekday evenings, and community #28, which is the sole community, for the most part, preferring to play on weekday mornings. With respect to the activity score, large variations between the different communities can be observed. However, the principal communities show high levels of activity among their members, possibly due to the opportunities to partake in matches induced by the higher average numbers of friends. On the other hand, some smaller communities (#15, #25 and #28) also show very high levels of activity. On the contrary, there are also communities with noticeably lower average activity scores, such as, #31, and, especially, #12 (1.34 on average!). Furthermore, the PC-focused communities (#22 and #26) also exhibit low levels of activity. This, again, probably relates to the issue described earlier.

Table 6.5.: Spearman rank correlation between various user-related variables and centrality measures. Moderate or larger correlations ($\rho > .3$) are written in boldface. Adapted from (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

	Destiny		`the100.io`					
	Level	Light Level	Activity Score	Active Games	Group Count	Sherpa Score	Karma	Age
Level	1							
Light level	.249*	1						
Activity Score	**.429***	**.451***	1					
Active Games	.235*	**.484***	**.405***	1				
Group Count	**.329***	**.421***	**.340***	**.414***	1			
Sherpa Score	.113*	.246*	.272*	.281*	.158*	1		
Karma	**.595***	**.615***	**.759***	**.451***	**.428***	**.312***	1	
Age	.156*	.193*	.170*	.182*	.223*	.072*	.247*	1
Degree Centrality	.285*	**.357***	**.666***	**.327***	**.312***	.214*	**.591***	.259*
Betweenness Centrality	.240*	**.315***	**.630***	.284*	.253*	.217*	**.515***	.198*
Closeness Centrality	**.383***	**.402***	**.617***	**.331***	.277*	.236*	**.679***	.248*
Eigenvector Centrality	**.327***	**.304***	**.480***	.284*	.240*	.181*	**.537***	.208*

Cases with missing values excluded, $^*p < .00083$, Bonferroni corrected

0 1

Correlations

Table 6.5 depicts the summary of Spearman's rank correlation coefficients between various user-level measures. In regards to demographic data, here, the user's gender was excluded, as almost 50% of users did not publicly report their gender, and the distribution resulting from the reports of the remaining users was heavily skewed towards *male*. Note, that correlations between different centrality measures were also excluded as they do not provide much additional insight, and are often highly correlated, anyway (Valente, Coronges, Lakon, & Costenbader, 2008). This leaves the demographic information of age, in-game metrics related to *Destiny* (level and *Light level*), `the100.io`-related measures (activity score, active games, group count, *sherpa* score, and *karma*) and the four centrality measures (degree centrality, betweenness centrality, closeness centrality and eigenvector centrality). All correlations were shown to be significant with $p < .00083$. The results discussed here are going be limited to at least moderate (cf. Cohen, 1988) correlation coefficients. This means that only results with $\rho > .3$ are reported.

Immediately, it can be seen that both *Destiny*-centered metrics, level and *Light level*, show moderate positive correlations with different centrality measures. This means, that players who are more socially embedded within the `the100.io` network have also progressed further within the game, *Destiny*. However, it cannot be determined whether players reach higher levels of achievement due to being highly connected, thus, having more opportunities to play with others and progressing through the game, or if players are better connected because they are high-level players. In addition, the activity score as provided and calculated by `the100.io` also shows moderate to high levels of correlation with the centrality measures – indicating that players more central to the network are also "more active" as determined by `the100.io`. Furthermore, users with higher activity scores exhibit more active games, meaning that their overall activity (since sign-up), coincides with high activity (number of active games) during the period prior to sampling.

It is worth noting, that the group count metric only weakly to moderately correlates with the various centrality measures. The strongest of those correlations, that is to say, the moderate correlation between group count and degree centrality, indicates that while being a member of more groups predicts more friendships, it does not result in considerably more friendships or vice versa. This may well be due to players mainly seeking friendships within only one group, their primary group with which they play.

Furthermore, *karma* is highly correlated with all of the centrality measures. The strongest of these correlations can be observed with closeness centrality, that is, more central players exhibit higher *karma* values. Since *karma* may only be awarded once from each player to each other player, a higher degree centrality, that is, a larger number of friends increases the likelihood of receiving *karma* resulting in a higher overall score for *karma*. Additionally, *karma* is highly correlated with both activity score and the number of active games. Again, this is to be expected as participating in more activities exposes a player to more chances of receiving *karma*. Nevertheless, it cannot be said definitively whether *karma* can act as an incentive for making more friends and/or for playing in more game sessions. While it should be noted here that higher-level players, that is, those with higher *Guardian* levels and *Light levels*, also exhibit higher values of *karma* it is not clear at this point if this is impacted by high-level players being more active, that is, being a member of more groups and showing more active games and by these players being more central to the social network. Lastly, the *sherpa* score is only moderately correlated with karma, that is, a higher *sherpa* score only predicts a slightly higher *karma* value and vice versa. No matter which of the centrality measures is used, *sherpas* are only marginally more central to the network than non-*sherpas*. At first glance, this is an unexpected outcome, since *sherpas*, that is, players willing to help other players, would presumably make for alluring friendships as they can help and offer guidance in unfamiliar situations. The fact that there is only a weak correlation could possibly be effected by the audience found on `the100.io`. From the dataset it can be concluded that `the100.io` mostly attracts comparatively experienced players – as indicated by high value for *Light* and Guardian levels. Another confounding factor in this regard could be that `the100.io` does not indicate an easy way for finding *sherpas* for a user's specific platform allowing them to specifically target *sherpas* in their search for friendships.

After investigating user-level measures and their correlation in this section, in the next section group-level measures and analyses performed on the

the100.io dataset will be presented.

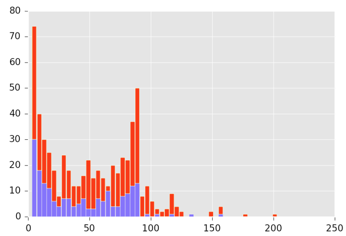
6.1.3. Groups

In this section, group-level analyses will be presented, some of which have already been published by Schiller et al. (2018). The the100.io dataset contains information for all groups, and games scheduled on the100.io since its inception up to the time of collection, December 16, 2016. A grand total of 637,823 unique game sessions was scheduled by 218,214 players in 2,468 groups. As the following analyses will focus on the FPS game *Destiny*, and because the100.io offers its matchmaking services for other games as well, groups that did not report to playing *Destiny*, as well as game sessions scheduled for video games other than *Destiny* were removed and not taken into account for the analyses discussed here. Furthermore, groups on the100.io that are composed of two or fewer players were removed, as well as groups for which activity information was missing. Furthermore, game sessions for which associated group information was missing were excluded from further consideration. After applying these cleaning steps, 586 groups remained, 390 of which were labelled as serious groups, while the remaining 196 of them were marked casual. Considering associated platforms, the following results were obtained (highest number of groups to lowest): 252 groups are designated PS4 groups, while 216 are characterized as groups playing on the Xbox One platform. The platforms PS3 and Xbox 360 both have 42 groups each associated with them. Finally, 34 groups are dedicated to playing on the PC.

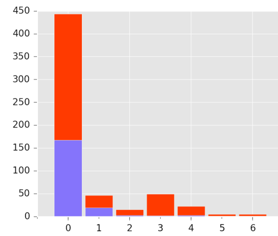
Overview and Descriptive Statistics

These 586 groups consist of 26,317 players which indicated to have played or planned to play a combined number of 1,493,599 games from the inception of the100.io until the time of data collection. While friendships on the100.io are typically formed by one party sending a request and another party accepting this request, it is worth noting that some information is already exchanged between players upon sending the request. For this reason, there may be too little incentive to actively accept friendship requests. Hence, for the following analyses, two players are considered to be friends if at least one party sent the other a friendship request, as at least one of the users explicitly expressed interest in connecting to the other.

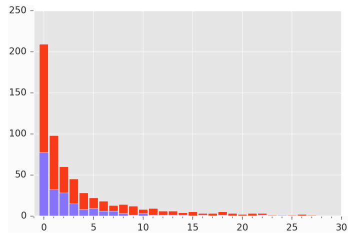
Figure 6.8 shows histograms for the distributions of basic group-level data. The histograms' information is categorized according to the type of group, that is, serious (■) and casual (■).
One of the measures, group size, is shown in Figure 6.8a. From the chart, it can be observed that most groups' sizes vary in the range from 3 to 100, with peaks on both ends of this range. After the peak at 100, that is, above 100, a drop in groups with corresponding sizes can be observed. This drop can be explained by the inner workings of the platform `the100.io`. The website tries to form groups of 100 players. However, users can invite other players and players may join groups on their own volition, resulting in group sizes greater than 100.
Another measure that is directly available from `the100.io` is the number of moderators a group has (Figure 6.8b). Here, it can be noted that most groups do not have any moderators at all. This may well be due to the requirement of reaching a specific level of activity before a player can become a group's moderator. The small peak indicating groups with exactly three moderators can be explained by `the100.io` allowing at most three non-paying users to become moderators. For further information regarding moderators on `the100.io`, please see subsection 4.1.2.
Along the same lines, most groups do not have a single *sherpa* among their users as can be seen from Figure 6.8c. However, compared to the distribution of moderators, the *sherpa* distribution exhibits a long tail (Shirky, 2003), that is increasingly smaller numbers of groups show higher and higher numbers of *sherpas* within their ranks. Overall, it can be said that groups contain more *sherpas* than moderators.
Generally, users of the `the100.io` platform have made significant progress within the game. This is reflected in the groups' average level (Figure 6.8d) – most groups exhibit an average level in the range of 35 to 40, with 40 being the maximum level achievable.
Furthermore the skewed distribution of player progress towards high achievement is noticeable when looking at the groups' average *Light levels* (Figure 6.8e). While the values of the distribution are more dispersed than the one for the average level, still, a large fraction of groups show



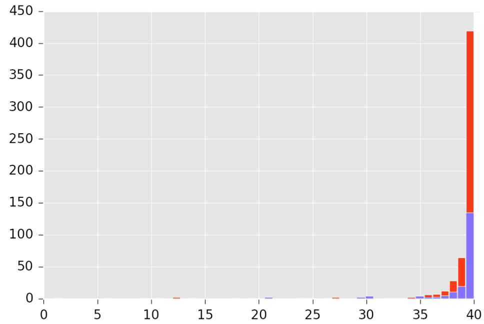
(a) Group Size



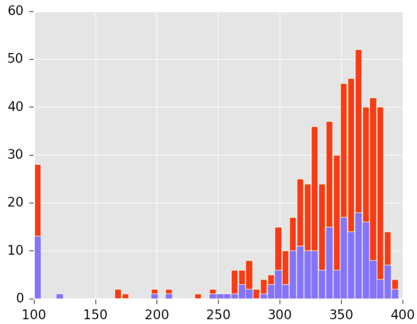
(b) Number of Moderators



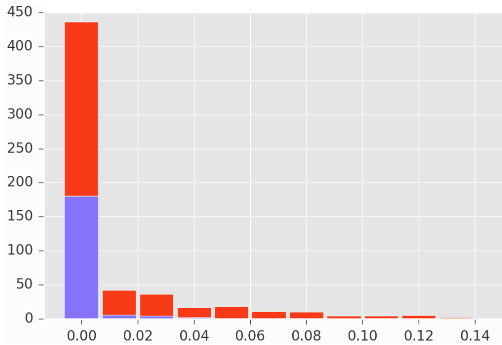
(c) Number of Sherpas



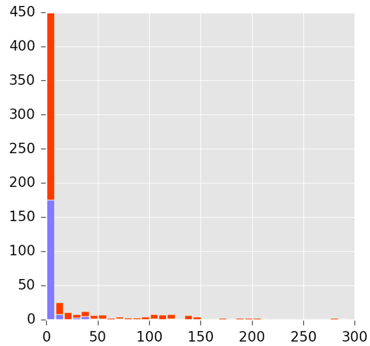
(d) Average Level



(e) Average Light Level



(f) Density



(g) Activity Score

Figure 6.8.: Histograms of distributions of group-related characteristics for serious (■) and casual groups (■). Adapted from Schiller et al. (2018).

average *Light levels* in the range of 300 to 400. It has to be noted that 400 is the maximum possible *Light level* a player can achieve.
In terms of density, most groups are either disconnected or very loosely connected by their contained friendships. This circumstance is shown in Figure 6.8f with very low scores for the groups' densities. Finally, as seen from Figure 6.8g, most groups are entirely inactive or display very low activity scores. While a small fraction of groups achieves activity scores of up to 200, groups scoring higher than that become exceedingly rare.

Correlations

Table 6.6 shows results from Spearman rank correlations applied to group-level data as well as to network measures – with a focus on *sherpas* and moderators. Spearman's rho was selected as correlation measure due to the variables under consideration not being normally distributed – as suggested by, for example, Field (2013). Furthermore, the variable describing groups' play style was excluded due to being a dichotomous (or, if encoded, binary) variable. Some of the correlations reported in Table 6.6 were calculated on slightly smaller samples, due to some of the groups missing data required for calculating correlation. For example, correlations involving the global clustering coefficient C, could only be calculated using a sample of 270 groups, as only those had at least a single triad, that is, a set of three users that are connected by links within the network. If a group does not contain at least one triad, C is undefined. To account for the multitude of comparisons, a Bonferroni-corrected confidence level of .00091 (cf. Field, 2013) was used to test for statistical significance. Subsequently, only strong correlations with $|\rho| > .5$ (cf. Cohen, 1988) will be discussed in detail.

Following from the fact that the group-level activity score, unlike the player-level activity score, is calculated based on activities within the two weeks prior to the current day, it is to be expected that a high degree of correlation between active games – which are also calculated in a similar fashion – and group activity score is found. For further information on metrics retrieved from `the100.io`, please see subsection 4.1.2. Since the activity score and the number of active games are so highly correlated, further results will not distinguish between the two when reporting significant correlations.

Group size is strongly positively correlated with activity. Similar to the number of friendships discussed for the user-level correlations earlier, this may be due to a higher likelihood of exposure to activities on `the100.io`. Generally, it can be assumed that more players within a group lead to more games being scheduled within a group which – in turn – leads to higher group activity.
Furthermore, both measures, *number of moderators* and *number of sherpas*, are strongly correlated with group size. The more users are in a group, the more *sherpas* and moderators are also members of the group, and vice versa.
Contrary to the density observed in communities constructed from user information (discussed earlier), for groups, density increases as the group's number of members increases; players that are part of larger groups connect to relatively more players via friendships. Furthermore, density also positively predicts activity.
The average level of players does not show any strong correlations, which can be explained by the low dispersion of the average-level distribution (cf. Figure 6.8d) with almost all groups exhibiting average levels of or close to the maximum level of 40.
The average *Light level*, on the other hand, showed a strong correlation with group activity. In this regard, higher average *Light levels* seem to coincide with more active groups.
Both measures of connectedness, the average degree centrality of moderators as well as of *sherpas*, are strongly positively correlated with activity. However, these measures are also closely correlated with both group sizes and the number of moderators and *sherpas*, respectively.

Multiple Regression Analysis

This initial analysis showed strong correlations for a wide range of variables. Group activity seemed to be strongly impacted by all of these variables: group size, number of moderators, number of *sherpas*, density, connectedness of *sherpas*, connectedness of moderators, as well as *Light level*. To help better understand the influence of each of these factors a multiple linear regression was conducted. During the process, a model for predicting group activity from the measures *number of moderators*, *group size*, *number of sherpas*,

Table 6.6.: Spearman rank correlations between different group-related characteristics. Correlations with $|\rho| > .5$ are written in boldface (Schiller et al., 2018).

	Activity Score	Active Games	Group Size	No. of Moderators	No. of Sherpas	Density	C	$\overline{\text{dc}}$ of Moderators	$\overline{\text{dc}}$ of Sherpas	$\overline{\text{level}}$	$\overline{\text{light level}}$
Activity Score	1										
Active Games	**.958***	1									
Group Size	**.687***	**.676***	1								
No. of Moderators	**.748***	**.746***	**.606***	1							
No. of Sherpas	**.695***	**.698***	**.751***	**.676***	1						
Density	**.610***	**.613***	**.591***	**.683***	**.682***	1					
C	.405*	.424*	.309*	.465*	.425*	**.589***	1				
$\overline{\text{dc}}$ of Moderators	**.721***	**.723***	**.570***	**.927***	**.648***	**.715***	**.522***	1			
$\overline{\text{dc}}$ of Sherpas	**.672***	**.679***	**.578***	**.763***	**.724***	**.798***	**.573***	**.789***	1		
$\overline{\text{level}}$	-.085	-.084	-.234*	-.065	-.150*	-.073	-.091	-.080	-.059	1	
$\overline{\text{light level}}$	**.587***	**.580***	.462*	**.556***	**.533***	.440*	.344*	**.531***	.491*	.174*	1

dc = degree centrality, C = clustering coefficient, averaged values denoted by overlines, $^*p < .00091$ (Bonferroni adjusted) -1 0 1 *connectedness of sherpas*, *connectedness of moderators*, as well as *density* was developed. Basic regression coefficients are shown in Table 6.7.

Table 6.7.: Multiple linear regression of group characteristics on group activity (Schiller et al., 2018).

Predictor	**Estimate**	**Std. Error**	β	**t-value**
Number of Moderators*	11.10	1.47	0.31	7.57
Group Size*	0.16	0.04	0.12	3.68
Number of Sherpas	0.57	0.38	0.06	1.50
$\overline{\text{dc}}$ of Sherpas	-164008.24	100514.08	-0.07	-1.63
$\overline{\text{dc}}$ of Moderators*	534756.40	51001.41	0.50	10.49
Density	33.51	34.37	0.03	0.98

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

Three of the six predictor variables show a significant ($p < .001$) zero-order correlation with group activity. These variables are *group size*, *number of moderators* and *average degree centrality of moderators*. The resulting three-predictor model can account for 72.67% of the variance in group activity, $F(4, 574) = 254.5$, $p < .001$, with an adjusted R-squared of 0.7267 (Schiller et al., 2018). After presenting results for general statistical approaches and analyses on both user-level and group-level, in the following section the results from applying AA to groups and users found on `the100.io` will be discussed.

6.2. Archetypes

As a next step, AA was applied to both users and groups. In this way, both similarities and differences between entities, that is, users and groups, were discovered. The details and results of applying AA to the `the100.io` dataset will be presented in this section. To briefly recall, AA is used to find k archetypes, or archetypal points, in a set of n points that exist within a m-dimensional feature space. Considering the data points $x$to be encoded as a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ with each column representing one data point, then AA aims to find the two matrices $\mathbf{Z} \in \mathbb{R}^{m \times k}$ and $\mathbf{A} \in \mathbb{R}^{k \times n}$ which minimize the term $\|X - ZA\|_F$. Then $\mathbf{Z}$ contains the archetypes and the column-stochastic matrix $\mathbf{A}$ describes the belongingness between archetypes and data points. Each column a_i in $\mathbf{A}$ corresponds to a data point x_i in $\mathbf{X}$ and each column in a exists in $(k-1)$-simplex. Due to the fact that usually $k \ll m$, that is, the number of archetypes is much smaller than the number of features in the feature space, AA can also be used for dimensionality reduction. In this thesis, however, it will be used mainly for finding extremal points to help better understand the user and group behavior as expressed on `the100.io`.

6.2.1. Players

For most data analysis tasks feature selection is one important aspect. In this regard, AA is no different from traditional data analysis tools. For player-level analysis the features *sherpa* score, *activity score*, *karma*, *friend count*, *group count*, and the *player level* were considered. This selection of features represents both features, *Destiny*-related on one hand, and features connected to `the100.io` on the other, while it also keeps the overall number of features at a sensible level. This approach – of not keeping all features – was chosen for two reasons. Firstly, it keeps the obtained results interpretable. Secondly, features with a high number of missing values, for example, *Light level*, could be removed while still maintaining a high number of data points. To emphasize this point, keeping *Light level* in the feature set would have meant that more than 10,000 players ($\approx$ 25% of all players) would have had to be removed from consideration for AA. In preparation for AA the user data was processed further. Any users missing values for one or more for the selected features (listed above) were excluded from this analysis. Additionally, for the features *activity score*, *friend count*, *karma*, outliers were excluded by only considering players with values below or at the 99$^{\text{th}}$ percentile. After these steps, close to 43,000 players remained for consideration in AA. The features were then scaled by applying min-max normalization as

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \tag{6.1}$$

When entries contain the value 0 for the feature under consideration, this simplifies to

$$x' = \frac{x}{\max(x)} \tag{6.2}$$

which was the case for all features in this player-level analysis. Following these steps, AA itself was applied using Python and, specifically, the module `py_pcha` (Aslak, 2016), a Python implementation of the algorithm proposed by Mørup and Hansen (2012). This algorithm is adequate for being applied to large datasets. In order to find the correct number of archetypes, AA was run using different values for the number of archetypes, k. The values tested ranged from two to ten.

So as to determine the correct number of archetypes a scree plot (Figure 6.9) was composed. The plot shows the number of archetypes, k, on the x-axis and the fraction of variance explained by the corresponding k-cluster solution on the y-axis. The elbow method (cf. Thorndike, 1953) suggests that either a four-cluster solution or one that uses five archetypes fits the data well. It can be seen that the four-cluster solution explains over 98% of the variance while the five-cluster solution can account for approximately 99% of the variance. After examining both solutions and evaluating their interpretability, the five-cluster solution was chosen.

An overview of descriptive statistics for each of the clusters can be seen in Figure 6.10. Here it should be noted that visual inspection suggests that the four-cluster solution results in archetypes `A4` and `A5` collapsing into a single cluster. For the purpose of giving an overview of players' archetypes, Table 6.8 shows the distribution of players when assigned to their predominant archetype. Furthermore, Figure 6.11 shows the players' belongingness to the archetypes, as well as their distribution across different archetypes. In the following paragraphs, the five different user clusters will be described briefly.

`A1`: This archetype shows a pronounced difference when compared to all other archetypes: it contains users with high *sherpa* scores. Furthermore, players matching this archetype exhibit a high character level. These two observations probably coincide as players which are able

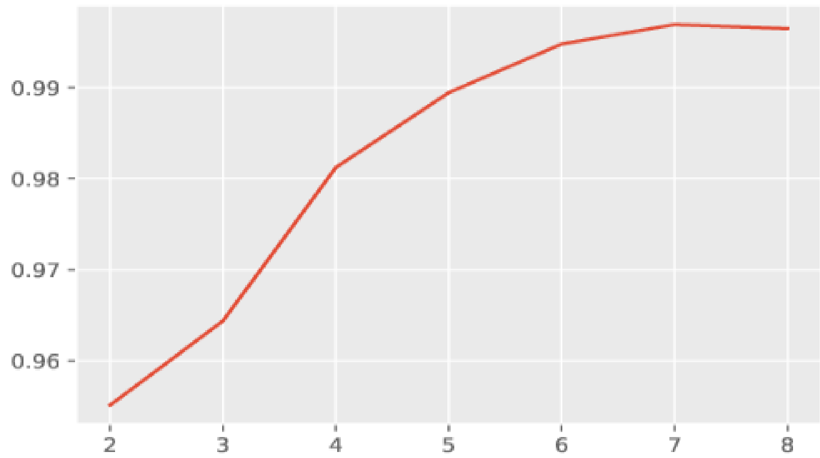


Figure 6.9.: Variance explained by principal convex hull analysis for two to eight clusters (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

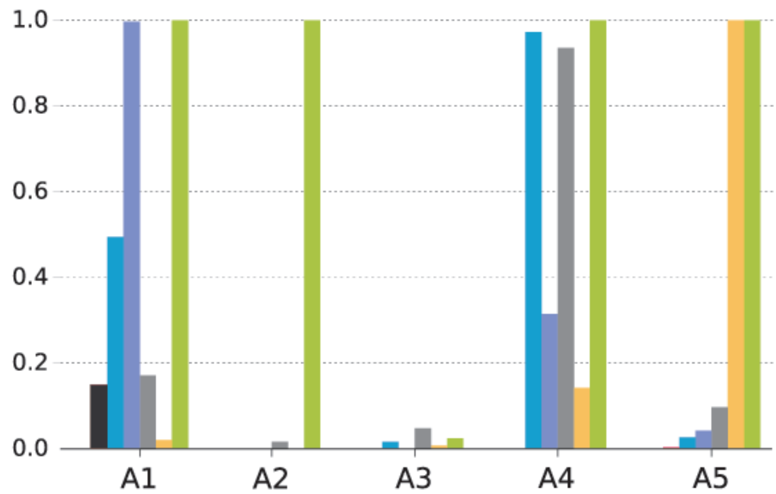


Figure 6.10.: Archetypal analysis using six features related to `the100.io` and *Destiny* (■ = Sherpa Score, ■ = Activity Score, ■ = Karma, ■ = Friend Count, ■ = Group Count, ■ = Level). Adapted from Wallner, Schinnerl, Schiller, Pirker, and Drachen (2019).

to help others, that is, *sherpas* are expected to have high levels of expertise themselves. Meanwhile, players in this archetype are only part of a small number of groups and do not have the highest number of friends (cf. A4). This observation coincides with the correlations found (cf. Table 6.5) – *sherpa* score and number of friends, expressed by the degree centrality, as well as *sherpa* score and group count are only weakly correlated. Considering all archetypes, A1 also exhibits the highest *karma* value out of all of them. Since *karma* is designed to be used as a reward for helpful or friendly players, observations indicate that users of the100.io use it for its intended purpose – to express their appreciation for other players.

A2 & A3: Players associated with either of these two archetypes can best be described as those who are not actively using the LFG platform the100.io. Either they have *not yet* used the platform actively or they lost interest shortly after their initial sign-up. This is demonstrated by low values for activity score, and few friends and groups. The principal difference between players in A2 and A3 is the level as determined by their Guardian. Players associated with A2 show a high character level indicating that they are experienced players while players in A3 only exhibit low values for their character level suggesting that they are new to the game *Destiny*. Table 6.8 shows that players primarily linked to A2 (experienced player) by far outnumber those who are connected to A3 (inexperienced players). This can be explained by the overall distribution of players showing that only a small fraction of players is found within the lower character levels.

A4: This archetype can be described as the one related to the *power users* of the100.io. These users exhibit the highest activity scores, and have the most friends. Furthermore, users connected to this archetype show high character levels. A high activity score would indicate that these users join a multitude of game sessions offering them opportunities to level up their characters. Additionally, the number of friendship strongly positively correlates with *karma* (cf. Table 6.5), hence, players belonging to this archetype display moderately high values for their *karma* score.

A5: The final archetype discussed here, is characterized by a high group count and a low friendship count. This suggests that players belonging to this archetype have a hard time finding others with whom to play.

This is indicated by the low activity score suggesting that players related to this archetype do not join many sessions. It is possible that these players are not satisfied with their current options of groups and, thus, are searching for others which they can join.

	Number of Players
A1	1,749
A2	39,588
A3	616
A4	1,289
A5	409

Table 6.8.: Distribution of players when assigned to their dominant archetype (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

From examining archetypes related to active players, that is, `A1`, `A4` and – to some extent – `A5`, we can see that these archetypes also exhibit high character levels. This implies that frequent users of `the100.io` improve the gameplay of *Destiny* by leveling up, or that the services offered by `the100.io` attract more advanded audiences. After examining the results of applying AA to users, in the next section similar steps as they were applied to group-level data will be described.

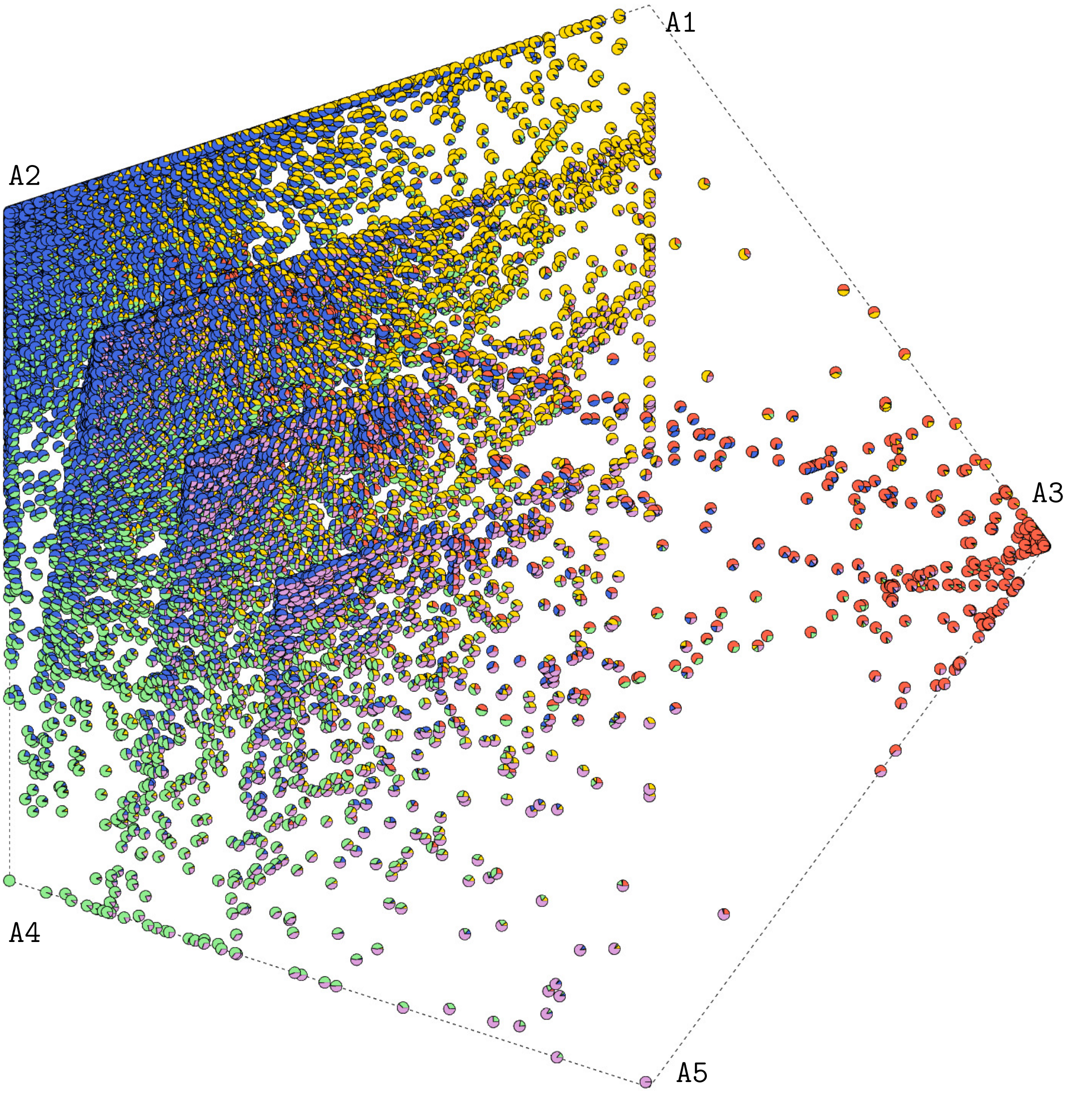


Figure 6.11.: Players degree of membership to each of the five archetypes (■ = A1, ■ = A2, ■ = A3, ■ = A4, ■ = A5). Each corner of the pentagon represents one archetype. Each pie-chart represents one player, placed by weighting the positions of the corners based on the player's belongingness coefficients (Wallner, Schinnerl, Schiller, Pirker, & Drachen, 2019).

6.2.2. Groups

Since AA was already discussed at great length in Chapter 2 and briefly summarized in the previous section about applying AA to players, here it will be refrained from restating any mathematical or theoretical foundations. Before being able to investigate any archetypal structures on a group-level, certain features had to be selected. In order to promote interpretability, again, it was tried to keep the number of features at a reasonable level. For the subsequent analyses these features were selected: *activity score* due to being related to the group's overall activity, measures describing the group on a structural level, *group size*, *number of moderators*, *number of sherpas*, and *density*, as well as metrics describing how experienced players are within the game *Destiny*, *average level*, and *average Light level*. Once more, entities – in this case, groups – containing invalid information were excluded from further analyses. As all other measures were either calculated and offered by `the100.io`, for example, group size, and number of *sherpas*, or calculated on the constructed network, that is, density, most excluded groups were rejected due to invalid values for their average level or average *Light level*. After removing invalid groups, a total of 573 groups remained.

In another discovery phase, different values for the number of group archetypes, k, were tested and the resulting clustering was investigated. Figure 6.12 shows the number of clusters on the x-axis, and both the variance explained by the AA model on the primary y-axis (left), and the distortion (cf. Selim and Ismail, 1984) of the k-means model on the secondary y-axis (right). Despite AA and k-means having different targets for their respective optimization while also employing different search parameters, they found similar distributions for the group clustering. When applying hard clustering to the dominant archetype, similar associations with clusters could be found (see Table 6.9). k-means clustering shows a pronounced "elbow" (cf. Thorndike, 1953) pointing towards a four-cluster solution. In contrast, the scree plot for AA does not indicate a definitive number of clusters to choose. However, selecting higher values for k, such as, five or six, only results in `A2` being fragmented into smaller clusters. Since k-means was applied in order to support the selection of $k = 4$ archetypes, and because it yielded similar results (see Table 6.9, Figure 6.13), the specific results of k-means clustering will not be described in greater detail.

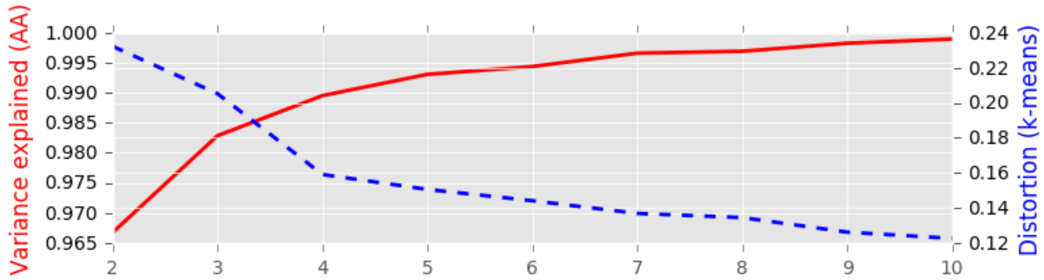


Figure 6.12.: Scree Plot for AA (red) and k-means (blue, dashed) Clustering for two to ten Clusters (Schiller et al., 2018).

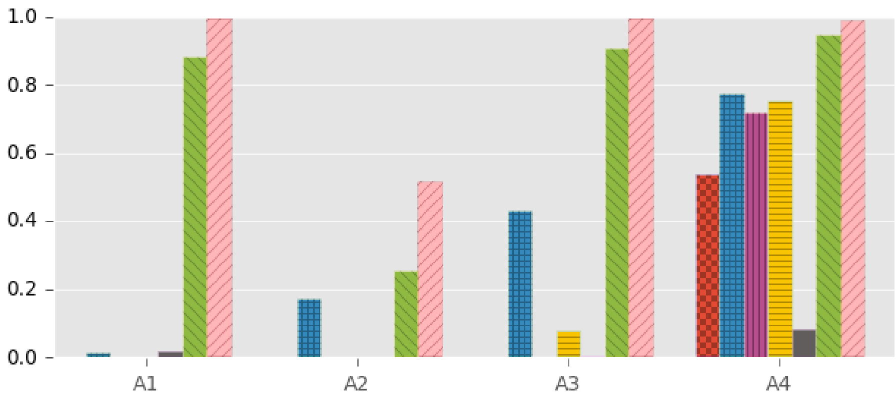
	A1	**A2**	**A3**	**A4**
AA	258	32	223	60
***k*-means**	240	31	217	85

Table 6.9.: Comparison of cluster sizes between AA and k-means

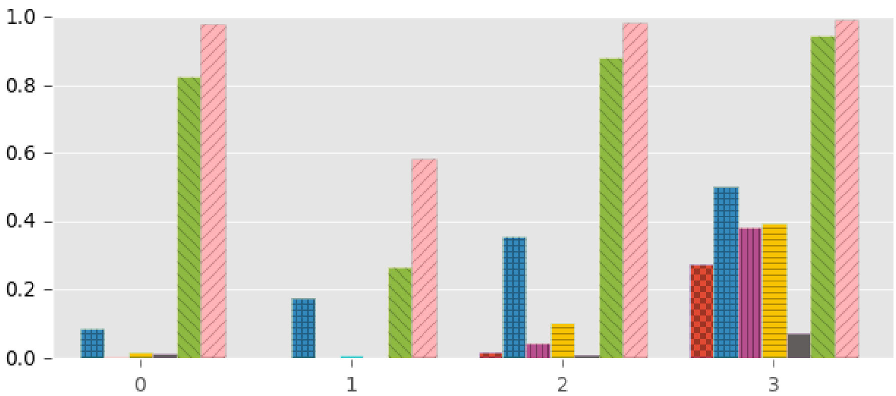
Figure 6.13 shows the profiles of clustering using $k = 4$ clusters applying both, AA (Figure 6.13a), and k-means clustering (Figure 6.13b). It should be noted again that AA does not result in hard clustering, leading to groups being a combination of archetypes – they belong to each archetype to a varying degree. Hereafter, the four different kinds of groups will be described in short.

`A1`: This archetype can best be described as a small, inactive group. Possibly due to the low activity within the group, there are no moderators present either. Furthermore, these groups also do not have any *sherpas* among their members. While members of this group show high levels of in-game experience – they exhibit relatively high values for average level and *Light level* – this circumstance may be explained by the nature of this dataset: users of the `the100.io` platform, generally, display high levels of in-game expertise. Players that are assigned to groups matching this archetype are most likely at risk of leaving the group or the `the100.io` website, altogether.

`A2`: This group archetype can be seen as a step up from `A1` with minor levels of activity. It is characterized by larger group sizes than displayed by



(a) Archetypal analysis (AA)



(b) *k*-means

Figure 6.13.: Group Profiles of four-cluster solutions (= Activity Score, = Group Size, = No. of Moderators, = No. of Sherpas, = Density, = Avg. Light Level, = Avg. Level). Adapted from Schiller et al. (2018).

`A1`, while still not containing noticeable numbers of either *sherpas* or moderators. Furthermore, `A2` exhibits inexperienced players which is indicated by the lowest average level and *Light level* found in any of the archetypes. Groups belonging to this archetype also feature low values for density, pointing towards few friendships within the group.

`A3`: Groups belonging to this archetype are similar to `A2` and can be described as evolved versions of the archetype. These groups are usually bigger than groups in `A2` and show increased levels of activity over the former. Furthermore, this archetype contains more veteran players than `A2` as suggested by the higher average level and *Light level*. Additionally, this archetype encompasses groups which have a small number of *sherpas* as their members. Lastly, these kinds of groups show a higher density than groups in `A2`, meaning that more friendship connections within the group were formed.

`A4`: This last archetype is characterized by high activity, and the largest group size of any of the archetypes. Furthermore, groups belonging to `A4` contain many in-group friendships and exhibit the largest numbers of *sherpas* and moderators. This archetype also exhibits the most advanced type of player groups as indicated by the highest average *Light level*. Similar to `A1` and `A3`, the archetype `A4` also displays very high values for the average character level which is explained by the dataset containing very advanced players. From a game design standpoint, it seems highly desirable to have a large portion of groups belonging to this archetype as the high activity can be seen as a measure for high player engagement.

After offering a descriptive overview of the groups' archetypes, prototypical groups were examined. Figure 6.14 shows a visualization for each of the group archetypes. It should be noted that each of the selected groups depicted is very similar to a pure archetype with a belongingness coefficient of $> .98$ to their respective archetype. In the visualization each segment of the surrounding chord chart represents one player that is a member of the examined group. The player's function within the group and on `the100.io` is encoded as the segment's color, with describing a "normal" member and indicating a player acting as a *sherpa*, that is, a player with a non-zero *sherpa* score. A segment colored marks a group's moderator, while the color marks a player as both a *sherpa* and moderator of the group. A connection between two segments indicates a friendship relation formed on `the100.io`. The central circle's background color encodes the group's activity score with lighter background indicating less activity and a darker color expressing higher levels of activity. As seen in Figure 6.14a,

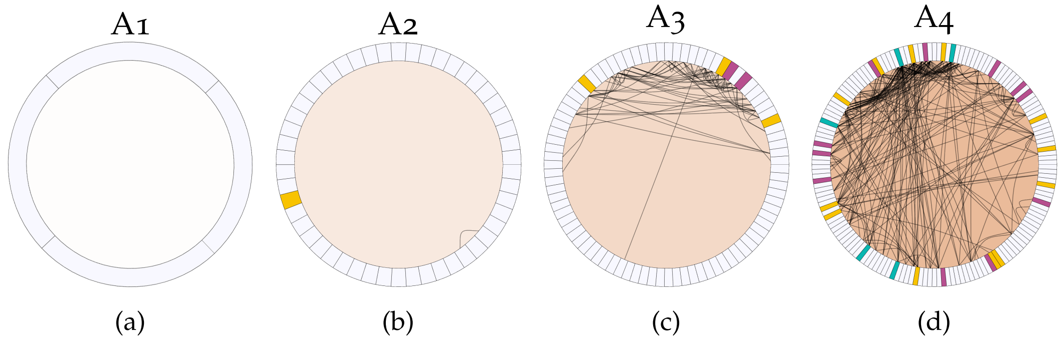


Figure 6.14.: Prototypical groups for each of the four archetypes. All groups have a belongingness coefficient greater than 0.98 with respect to the archetype in question. Each sector represents one group member colored with respect to the role in the group (■ = moderator, ■ = sherpa, ■ = sherpa & moderator). Friendships are indicated as lines. The background color of the inner circle reflects the group's activity score (0 ▬ 352). Adapted from Schiller et al. (2018).

the prototypical group for the first archetype is quite small in size with only four members, none of whom are *sherpas* or moderators. The formation of such groups is conceivable in two ways; either by `the100.io` having newly formed the group and members having to wait for others to join, or by an existing group disbanding and members leaving. This prototypical group also shows little to no activity. This observation could be explained by the low number of members which could result in fewer opportunities to play together. Furthermore, none of the members formed any friendships within the group. A prototypical group for `A2` is shown in Figure 6.14b. Here, the group size is already larger when compared to the prototype of `A1`. Additionally, this group exhibits a higher level of activity and contains a player which acts as a *sherpa*. However, this prototypical group only possesses a single friendship connection within its confines. With the prototype for the third archetype, we see another step towards a preferable group structure (Figure 6.14c) when compared to `A2`. Here, the group size is almost double the number of the prototypical group for the second archetype. Furthermore, an increase in numbers of *sherpas* can be observed, as well as the fact that moderators are present in the group. Additionally, this example group shows a more densely connected group structure with more friendships between players. Lastly, it can be seen that this group is more active than the former two prototypical groups. The last example, depicting a group with high belongingness to the archetype A4 is shown in Figure 6.14d. Once more, this prototypical group shows an increase in the number of members –it is almost double the size– when compared to the prototype of A3 (cf. Figure 6.14c). The A4 group also displays a greater number of both *sherpas* and moderators. Simultaneously, players taking on the dual role of *sherpa-moderator* are present in this type of group. Again, an increase in the group's activity can be seen when comparing it to the other archetypes. Lastly, it has to be noted that the prototypical group for the A4 archetype shows an even more densely connected group with more friendships than any other of the discussed groups.

In another analysis, hard clustering to the results of AA was conducted. In the process, each group was assigned to its dominant archetype, that is, the archetype with the highest belongingness coefficient. The overall descriptive statistics for these group clusters can be seen in Table 6.10. Here the results discussed so far are shown in more detail. It can be seen that A4 has the highest activity, both on average and when considering the maximum. Furthermore, we see that archetypes A1 and A2 contain relatively small groups with group sizes spanning from three to 85, while A3 contains larger groups and A4 contains larger groups, still. Additionally, it can be seen that A2 contains the least experienced players with low average character and *Light level*. Lastly, we see the relatively high density exhibited by A4 which has to be considered exceptional in light of the larger average group sizes.

Table 6.10.: Groups belonging to the different archetypes based on the highest membership value together with descriptive statistics of these groups. Adapted from Schiller et al., 2018.

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

6.3. Discussion

In regards to the distribution of playtime over the span of a week, it was found that users of `the100.io` tend to be more active on weekdays than on weekends. Overall, activity peaks on Tuesday evenings which coincides with weekly resets. At the time of scraping the data, weekly resets were performed at 2:00 AM Pacific Time (Bungie, 2017). During this reset, activities and rewards reset allowing players to earn new weekly rewards. Regardless of the day of the week, the most popular playtime seems to be 6:00 PM with high activity being observable over the course of the afternoon and evening. Therefore, `the100.io`, or LFG sites in general, could consider encouraging players to schedule sessions along trends of activity, thus, facilitating more satisfying game sessions and better player experience. When looking at the dataset's overview it has to be noted that `the100.io` appears to attract an audience of advanced, high-level players. This trend is observable across both serious and casual players with no pronounced difference between the two in this regard. Therefore, even players self-reporting their play style as "casual" to `the100.io` are most likely not casual players when considering the entire player base of *Destiny* as they have progressed far into the game and are highly engaged. Furthermore, it should be noted that selecting a suitable relation to encode as a link between nodes in a social network, is one of the challenges in social network analysis (SNA) (cf. Van De Bovenkamp et al., 2013). Especially in game-related analysis, it can be difficult as application programming interfaces (APIs) usually do not offer means to directly query social connections and a feature has to be selected which is then represented by links between players. In this regard, Ducheneaut et al. (2007) chose to create a link between two players if they were members of the same guild and online at the same time, while Iosup et al. (2014) connected players when they were playing in the same match. Since `the100.io` aims to help players overcome the shortcomings of *Destiny*'s in-game social features, the analyses presented here offer insights into online matchmaking services rather than into social structures within the game itself. Due to the dataset containing all information from the inception of `the100.io` until the time of data collection, it is not be expected that there would be a systemic sampling error. While players in online games tend to create multiple characters (cf. Ducheneaut and Moore, 2004; Williams et al., 2006), it seems unreasonable to assume that users of `the100.io` create multiple accounts on the platform as `the100.io` tries to match users based on their preferences and not necessarily a per-character basis. Furthermore, it seems easier for a player to only manage one `the100.io` account instead of multiple. As SNA was a part of the analyses presented here, it should be noted that approximately 75% of users did not connect to any other users by means of a friendship. Even when only considering the LCC of the friendship network, it should be noted that most users only maintain a small number of friendships. This finding is surprising as befriending another player on `the100.io` has the added benefits of being alerted when a friend is online and being informed about friends' upcoming sessions in order to join them. This observation seems to confirm players' want to be "alone together" (Ducheneaut et al., 2006) – they are playing concurrently but are not interested in interacting with each other. While Ducheneaut et al. observed this phenomenon for *World of Warcraft (WoW)*[6], Shen (2014) reported similar findings for *EverQuest II*[7]. Albeit, social features are offered to players and, even though, social behavior is often encouraged players opt-out from interacting with others and try to mitigate any dependence on others. Similarly, users of `the100.io` seem to be more interested in being offered ad-hoc team mates rather than building long-term relationships with others. If players' aim was to be able to play more and to improve player experience by playing with better team mates they should aim at finding more friends – players who are better embedded in the social network, exhibit higher levels of activity. This finding is consistent with Jia et al. (2015) who made similar observations in their analysis related to the multiplayer online battle arena (MOBA) game *Defense of the Ancients (DotA)*[8] and the now-defunct website *Dota-League*. As there is a strong positive correlation between the number of friendships and activity, maybe players who are at risk of leaving the game could be nudged towards adding new friends with whom to play. Due to the nature of the dataset at hand, it cannot be definitively said whether having more friends leads to more activity due to more opportunities to play together, or if being more active, that is, playing for longer periods of time leads to building more friendships.

[6] Blizzard Entertainment, 2003. https://worldofwarcraft.com.
[7] Sony Online Entertainment, 2004. https://www.everquest2.com.
[8] Eul, Steve Feak, IceFrog, 2003.

Additionally, positive effects in both directions with feedback loops are conceivable. Furthermore, *karma*, a mechanic implemented by `the100.io` to award helpful or friendly players seems to be a useful tool for community building. The public display of the *karma* score appears to incentivize good player behavior. It was found that *karma* is highly correlated with centrality, activity and in-game experience measures. Additionally, *sherpas* exhibit slightly higher *karma* than non-*sherpas*. However, since *sherpas* do only display slightly higher connectivity than *non-sherpas*, the determining factor for receiving *karma* seems to be embeddedness within the social structures rather than acting as a *sherpa*. While the *sherpa* score is a site-created measure, *karma* is a user-driven and community-driven metric. The results presented here, seem to offer some support for giving users means to police the community and incentivize desirable behavior. Overall, *karma* seems to work more reliably than *sherpa* scores. At first glance, the results in regards to group sizes appear to defy theories and results related to upper limits for social networks. For example, Dunbar (1993) suggested an upper limit for maintainable social connections to be roughly 150 – 200, now known as Dunbar's Number. This hypothesis has seen some corroboration in social media, for example on *Twitter* (Gonçalves et al., 2011). When focusing on video game groupings, Ducheneaut et al. (2007) found a much lower limit when examining guilds within *WoW*. Most guilds exhibit sizes of around 35 with guilds becoming unstable once they reach 60 members or more. The difference with the findings in this thesis and as published by Schiller et al. (2018) can be explained by the distinct purposes, guilds and groups on LFG sites, serve. The former are meant to be long-term connections with more personal or, even, intimate relationship – guild members are often encouraged to trade or exchange items by means of a shared stash – while the latter are designed to offer ad-hoc entertainment by facilitating fast-paced grouping and matchmaking. For this reason, it seems desirable to create groups as large as possible as this likely exposes users to more chances of joining an upcoming or ongoing session. The group-level results found that larger groups tend to be more organized with higher numbers of moderators. Moderators in combination with their connectedness, seem to be the largest predictor of group activity. However, overall connectedness of the group seems to not affect the activity in a meaningful way. Therefore, it appears to be desirable to ensure that smaller groups also have a moderator. Possibly this can be facilitated by lowering the barrier to become a moderator within groups that do not have any (cf. subsection 4.1.2).

In this section, some caveats in regards to the analyses were discussed. Furthermore, clarifications and explanations for observed phenomena were offered, parallels to published works were drawn and differences were depicted. Additionally, it was attempted to give possible explanations where results differ from published studies. In the following section, some limitations of the present analyses will be discussed.

6.4. Limitations

First, it has to be noted that `the100.io` attracts an audience of advanced high-level players. This trend is observable across both serious and casual players with no pronounced difference between the two in this regard. Therefore, even players self-reporting their play style as "casual" to `the100.io` are most likely not casual players when considering the entire player base of *Destiny* as they have progressed far into the game and are highly engaged. The network analyzed here was constructed from friendships expressed explicitly on `the100.io`. As there are no other features to support friendship connections, the true strength of a link cannot be verified within the scope of this thesis. While the number of active games was analyzed and it impacts the activity score of both players and groups, within the `the100.io` dataset itself there is no sensible way to check whether a game session scheduled on the platform actually took place or how many of the users signed up for them actually participated in the end. Additionally, when assuming that the game sessions happened it cannot be said how long they were ongoing. It is conceivable that some portion of them was played until the very end while others fell apart due to players leaving the lobby, disconnecting or similar reasons. Furthermore, the analyses at hand were targeted at a single game, *Destiny*. Since the game-related measures, character level, and *Light level*, did not have a substantial impact on activity it is possible that the presented results can also be applied to other games. However, further testing will be necessary before this could be concluded.

6.5. Summary

In this chapter analyses performed on the two datasets introduced in Chapter 4 were presented. First, general measures and descriptive stats were discussed. In this regard for the *Destiny* dataset magnitudes for numbers such as the the number of unique players, the number and distribution of in-game characters, *Guardians*, across character classes, and the number of game sessions were presented. Furthermore, the number and distribution of matches across the different game modes, as well as ranges for a wide variety of in-game performance metrics, such as K/D ratio, number of kills or deaths, as wells as the number of medals awarded were discussed. Additionally, surprising measures were presented. These included the longest match being 581,640 seconds long, thus, almost spanning an entire week, as well as high scores for kills, being over 1,200 and a player's unusually high K/D ratio of 49. Next, a short overview of the `the100.io` data was given. This overview included reporting the number of observations for all of these measures: groups, player, game sessions, and friendships. Furthermore, the numbers for distinctive players, such as, moderators and *sherpas* were presented. Additionally, it was shown that while `the100.io` facilitates matchmaking for other games as well, *Destiny* is the most popular game which is indicated by its reach, having more than ⅔ of groups list it on their profile as a game they are interested in playing. Next, the players' time zones and preferred platform were examined and discussed. The players overwhelmingly indicated US & Canada time zones as theirs and answered that they would prefer current-generation platforms, that is, PS4 and Xbox One over previous-generation platforms consisting of PS3 and Xbox 360 which, in turn, were more popular than the PC platform. In another analysis of players' self-reported data, their preferred play time was investigated. Here, only a very small minority of players preferred to only play on weekends while most players favored playing on weekday late-nights and weekends. Further investigation showed a daily distribution of scheduled which repeats for every day of the week. This distribution is characterized by the number of sessions rising from 12 midnight to 12 noon with a small drop afterwards until approximately 2 PM. At this point, the number of scheduled sessions starts to rise again culminating in the highest amount of games at 6 PM after which it falls again. The highest peak of any of the days of the weeks is observed on Tuesday which coincides with weekly resets. Unexpectedly, weekends show noticeably lower levels of activity. Next, descriptive statistics of numerical features, that is, ranges, averages and standard deviations were presented, as well as the number of entries missing values for each feature. It should be noted, that most often the feature "play style" was missing, followed by the "profanity" preference and the in-game metric of the *Light level*. Following these descriptive statistics and high-level analyses, the social network constructed from friendships formed on `the100.io` was described. Since the resulting graph was disconnected, only the LCC was analyzed – it contained approximately 85% of users. Next, community detection was applied to this sub-graph. Visual inspection of the resulting network showed a clear divide between the two platforms PlayStation and Xbox. Two principal communities formed which each consisted predominantly of players preferring one of the current-generation platforms or the other – PS4, or Xbox One. Further more two smaller communities formed which, again, were mostly dedicated to one platform – in this case the previous-generation platforms, Xbox 360, and PS3. All remaining communities overwhelmingly showed connections towards one or both of the principal communities while links between two smaller communities were rare. Furthermore, two communities were entirely dedicated to playing on PC, which is unexpected since *Destiny* was only released to console platforms. For the communities detected this way, descriptive statistics were offered. These include, for example, the number of users, the density, average level, activity as well composite measures for the preferred platform and the preferred play time. Following this SNA for users of `the100.io`, correlations between the different measures were calculated. In this analysis, strong correlations were found for activity and centrality measures, as well as for activity and *karma*, and for centrality and *karma*. After these user-level examinations, the focus for analysis was shifted towards the group level. First, descriptive statistics as well as graphs providing an overview of the group information were discussed. Next, correlations between group measures were calculated. At first inspection, most of the measures seemed to be strongly correlated with one another. Since the main focus of this thesis lies on activity, correlations with activity were investigated further. Using multiple regression it was shown that only three variables, that is, the number of moderators, the moderators' average degree and the group size show significant correlations with activity. Using these three variables, a derived model can account for approximately 73% of the variance in activity. Next, AA was applied to both players and groups. For clustering players a five-cluster solution was chosen. Looking at the distribution of players hard-clustered to their dominant archetype it was found that a single archetype dominated all other containing approximately 40,000 of roughly 45,000 players. When applying AA to groups, finding the correct number of archetypes was not as obvious. For this reason, an additional k-means clustering was undertaken which indicated that a four-cluster solution would describe the data well. For each of the found archetypes, a prototypical group with high belongingness was examined and discussed. Lastly, an overview of descriptive statistics of groups assigned to their dominant archetypes was given. In the discussion section following the analysis further clarifications and caveats were discussed. Furthermore, parallels to other publications were drawn while differences were highlighted and possible explanations for them were given. Concluding this chapter, limitations of the current study were highlighted. In this regard, it has to be said that all analyses shown only relate to the users of `the100.io`. Since these players have progressed far into the game, the results cannot be easily generalized to the entire player base of *Destiny*. Furthermore, the strength of friendship ties could not be verified using the data at hand. Lastly, there is also no conceivable way of checking whether a game scheduled on `the100.io` took place or how it developed over time.

7. Conclusion and Outlook

At the end of this thesis and after showing analyses and discussing results in the last chapter, this chapter will tackle conclusions which can be drawn from the results. These conclusions will be discussed in detail as well as calls to action. Finally, avenues for future research will be described.

7.1. Conclusion

In this thesis, it has been shown that multiplayer online games (MOGs), and massively online multiplayer games such as *Destiny* require users to be able to find companions with whom they can play. Therefore, it is important to facilitate this kind of social play in an easy and accessible manner. Being able to analyze and understand social structures can help create better player experience and increase player retention. This way it is not only of academic interest but relevant for game companies. Understanding patterns in player and group behavior gives game developers insights into players' needs and how they interact with the game and with each other. Hence, knowing their player base, helps game companies since it enables them to design challenges and incentives around the players. For example, being able to identify players or groups which exhibit low levels of activity before they leave or disband can help game developers as they can employ countermeasures in time, such as suggesting friendships to players with similar levels of experience and similar play schedules. Additionally, an inexperienced player could be offered a friendship connection to a *sherpa* who helps him by making him familiar with the game. For groups, recommendations could be facilitated which try to move an inactive group from one archetype over to another until it, hopefully, arrives at a high level of activity. These high-activity groups indicate that players are invested in the game and enjoy playing it. As such, the thesis at hand offers a first step towards understanding game communities and help foster a long-term player base. In this regard, it was shown how activity of both players and groups can be impacted by being connected and –in the group case – by the presence of active group moderation. Archetypal analysis was applied to find patterns of both user and group behavior and to find actionable differences between the groups.

7.2. Future Work

In future work, the effect of matchmaking websites such as `the100.io`– the one analyzed in thesis – on in-game measures should be investigated. This not only encompasses comparisons between users who use a looking for group (LFG) service and those who do not, but should include comparisons between progression rates and performance before and after joining a matchmaking site. Furthermore, comparative studies examining different matchmaking websites are conceivable to find differences and similarities between them. Temporal analyses can be considered another route worth investigating. In this regard, evolution of users and groups, the friendships formed and group compositions could be examined. Another possible network which could be constructed is given by the relationship between players who join the same game sessions (cf. Iosup et al., 2014). Building on the results presented here, also a recommender system could be devised which offers suggestions to players. This recommendations could include friend suggestions for players and suggesting players to add to a group for group moderators to move the group towards an optimal composition. Lastly, and since `the100.io` added support for other games, similar analyses as presented here could be applied to other games. Doing this, similarities and differences in the formation of groups and communities could be examined. Furthermore, the impact of game modes present in the game on the structure of the social network can be explored.

List of Figures

1.1. Thesis Structure 4

2.1. Congruities: Myers-Briggs Type Indicators and Keirsey Temperaments 10
2.2. Keirsey Temperament Matrix 11
2.3. Interest Graph according to Bartle 16
2.4. Action-based KeyGraphs for different player types 21
2.5. Seven Bridges of Königsberg 26
2.6. A Network Exhibiting the Community Structure Property . . 28
2.7. Balanced Triads 31
2.8. Illustration of relaxed PCH/AA-δ 44
2.9. Comparison: vehicle clustering k-means / k-maxoids 50

3.1. Destiny Gameplay 57
3.2. Destiny Classes 58

5.1. Data Preprocessing Procedure 75
5.2. Destiny Dataset: Data Model 84
5.3. the100.io Dataset: Data Model 86

6.1. Distribution of Games Across Different Game Modes 93
6.2. Class Distribution of Guardians 94
6.3. Distribution of Video Games Supported by the100.io 96
6.4. User-Reported Preferred Platform 98
6.5. Distribution of Weekly Activity 101
6.6. Description of Player-Level Measures 105
6.7. Visualization of the Community Structure on the100.io 107
6.8. Histograms of Group-level Characteristics. 114
6.9. Scree Plot Showing Variance Explained for Different Numbers of Clusters (User-Level) 121

6.10. Archetypal Analysis - Player Profiles. 121
6.11. Degree of Archetype Membership for Players 124
6.12. Scree Plot for AA and k-means Clustering (Group-Level) . . . 126
6.13. Group Profiles 127
6.14. Prototypical Groups according to AA. 129

List of Tables

2.1. MBTI - Bartle Player Types Overlap 17
2.2. MBTI - Bartle Player Types Correspondences 18
2.3. Bartle's Player Types - Factor Analysis: Components 19
2.4. Comparison: Psychology in Games 24
2.5. Social Analyses – A Comparison 38
2.6. Comparison of Clustering Algorithms 53

4.1. Key Features of the the100.io Dataset 66
4.2. Key Characteristics of the Destiny Dataset 70

6.1. Distribution of Preferred Playtime 99
6.2. Descriptive Statistics of User-Level Data 103
6.3. Number of Missing Values from User-Level Data 103
6.4. Community Descriptions. 108
6.5. Spearman Rank Correlation Between User-Related Variables. 109
6.6. Spearman Rank Correlations Between Group-Related Characteristics. 117
6.7. Multiple Linear Regression of Group Characteristics on Group activity. 118
6.8. Distribution of Players' Main Archetype. 123
6.9. Comparison between Clusterings: AA and k-means (Group-Level) 126
6.10. Aggregate Group Statistics after AA. 131

List of Listings

5.1. Hash-only Object 77
5.2. Medal Entry with Weights 78
5.3. User Data 79
5.4. Match Data 81
5.5. Original Variable Encoding 82
5.6. Weapon Statistics 82

Acronyms

AA archetypal analysis. 2, 41–47, 49, 51, 52, 55, 92, 118–120, 123, 125–127, 130, 138, 142, 143
API application programming interface. 3, 35, 66, 69, 73, 74, 76, 77, 82, 90, 132

DotA Defense of the Ancients. 33, 36, 133
DTR Destiny Tracker. 82, 83, 90

F2P free-to-play. 6, 46
FPS first-person shooter. 37, 46, 56, 64, 96, 112

GMS Gaming Motivation Scale. 18, 20–24, 54

HDF hierarchical data format. 74, 76, 77, 80, 83, 85, 86, 90

IGD internet gaming disorder. 14, 22, 24
IPIP International Personality Item Pool. 13, 22

LCC largest connected component. 27, 36, 37, 99, 104, 106, 133, 137
LFG looking for group. 2, 64–66, 73, 74, 95, 104, 122, 132, 134, 140

MBTI Myers-Briggs type indicator. 7, 9, 10, 12, 13, 17, 18, 54
MMORPG massively multiplayer online role-playing game. 15, 18, 21, 47, 56, 64
MOBA multiplayer online battle arena. 33, 35, 133
MOG multiplayer online game. 31, 34, 139

NEO PI-R revised NEO Personality Inventory. 12, 13
NMF non-negative matrix factorization. 42, 43, 48, 55
NPC non-player character. 20, 57, 61

PCA principal component analysis. 18, 43

PS3 PlayStation 3. 97, 98, 106–108, 112, 136, 137
PS4 PlayStation 4. 97, 98, 104, 107, 108, 112, 136, 137
PvE player-versus-environment. 5, 20, 51, 59, 60, 62, 64
PvP player-versus-player. 5, 37, 51, 59, 60, 62, 64, 71, 73, 83, 92

RPG role-playing game. 14, 15, 56, 64

SIVM simplex volume maximization. 42, 49
SNA social network analysis. 2, 3, 25, 28, 29, 31, 36, 38, 39, 54, 132, 133, 137

TIPI Ten-Item Personality Inventory. 13, 15

WoW World of Warcraft. 21, 22, 24, 31, 32, 38, 48, 133, 134

Appendix

A. Bartle Test

- Are you more comfortable, as a player on a MUD:
  - +S Talking with friends in a tavern?
  - +A Out hunting orcs by yourself for experience?
- Which is more enjoyable to you?
  - +A Killing a big monster
  - +S Bragging about it to your friends?
- Which do you enjoy more in MUD quests:
  - +S Getting involved in the storyline
  - +A Getting the rewards at the end?
- Which would you rather be noticed for on a MUD?:
  - +A Your equipment
  - +S Your personality
- Would you rather be:
  - +S Popular
  - +A Wealthy
- Which do you enjoy more on a MUD?:
  - +S Getting the latest gossip
  - +A Getting a new item
- Which would you rather have, as a player on a MUD?:
  - +S A private channel, over which you and your friends can communicate
  - +A Your own house, worth millions of gold coins
- Which would you enjoy more as a MUD player?
  - +S Running your own tavern?

+E Making your own maps of the world, then selling them?

- What's more important in a MUD to you?

  +S The number of people
  +E The number of areas to explore

- What's more important to you:

  +S The quality of roleplaying in a mud
  +E The uniqueness of the features, and game mechanic

- You are being chased by a monster on a MUD. Do you:

  +S Ask a friend for help in killing it
  +E Hide somewhere you know the monster won't follow

- You're a player on a mud, and you want to fight a really tough dragon. How would you approach this problem?

  +S Get a big group of players to kill it.
  +E Try a variety of weapons and magic against it, until you find its weakness.

- You're a player on a mud, and about to go into an unknown dungeon. You have your choice of one more person for your party. Do you bring:

  +S A bard, who's a good friend of yours and who's great for entertaining you and your friends
  +E A wizard, to identify the items that you find there

- Is it better to be:

  +K Feared
  +S Loved

- Someone has PK'ed you. Do you want to:

  +S Find out why, and try to convince them not to do it again
  +K Plot your revenge

- Which is more exciting?

  +S A well-roleplayed scenario
  +K A deadly battle

- Which would you enjoy more?

  +K Winning a duel with another player

+S Getting accepted by a clan

- Would you rather

  +K Vanquish your enemies
  +S Convince your enemies to work for you, not against you

- What's worse:

  +K To be without power
  +S To be without friends

- Would you rather:

  +S Hear what someone has to say
  +K Show them the sharp blade of your axe

- On a MUD, a new area opens up. Which do you look forward to more?

  +E Exploring the new area, and finding out its history
  +A Being the first to get the new equipment from the area

- On a MUD, would you rather be known as:

  +E Someone who can run from any two points in the world, and really knows their way around.
  +A The person with the best, most unique equipment in the game

- Would you rather:

  +A Become a hero faster than your friends
  +E Know more secrets than your friends?

- Would you rather:

  +E Know where to find things
  +A Know how to get things?

- Which would you rather do:

  +E Solve a riddle no one else has gotten
  +A Getting to a certain experience level faster than anyone else

- Do you tend to:

  +E Know things no one else does
  +A Have items no one else does

- On a MUD, would rather join a clan of:

+E Scholars
+K Assassins

- Would you rather win:

  +E A trivia contest
  +K An arena battle

- If you're alone in an area, do you think:

  +E It's safe to explore
  +K You'll have to look elsewhere for prey

- On a MUD, would rather be known for

  +E Knowledge
  +K Power

- Would you rather:

  +K Defeat an enemy
  +E Explore a new area

- You learn that another player is planning your demise. Do you:

  +E Go to an area your opponent is unfamiliar with and prepare there
  +K Attack him before he attacks you

- You meet a new player. Do you think of him as:

  +E Someone who can appreciate your knowledge of the game
  +K As potential prey

- On a MUD, would you rather:

  +A Have a sword twice as powerful as any other in the game
  +K Be the most feared person in the game

- On a MUD, would you be more prone to brag about:

  +K How may other players you've killed
  +A Your equipment

- Would you rather have:

  +K A spell to damage other players
  +A A spell that increases the rate at which you gain experience points?

- Would you rather have:
  - +A Two levels of experience
  - +K An amulet that increases the damage you do against other players by 10%.
- Would you rather receive as a quest reward:
  - +A Experience points
  - +K A wand with 3 charges of a spell that lets you control other players, against their will. (charm person)
- When playing a video game, is it more fun to:
  - +A Have the highest score on the list?
  - +K Beat your best friend one-on-one?

Bibliography

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases VLDB* (Vol. 1215, pp. 487–499).

Allen, C. (2004). The dunbar number as a limit to group sizes. Retrieved March 28, 2019, from http://www.lifewithalacrity.com/2004/03/the_dunbar_numb.html

Alsén, A., Runge, J., Drachen, A., & Klapper, D. (2016). Play with me? understanding and measuring the social aspect of casual gaming. In *Proc. of aaai aiide player analytics workshop* (pp. 115–121).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: APA.

Andreasen, E. & Downey, B. (2001). The mud personality test. *The Mud Companion*, 33–35.

Aslak, U. (2016). py_pcha. https://github.com/ulfaslak/py_pcha.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. Retrieved March 28, 2019, from http://mud.co.uk/richard/hcds.htm

Bauckhage, C. & Sifa, R. (2015). K-maxoids clustering. In *LWA* (pp. 133–144).

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730. doi:10.1121/1.1906679

Biggs, N. L., Lloyd, E. K., & Wilson, R. J. (1999). *Graph theory 1736-1936*. Clarendon Press.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, *2*(1), 113–120. doi:10.1080/0022250X.1972.9989806

Bradley, P. S. & Fayyad, U. M. (1998). Refining initial points for k-means clustering. (pp. 91–99). Morgan Kaufmann.

Braun, B., Stopfer, J. M., Müller, K. W., Beutel, M. E., & Egloff, B. (2016). Personality and video gaming: Comparing regular gamers, non-gamers, and gaming addicts and differentiating between game genres. *Computers in Human Behavior*, *55*, 406–412. doi:10.1016/j.chb.2015.09.041

Briggs Myers, I. & Briggs Myers, P. (1995). *Gifts Differing: Understanding Personality Type*. Consulting Psychologists Press.

Bright, D. A., Hughes, C. E., & Chalmers, J. (2012). Illuminating dark networks: A social network analysis of an australian drug trafficking syndicate. *Crime, Law and Social Change*, *57*(2), 151–176. doi:10.1007/s10611-011-9336-z

Bungie. (2017). Destiny: Daily and weekly resets. Archived version from September 24, 2018. Retrieved August 7, 2019, from https://web.archive.org/web/20180924014411/https://www.bungie.net/en/Help/Article/45956

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, *70*, 066111. doi:10.1103/PhysRevE.70.066111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates. doi:10.4324/9780203771587

Costa, P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources Odessa, FL.

Csikszentmihalyi, M. (2008). *Flow: The psychology of optimal experience*. Harper Perennial Modern Classics.

Cutler, A. & Breiman, L. (1994). Archetypal analysis. *Technometrics*, *36*(4), 338–347. doi:10.1080/00401706.1994.10485840

Destiny Wiki community. (2014). Destiny Wiki. Retrieved July 12, 2019, from https://www.destinygamewiki.com

Destinypedia community. (2013). Destinypedia. Retrieved July 12, 2019, from https://www.destinypedia.com

Diestel, R. (2005). *Graph theory* (3rd ed.). Berlin New York: Springer.

Drachen, A., Canossa, A., & Yannakakis, G. N. (2009). Player modeling using self-organization in Tomb Raider: Underworld. In *IEEE Symposium on Computational Intelligence and Games* (pp. 1–8). IEEE. doi:10.1109/CIG.2009.5286500

Drachen, A., Green, J., Gray, C., Harik, E., Lu, P., Sifa, R., & Klabjan, D. (2016). Guns and guardians: Comparative cluster analysis and behavioral profiling in Destiny. In *IEEE Conference on Computational Intelligence and Games* (pp. 1–8). IEEE. doi:10.1109/CIG.2016.7860423

Drachen, A., Lunquist, E., Kung, Y., Rao, P., Klabjan, D., Sifa, R., & Runge, J. (2016). Rapid prediction of player retention in free-to-play mobile games. In *Proc. of AAAI AIIDE*.

Drachen, A., Sifa, R., Bauckhage, C., & Thurau, C. (2012). Guns, swords and data: Clustering of player behavior in computer games in the wild. In *IEEE Conference on Computational Intelligence and Games* (pp. 163–170). IEEE. doi:10.1109/CIG.2012.6374152

Dreżewski, R., Sepielak, J., & Filipkowski, W. (2015). The application of social network analysis algorithms in a system supporting money laundering detection. *Information Sciences*, *295*, 18–32. doi:10.1016/j.ins.2014.10.015

Ducheneaut, N. & Moore, R. J. (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* (pp. 360–369). ACM. doi:10.1145/1031607.1031667

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). 'Alone together?': Exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 407–416). CHI '06. Montréal, Québec, Canada: ACM. doi:10.1145/1124772.1124834

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2007). The life and death of online gaming communities: A look at guilds in World of Warcraft. In *Proceedings of the SIGCHI conference on human factors in computing systems - CHI '07*. ACM Press. doi:10.1145/1240624.1240750

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(04), 681. doi:10.1017/s0140525x00032325

Ediger, D., Jiang, K., Riedy, J., Bader, D. A., Corley, C., Farber, R., & Reynolds, W. N. (2010). Massive social network analysis: Mining twitter for social good. In *2010 39th International Conference on Parallel Processing* (pp. 583–593). doi:10.1109/ICPP.2010.66

Ermi, L. & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum imperialis Petropolitanae*, *8*, 128–140. Retrieved from http://eulerarchive.maa.org//pages/E053.html

Field, A. (2013). *Discovering statistics using ibm spss statistics* (4th ed.). Sage Publications Ltd.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*, 35–41. doi:10.2307/3033543

Gajadhar, B., de Kort, Y., & IJsselsteijn, W. (2008). Influence of social setting on player experience of digital games. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3099–3104). CHI EA '08. Florence, Italy: ACM. doi:10.1145/1358628.1358814

Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. of the National Academy of Sciences*, *99*(12), 7821–7826. doi:10.1073/pnas.122653799

Global Music Report 2019: State of the Industry. (2019). International Federation of the Phonographic Industry (IFPI). London, GB.

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of personality and social psychology*, *59*(6), 1216.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*(1), 7–28.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, *40*(1), 84–96.

Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number. *PLOS ONE*, *6*(8), 1–5. doi:10.1371/journal.pone.0022656

Gosling, S. D. [Samuel D], Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1

Gow, J., Colton, S., Cairns, P., & Miller, P. (2012). Mining rules from player experience and activity data. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Graham, L. T. & Gosling, S. D. [Samuel D.]. (2013). Personality profiles associated with different motivations for playing world of warcraft. *Cyberpsychology, Behavior, and Social Networking*, *16*(3), 189–193. doi:10.1089/cyber.2012.0090

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Harary, F. (1999). *Graph theory*. Reading, Mass: Perseus Books.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc. doi:10.1037/10628-000

Hirsh, J. B. & Peterson, J. B. (2008). Predicting creativity and academic success with a "fake-proof" measure of the big five. *Journal of Research in Personality*, *42*(5), 1323–1333. doi:10.1016/j.jrp.2008.04.006

Iosup, A., Van De Bovenkamp, R., Shen, S., Jia, A. L., & Kuipers, F. (2014). Analyzing implicit social networks in multiplayer online games. *IEEE Internet Computing*, *18*(3), 36–44. doi:10.1109/MIC.2014.19

Jia, A. L., Shen, S., Van De Bovenkamp, R., Iosup, A., Kuipers, F., & Epema, D. H. J. (2015). Socializing by gaming: Revealing social relationships in multiplayer online games. *ACM Transactions on Knowledge Discovery from Data*, *10*(2), 11:1–11:29. doi:10.1145/2736698

Johnson, D. & Gardner, J. (2010). Personality, motivation and video games. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (pp. 276–279). OZCHI '10. Brisbane, Australia: ACM. doi:10.1145/1952222.1952281

Kaufman, L. & Rousseeuw, P. (1987). Clustering by means of medoids. *Data Analysis based on the L1-Norm and Related Methods*, 405–416.

Keegan, B., Ahmed, M. A., Williams, D., Srivastava, J., & Contractor, N. (2010). Dark gold: Statistical properties of clandestine networks in massively multiplayer online games. In *2010 IEEE Second International Conference on Social Computing* (pp. 201–208). IEEE. doi:10.1109/SocialCom.2010.36

Keirsey, D. (1998). *Please Understand Me II: Temperament, Character, Intelligence*. Prometheus Nemesis Book Co.

Kennerly, E. (2004). Elements of the Psyche: Does Myers-Briggs trump Bartle? Retrieved March 29, 2019, from http://finegamedesign.com/personality.html

Kent, A., Berry, M. M., Luehrs, F. U., & Perry, J. W. (1955). Machine literature searching VIII. operational criteria for designing information retrieval systems. *American Documentation*, *6*(2), 93–101. doi:10.1002/asi.5090060209

Klimmt, C., Blake, C., Hefner, D., Vorderer, P., & Roth, C. (2009). Player performance, satisfaction, and video game enjoyment. In S. Natkin & J. Dupire (Eds.), *Entertainment Computing – ICEC 2009* (pp. 1–12). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kokkinakis, A. V., Lin, J., Pavlas, D., & Wade, A. R. (2016). What's in a name? ages and names predict the valence of social interactions in a massive online game. *Computers in Human Behavior*, *55*, 605–613. doi:10.1016/j.chb.2015.09.034

Koschade, S. (2006). A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence. *Studies in Conflict & Terrorism*, *29*(6), 559–575.

Krause, J., Croft, D. P., & James, R. (2007). Social network theory in the behavioural sciences: Potential applications. *Behavioral Ecology and Sociobiology*, *62*(1), 15–27. doi:10.1007/s00265-007-0445-8

Lehenbauer-Baum, M., Klaps, A., Kovacovsky, Z., Witzmann, K., Zahlbruckner, R., & Stetina, B. (2015). Addiction and engagement: An explorative study toward classification criteria for internet gaming disorder. *Cyberpsychology, behavior and social networking*, *18*, 343–349. doi:10.1089/cyber.2015.0063

Lin, C., Wu, L., Wen, Z., Tong, H., Griffiths-Fisher, V., Shi, L., & Lubensky, D. (2012). Social network analysis in enterprise. *Proceedings of the IEEE*, *100*(9), 2759–2776. doi:10.1109/jproc.2012.2203090

Luce, R. D. & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, *14*(2), 95–116. doi:10.1007/BF02289146

Lusseau, D. & Newman, M. E. J. (2004). Identifying the role that animals play in their social networks. *Proceedings. Biological sciences*, *271 Suppl 6*(Suppl 6), S477–S481. doi:10.1098/rsbl.2004.0225

Maher, B. (2016). Can a video game company tame toxic behaviour? *Nature*, *531*(7596). doi:10.1038/531568a

Märtens, M., Shen, S., Iosup, A., & Kuipers, F. (2015). Toxicity detection in multiplayer online games. In *2015 International Workshop on Network and Systems Support for Games (NetGames)* (pp. 1–6).

Mason, W. & Clauset, A. (2013). Friends FTW! Friendship, collaboration and competition in Halo: Reach. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 375–386). CSCW '13. San Antonio, Texas, USA: ACM. doi:10.1145/2441776.2441820

Milgram, S. (1967). The small world problem. *Psychology today*, *2*(1), 60–67.

Mørup, M. & Hansen, L. K. (2012). Archetypal analysis for machine learning and data mining. *Neurocomputing*, *80*, 54–63. Special Issue on Machine Learning for Signal Processing 2010. doi:10.1016/j.neucom.2011.06.033

Nacke, L., Drachen, A., Kuikkaniemi, K., Niesenhaus, J., Korhonen, H., M. Van Den Hoogen, W., . . . De Kort, Y. (2009). Playability and player experience research. In B. Atkins & H. Kennedy (Eds.), *Proceedings of DiGRA 2009*. Breaking new ground: innovation in games, play, practice and theory. DiGRA.

Nacke, L. & Lindley, C. (2008). Flow and immersion in first-person shooters: Measuring the player's gameplay experience. (pp. 81–88). doi:10.1145/1496984.1496998

El-Nasr, M. S., Drachen, A., & Canossa, A. (Eds.). (2013). *Game analytics*. Springer London. doi:10.1007/978-1-4471-4769-5

Nazir, A., Raza, S., & Chuah, C.-N. (2008). Unveiling facebook: A measurement study of social network based applications. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement* (pp. 43–56). IMC '08. Vouliagmeni, Greece: ACM. doi:10.1145/1452520.1452527

Newzoo. (2019). Most popular core pc games — global. Retrieved June 19, 2019, from https://newzoo.com/insights/rankings/top-20-core-pc-games

Ohsawa, Y., Benson, N. E., & Yachida, M. (1998). Keygraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In *Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries -ADL'98-* (pp. 12–18). doi:10.1109/ADL.1998.670375

Pearce, C. (2009). *Communities of Play: Emergent cultures in Multiplayer Games and Virtual Worlds*. Cambridge, Mass: MIT Press.

Peever, N., Johnson, D., & Gardner, J. (2012). Personality & video game genre preferences. In *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System* (20:1–20:3). IE '12. Auckland, New Zealand: ACM. doi:10.1145/2336727.2336747

Pelleg, D. & Moore, A. W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML* (Vol. 1, pp. 727–734).

Pirker, J., Rattinger, A., Drachen, A., & Sifa, R. (2018). Analyzing player networks in Destiny. *Entertainment Computing*, *25*, 71–83. doi:10.1016/j.entcom.2017.12.001

Pollard, D. (1981). Strong consistency of *k*-means clustering. *The Annals of Statistics*, *9*(1), 135–140.

Poor, N. D. (2014). Collaboration via cooperation and competition: Small community clustering in an MMO. In *47th Hawaii International Conference on System Sciences* (pp. 1695–1704). doi:10.1109/HICSS.2014.217

Poor, N. D. (2015). What MMO communities don't do: A longitudinal study of guilds and character leveling, or not.

Prell, C., Hubacek, K., & Reed, M. (2009). Stakeholder analysis and social network analysis in natural resource management. *Society & Natural Resources*, *22*(6), 501–518. doi:10.1080/08941920802199202

Rapoport, A. (1953). Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, *15*(4), 523–533.

Rattinger, A., Wallner, G., Drachen, A., Pirker, J., & Sifa, R. (2016). Integrating and inspecting combined behavioral profiling and social network models in Destiny. In G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, & H.-S. Yang (Eds.), *Proc. Entertainment Computing - ICEC 2016* (pp. 77–89). Springer. doi:10.1007/978-3-319-46100-7_7

Rigby, S. & Ryan, R. (2007). The player experience of need satisfaction (pens) model. *Immersyve Inc*, 1–22.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360. doi:10.1007/s11031-006-9051-8

Schiller, M. H., Wallner, G., Schinnerl, C., Monte Calvo, A., Pirker, J., Sifa, R., & Drachen, A. (2018). Inside the group: Investigating social structures in player groups and their influence on activity. *IEEE Transactions on Games*. doi:10.1109/TG.2018.2858024

Seif El-Nasr, M., Drachen, A., & Canossa, A. (2013). *Game analytics - maximizing the value of player data*. London: Springer. doi:10.1007/978-1-4471-4769-5

Selim, S. Z. & Ismail, M. A. (1984). K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(1), 81–87. doi:10.1109/TPAMI.1984.4767478

Shen, C. (2014). Network patterns and social architecture in massively multiplayer online games: Mapping the social world of EverQuest II. *New Media & Society*, *16*(4), 672–691. doi:10.1177/1461444813489507

Shirky, C. (2003). Power Laws, Weblogs, and Inequality. Retrieved August 5, 2019, from https://web.archive.org/web/20030220025754/http://www.shirky.com/writings/powerlaw_weblog.html

Sifa, R., Bauckhage, C., & Drachen, A. (2014a). Archetypal game recommender systems. In *Proc. of the 16th LWA Workshops: KDML, IR and FGWM* (Vol. 1226, pp. 45–56).

Sifa, R., Bauckhage, C., & Drachen, A. (2014b). The Playtime Principle: Large-scale Cross-games Interest Modeling. In *Proc. of IEEE CIG* (pp. 365–373).

Sifa, R., Drachen, A., & Bauckhage, C. (2015). Large-Scale Cross-Game Player Behavior Analysis on Steam. In *Artificial Intelligence and Interactive Digital Entertainment International Conference* (pp. 198–204). AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. AAAI Press.

Stafford, G., Luong, H., Gauch, J., Gauch, S., & Eno, J. (2012). Social network analysis of virtual worlds. In R. Huang, A. A. Ghorbani, G. Pasi, T. Yamaguchi, N. Y. Yen, & B. Jin (Eds.), *Proc. of the 8th International Conference Active Media Technology* (pp. 411–422). Springer. doi:10.1007/978-3-642-35236-2_41

Steinhaus, H. (1957). Sur la division des corps matériels en parties. *Bull. Acad. Pol. Sci., Cl. III*, *4*, 801–804.

Stepanyan, K., Borau, K., & Ullrich, C. (2010). A social network analysis perspective on student interaction within the twitter microblogging environment. In *2010 10th IEEE International Conference on Advanced Learning Technologies* (pp. 70–72). doi:10.1109/ICALT.2010.27

Sugar, C. A. & James, G. M. (2003). Finding the number of clusters in a dataset. *Journal of the American Statistical Association*, *98*(463), 750–763. doi:10.1198/016214503000000666

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proc. of the National Academy of Sciences*, *107*(31), 13636–13641. doi:10.1073/pnas.1004008107

Szell, M. & Thurner, S. (2010). Measuring social dynamics in a massive multiplayer online game. *Social networks*, *32*(4), 313–329.

Tekofsky, S., Van Den Herik, J., Spronck, P., & Plaat, A. (2013). Psyops: Personality assessment through gaming behavior. In *In Proceedings of the International Conference on the Foundations of Digital Games* (pp. 166–173).

Ter Wal, A. L. J. & Boschma, R. A. (2009). Applying social network analysis in economic geography: Framing some key analytic issues. *The Annals of Regional Science*, *43*(3), 739–756. doi:10.1007/s00168-008-0258-3

Thawonmas, R. & Iizuka, K. (2008). Visualization of online game players based on their action behaviors. *Int. J. of Computer Games Technology*, *2008*(1).

Theatrical Market Statistics 2016. (2017). Motion Picture Association of America (MPAA). Washington D.C., US.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. doi:10.1007/BF02289263

Thurau, C., Kersting, K., & Bauckhage, C. (2009). Convex non-negative matrix factorization in the wild. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 523–532). doi:10.1109/ICDM.2009.55

Thurau, C., Kersting, K., & Bauckhage, C. (2010). Yes we can: Simplex volume maximization for descriptive web-scale matrix factorization. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management* (pp. 1785–1788). ACM. doi:10.1145/1871437.1871729

Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social network analysis for organizations. *Academy of management review*, *4*(4), 507–519.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont.)* *28*(1), 16–26. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20505784

Valve Corporation. (2003). Steam. Digital Distribution Platform [Windows, macOS, Linux, iOS, Android, Windows Phone]. Bellevue, WA, USA: Valve Corporation. Retrieved from https://store.steampowered.com

Van De Bovenkamp, R., Shen, S., Iosup, A., & Kuipers, F. (2013). Understanding and recommending play relationships in online social gaming. (pp. 1–10). doi:10.1109/COMSNETS.2013.6465556

Vorderer, P., Hartmann, T., & Klimmt, C. (2003). Explaining the enjoyment of playing video games: The role of competition. doi:10.1145/958720.958735

Wallner, G., Schinnerl, C., Schiller, M. H., Pirker, J., & Drachen, A. (2019). Beyond the individual: Understanding social structures of an online player matchmaking website. *Entertainment Computing*. doi:10.1016/j.entcom.2019.01.002

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. doi:10.1038/30918

Weber, B. G., Mateas, M., & Jhala, A. (2011). Using data mining to model player experience. In *FDG Workshop on Evaluating Player Experience in Games*. ACM Press.

Wijman, T. (2019). The global games market will generate $152.1 billion in 2019 as the u.s. overtakes china as the biggest market. Retrieved June 19, 2019, from https://newzoo.com/insights/articles/the-global-games-market-will-generate-152-1-billion-in-2019-as-the-u-s-overtakes-china-as-the-biggest-market

Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E. (2006). From tree house to barracks the social life of guilds in World of Warcraft. *Games and culture*, *1*(4), 338–361. doi:10.1177/1555412006292616

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, *13*(4), 993–1018. doi:10.1111/j.1083-6101.2008.00428.x

Witt, E. A., Massman, A. J., & Jackson, L. A. (2011). Trends in youth's videogame playing, overall computer use, and communication technology use: The impact of self-esteem and the big five personality factors. *Computers in Human Behavior*, *27*(2), 763–769. Web 2.0 in Travel and Tourism: Empowering and Changing the Role of Travelers. doi:10.1016/j.chb.2010.10.025

Xu, J., Christley, S., & Madey, G. (2006). 12 - application of social network analysis to the study of open source software. In J. Bitzer & P. J. Schröder (Eds.), *The Economics of Open Source Software Development* (pp. 247–269). Amsterdam: Elsevier. doi:10.1016/B978-044452769-1/50012-3

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772–775. doi:10.1089/cpb.2006.9.772

Zuckerman, M., Porac, J., Lathin, D., & Deci, E. L. (1978). On the importance of self-determination for intrinsically-motivated behavior. *Personality and Social Psychology Bulletin*, *4*(3), 443–446.

Ludography

Bayer&Szell OG. (2004). *Pardus*. Game [Web browser]. Bayer&Szell OG, St. Pölten, AT. St. Pölten, AT: Bayer&Szell OG.

Blizzard Entertainment. (2003). *World of Warcraft*. Game [Windows, macOS]. Blizzard Entertainment, Irvine, CA, USA. Irvine, CA, USA: Blizzard Entertainment.

Blizzard Entertainment. (2010). *StarCraft 2: Wings of Liberty*. Game [Windows, macOS]. Blizzard Entertainment, Irvine, CA, USA. Irvine, CA, USA: Blizzard Entertainment.

Blizzard Entertainment. (2014). *Hearthstone: Heroes of Warcraft*. Game [Windows, macOS, iOS, Android]. Blizzard Entertainment, Irvine, CA, USA. Irvine, CA, USA: Blizzard Entertainment.

Blizzard Entertainment. (2016). *Overwatch*. Game [Windows, PlayStation 4, Xbox One]. Blizzard Entertainment, Irvine, CA, USA. Irvine, CA, USA: Blizzard Entertainment.

Bluehole Studio. (2011). *TERA: The Exiled Realm of Arborea*. Game [Windows, PlayStation 4, XboxOne]. NHN Corporation, Seongnam, KR. En Masse Entertainment, Seattle, WA, USA. Gameforge, Karlsruhe, DE. Bungdang-gu, Seongnam, KR: NHN Corporation.

Bungie. (2010). *Halo: Reach*. Game [Xbox 360]. Microsoft Game Studios, Redmond, WA, USA. Santa Monica, CA, USA: Microsoft Game Studios.

Bungie. (2014). *Destiny*. Game [PlayStation 3, PlayStation 4, Xbox 360, Xbox One]. Activision, Santa Monica, CA, USA. Santa Monica, CA, USA: Activision.

Crystal Dynamics. (2008). *Tomb Raider: Underworld*. Game [Windows, OS X, N-Gage 2.0, Nintendo DS, PlayStation 2, PlayStation 3, Wii, Xbox 360]. Eidos Interactive, Southwark, London, GB-ENG. Redwood City, CA, USA: Eidos Interactive.

EA DICE. (2010). *Battlefield: Bad Company 2*. Game [Windows, PlayStation 3, Xbox 360, iOS, Kindle Fire]. Electronic Arts, Redwood City, CA, USA. Stockholm, SE: Electronic Arts.

EA DICE. (2011). *Battlefield 3*. Game [Windows, PlayStation 3, Xbox 360]. Electronic Arts, Redwood City, CA, USA. Stockholm, SE: Electronic Arts.

Epic Games. (2017). *Fortnite*. Game [Windows, macOS, Nintendo Switch, PlayStation 4, Xbox One, iOS, Android]. Epic Games, Cary, NC, USA. Cary, NC, USA: Epic Games.

Eul, Steve Feak, IceFrog. (2003). *Defense of the Ancients*. Game [Windows, Mac OS X]. n/a: Self-Published.

Gygax, Gary and Arneson, David Lance. (1974). *Dungeons & Dragons*. Game. Wizards of the Coast, Renton, WA, USA. Wizards of the Coast.

Massive Entertainment. (2016). *The Division*. Game [Windows, PlayStation 4, Xbox One]. Ubisoft, Montreuil, FR. Malmö, SWE: Ubisoft.

Mojang. (2011). *Minecraft*. Game [Windows, macOS, Linux, Android, iOS, Windows Phone, Xbox 360, Xbox One, PlayStation 3, PlayStation 4, PlayStation Vita, Raspberry Pi, Universal Windows Platform, Wii U, Nintendo Switch, New Nintendo 3DS, tvOS, Fire OS]. Mojang, Stockholm, SE. Xbox Game Studios, Redmond, WA, USA. Sony Computer Entertainment, San Mateo, CA, USA. Stockholm, SE: Mojang.

Origin Systems. (1993). *Ultima Online*. Game [Windows, Linux]. Electronic Arts, Redwood City, CA, USA. Austin, TX, USA: Electronic Arts.

Respawn Enterntainment. (2016). *Titanfall 2*. Game [Windows, PlayStation 4, Xbox One]. Electronic Arts, Redwood City, CA, USA. Beverly Hills, CA, USA: Electronic Arts.

Riot Games. (2009). *League of Legends*. Game [Windows, macOS]. Riot Games, Los Angeles, CA, USA. Los Angeles, CA, USA.: Riot Games.

Sony Online Entertainment. (2003). *Star Wars Galaxies*. Game [Windows]. LucasArts, San Francisco, CA, USA. San Diego, CA, USA: LucasArts.

Sony Online Entertainment. (2004). *EverQuest II*. Game [Windows]. Sony Online Entertainment, San Diego, CA, USA. Ubisoft, Montreuil, FR. San Diego, CA, USA: Sony Online Entertainment.

Sony Online Entertainment. (2012). *PlanetSide 2*. Game [Windows, PlayStation 4]. Sony Online Entertainment, San Diego, CA, USA. San Diego, CA, USA: Sony Online Entertainment.

Valve Corporation. (2004). *Counter-Strike: Source*. Game [Windows, OS X, Linux]. Valve Corporation, Bellevue, WA, USA. Bellevue, WA, USA: Valve Corporation.

Valve Corporation. (2007). *Team Fortress 2*. Game [Windows, OS X, PlayStation 3, Xbox 360, Linux]. Valve Corporation, Bellevue, WA, USA. Bellevue, WA, USA: Valve Corporation.

Valve Corporation. (2012). *Counter-Strike: Global Offensive*. Game [Windows, OS X, PlayStation 3, Xbox 360, Linux]. Valve Corporation, Bellevue, WA, USA. Bellevue, WA, USA: Valve Corporation.

Valve Corporation. (2013). *Dota 2*. Game [Windows, OS X, Linux]. Valve Corporation, Bellevue, WA, USA. Bellevue, WA, USA: Valve Corporation.

Valve L.L.C. (2000). *Counter-Strike*. Game [Windows, Xbox, OS X, Linux]. Valve Corporation, Bellevue, WA, USA, Sierra Studios, Los Angeles, CA, USA. Bellevue, WA, USA: Valve L.L.C, Sierra Studios.

Wargaming Minsk. (2010). *World of Tanks*. Game [Windows, Xbox 360, Xbox One, PlayStation 4, iOS, Android, macOS]. Wargaming, Nicosia, CY. Nicosia, CY: Wargaming.