

# The Name in the Game: Patterns in Character Names and Gamer Tags

Rafet Sifa<sup>1</sup>, Anders Drachen<sup>2</sup> and Christian Thureau<sup>3</sup>

<sup>1</sup>R. Sifa is with Game Analytics, Lichtstrasse 45, 50825 Cologne, Germany and the University of Bonn (email: [rafet@gameanalytics.com](mailto:rafet@gameanalytics.com)).

<sup>2</sup>A. Drachen is with the PLAIT Lab, Northeastern University (360 Huntington Avenue, Boston MA 02115 USA) and the Department of Communication and Psychology, Aalborg University (A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark) (email: [andersdrachen@gmail.com](mailto:andersdrachen@gmail.com)) and Game Analytics (Refshalevej 147, 1., 1432 Copenhagen K., Denmark) (email: [anders@gameanalytics.com](mailto:anders@gameanalytics.com)).

<sup>3</sup>C. Thureau is with Game Analytics, Refshalevej 147, 1., 1432 Copenhagen K., Denmark (email: [christian@gameanalytics.com](mailto:christian@gameanalytics.com)).

**Abstract.** In online games, often the only truly unique thing about a player is the name associated with an online account or game character. In massively multiplayer online games like *World of Warcraft*, customization options are limited and in multiplayer games like *Battlefield*, there are no ways of customizing the appearance of the players' avatar. This lack of ability to visually distinguish oneself means that names become important in online games, and assists with building understanding of the people that play these games, which is notably important for games operating via Free-to-Play revenue models. Here, the first large-scale, cross-game analysis of virtual identities in games is presented, based on datasets spanning over eight million character names and gamer tags with associated behavioral data, from four major commercial game titles: the role-playing game *World of Warcraft* and the tactical shooters *Battlefield 2 Bad Company 2*, *Crysis* and *Medal of Honor*. The results highlight the inventiveness of the names players adopt for their characters or accounts, and describe two different patterns - or communities - of name usage: in *World of Warcraft*, player character names are distributed according to power laws, have semantic meaning (no numbers permitted in names), and names are influenced by the aesthetics and game function of characters, with some names even being predictors of particular classes and races. In the tactical shooters, where all gamer tags are unique, names comparatively more rarely have a clear semantic meaning (numbers and special characters are permitted and often used), and name components are not distributed according to power laws. However, there is to a degree of a non-random relationship between gamer tag and in-game behavior. These results indicate that the name chosen by players for their characters or tags are useful for profiling purposes and provide information for the design and development of games in terms of how names are perceived by the players.

**Keywords:** online identity, character name, gamer tag, character name, avatar.

## 1 Background, introduction and motivation

The digital game industry is currently undergoing what may turn out to be a major shift in the mode of operations, in the sense that business models are changing from traditional retail-based systems towards a greater variety of models that is taking advantage of the plethora of distribution platforms and the increased mobile network and broadband coverage worldwide. This to the extent that the retail-based model has come under pressure [45]. One of the clearest indicators of this change is the rise of the online game, i.e. digital games played over a network (LAN or Internet); and the introduction of revenue generation models such as Free-to-Play [39].

Online games are exactly what the term implies – games played via online networks across one or more players (typically multiple), interacting with or via a server. Online game forms such as the persistent-world type Massively Multiplayer Online Game (MMOG) (e.g. *World of Warcraft* (WoW), *Guild Wars*, *EVE Online*) or the multi-player online shooter (e.g. *Counterstrike*, *Team Fortress*), as well as casual games with varying degrees of persistence (e.g. *Farmville*, *Clash of the Clans*) and virtual worlds (*Second Life*). Even traditional single-player franchises (e.g. the *Mass Effect*-series) is increasingly seeing online components and online distribution of content post-launch (Down Loadable Content, DLC), championed by the use of distribution platforms such as Steam.

There are several factors that have been quoted as driving the move from offline to online play, including security and digital piracy [40], but one of the most discussed in game industry circles is the new revenue models opening up via online games, which alleviates one of the traditional problems with the retail model, which inserts an expensive middle link (the retailer) in between the publisher and the user, and limited the revenue streams available. In an online environment controlled by the publisher, new revenue streams open up, including the sale of virtual items [39].

Contemporaneously with the rise of the online game format, the analysis of player behavior has become important to monitor the population of persistent virtual worlds, or evaluating game designs based on actual user behavior. In essence, monitoring and analyzing the behavior of the users provides insights on the large scale, whereas traditional lab-based testing methods were limited in practice in the sample size attainable [39,41,42,43]. One of the key goals of these forms of game analytics is to profile and classify players based on their behavior. This is notably important in the Free-to-Play (F2P) revenue model, where locating the players who might purchase virtual items for real money, and examining the behavior leading to purchases, is of perennial interest to developers and investors alike [39]. Hitherto, research in industry and academia on player profiling has been limited to in-game behaviors and some demographic considerations, and ignores the virtual identity that a player (user) adopts in an online game context – for example the name of a virtual character or gamer tag.

Formally, a virtual identity or online identity is a social entity that an internet user establishes in an online community. In many digital game contexts, such as the Xbox Live service, gamer tags or character names are fixed after the initial selection, and anonymous (in some cases, e.g. *World of Warcraft*, player character names can be changed against payment). Because such names are fixed, players are accountable to others in the community, and a reputation mechanism is established, for example among players interacting in the context of MMOGs [1,49]. These digital games form virtual societies where users control avatars, or more precisely characters, interacting with the elements of the world as well as with other players [2, 3]. From a purely scientific perspective, online identities and user profiling are of interest because technological development has enabled people to interact via networks on a worldwide basis, not the least through Internet based applications [4]. A notable domain investigating online behavior is cybersecurity [44]. Digital online games form a vehicle of entertainment and in some cases also socialization, and the number

of users are increasing with the increased adoption of networked technologies in contemporary society [2,3,5]. At the fundamental level, understanding how people operate online, the identities they assume, motivations for behavior and their interactions with each other and various environments (web forums, online games, social sites ...), is vital to the evaluate the effect network technologies have on human society as such.

### **1.1 Motivation, contribution and results**

The fundamental goal of the research presented here is data-driven and explorative. This because within the games domain, there is not enough available empirical research to form scientific theories about player behavior or identities – at least this has to the best knowledge of the authors not been attempted and validated yet [45]. While theories outside the domain of digital games can be imported and applied in the context of games, doing so requires contextual data about the users that are not available for the current study (but see e.g. 50,51 for other situations where contextual data are used in conjunction with behavioral data). Finally, character names and gamer tags have not previously formed the subject of large-scale quantitative research, and it is therefore not known whether there are any patterns in the names used by players, and whether these patterns hold information valuable to game development, which provides another argument supporting an explorative approach

In this study, the focus is not on virtual identities as a whole, but a specific element of them: their names. Even more specifically, the research presented here only deals with names from gamer tags and player characters from digital games, not e.g. social online media or online forums in general.

There are a number of reasons – or motivations - why investigating virtual game identities are of interest and potentially useful in the applied science domain (i.e. *in addition* to any benefit derived from simply adding to the current knowledge about games), including:

- 1) Investigating the names chosen by players in online (or offline) games – their online identities - is of direct value because it informs about the people who play these games - whether as objects of research inquiry or for target group analysis in the context of game marketing.
- 2) Examining naming patterns provides information for the design and development of games in terms of how names, appearance and other features of characters are perceived by the players, and thus e.g. guidelines for how to utilize these features in the design of non-player characters (artificial agents) to evoke particular associations or reactions in/from the players – from NPCs in MMORPGs to bots in multi-player shooters.
- 3) Investigating names for the purpose of player (user) profiling is to the best knowledge of the authors a novel approach, and it is not known if the names chosen by players for their online identities in games can be used to draw inferences about their in-game behavior [e.g. 15]. If this is the case, name-mining may prove valuable to player profiling, irrespective of the specific purpose (research, community management, cybersecurity, design evaluation etc.).

The user-generated identities in all four game contexts identify the players in social interactions with each other. The two kinds of identifiers in the games are different – one being related to a persistent avatar, the other being related to an account for a player. Both serve the same overall purpose, i.e. the identifier used by players for interacting with each other. However, no claim is being made here that these two types of identities are identical or that results from an analysis of one type can be related to the other. Rather, the fundamental rules for name generation and uniqueness in the two classes of games necessitated the the development of different sets of data mining methodologies in order to locate the patterns that do exist in the datasets investigated. These

methodologies are presented here and form the first contribution of the study. The analysis presented tells a tale of two highly different “communities of names”, in the MMORPG *World of Warcraft* (WoW) vs. three tactical shooters *Crysis 2* (Crysis 2), *Medal of Honor* (MoH) and *Battlefield Bad Company 2* (BFBC2), with marked differences in the names adopted by the players

In WoW, names need to be unique only for individual instances of the game world, and yet the evidence presented points to a staggering variety in the names chosen by players of WoW – 3.8 million unique names out of 7.93 million – notably remarkable given the restrictions imposed on character name generation in the game. The result highlights the imaginativeness employed in naming characters, indirectly supporting earlier work such as [6, 19] in concluding that the choices made during the character creation process are important to the players. Character names in WoW follow the same kind of log-distribution as real-world names - despite MMOGs being only about 1.5 decades old and restrictive in terms of name choices.

Previous research on WoW does indicate that character name, appearance and functionality are somehow linked (e.g. [9]), however, previous research is limited in terms of sample size and adopts informal methodologies which prevents inference from these studies towards parts of or the entirety of, the population of players. The results presented here show that (for the games included in the analysis) the appearance of a character and its gameplay functionality is related to the names given to characters - to a degree where it may be possible to develop predictive models which can assist game developers to evaluate player populations based on character names and similar sparse data (notably in situations where rich datasets are not available [45]). Furthermore, the names given to characters on Role-Playing, RP, realms show differences from those of Player-vs-Player, PvP, and Player-vs-Environment, PvE, realms. Finally, as reported by Thureau and Drachen [32], who investigated the sources of inspiration for character names of 120,000 player character names, the sources of inspiration for character names are diverse (38 categories developed via explorative coding), and that names with a negative semantic meaning (e.g. “Nightmare”) are more than six times as common as those with a positive meaning (e.g. “Hope”).

For the three shooters in the current study, gamer tags must be unique for telemetry tracking to function, and therefore all are unique. While the data collected for these games form a fraction of the WoW dataset (86,005 total), they comprise – to the knowledge of the authors - a larger sample size than any previous study on gamer tags. The results of clustering analysis on these data indicate that gamer tags, based on string distance measures, can be clustered and “archetypical” names for the different clusters located, indicating that there is structure in the distribution of gamer tags. Furthermore, that the behavior of the player and the gamer tag chosen is related to some degree, with purity measures for clustering according to behavioral profiles reaching as high as 61%. Finally, gamer tags did not cluster according to the three games, indicating that for the three shooters at least, the choice of gamer tag is not determined by the game, although some non-randomness was apparent in the analysis (purity 34%).

## 2 Related Work

While gamer tags and character names are mentioned in numerous research publications focused on games, as well as game development books and conference presentation, there are very few of these that investigate these in any detail. The exceptions it has been possible to locate predominantly originate in qualitative research and use comparatively small sample sizes; which limit the explanatory strength of the results when viewed from a quantitative, statistical perspective [e.g. 9,19,20].

**Game analytics and behavioral profiling:** Where patterns in naming characters and tags, and any relationships between these and player behavior, and their potential value in user profiling, have not formed the focus of previous large-scale research, player behavior in its own right is a key area of focus in the domain of game data mining, not only for profiling purposes but for a variety of game development and game maintenance purposes [45]. Game data mining has in the previous few years emerged as a key component of game development, and has notably gained acceptance in the digital entertainment industry and the associated research domains with the shift from offline to online gaming [39,45]. Game development thus follows the general “big data” trend, taking advantage of notably telemetry tools to capture detailed player behavior [45]. The tracking and analysis of player behavior has mainly focused on either investigating the players themselves and their social interaction patterns [e.g. 13, 14], informing game development [15] or for the development and testing of data mining algorithms for complex datasets [16].

User behavior in the context of digital games generally collected in the form of telemetry, i.e. quantitative data describing player-game and player-player interaction, and is generally compiled in databases from logs provided by game clients, or alternatively by extracting data directly from game servers [15, 14, 16]. Any action initiated by either a user or the game software itself can be recorded, at any level of precision from low-level data such as button presses and positions within a virtual environment, in-game conversations etc.; to higher level aggregate information such as total number of times a player has died, total number of completed game levels, etc. In games hosted by third-party online distribution or hosting platforms, such as *Facebook* or *Steam*, additional information on users can be available (e.g. demographics, user point of entry, data on virality) as an additional source of information feeding into analyses such as classification and segmentation.

Building profiles or categories of players based on the playing and purchasing patterns forms a key aspect of online game data mining, whether formed by classification, funnel analysis, clustering, classification or other methods [15,39]. This because being able to evaluate how player progress through a game, convert from non-paying to paying users, etc., is generally important to evaluate whether a design functions the way it is intended, and for online games often as a method for optimizing revenue.

Outside the games domain, substantial research has been focused on virtual identities and user profiles in online environments, notably the case of Internet-based web communities [e.g. 11,26]. Fields such as Information Science, Network Science and Cyberpsychology have seen investigation of large-scale data mining of user profiles, e.g. performed on data from Facebook, Twitter or Google searches (the former and latter by the companies involved) [51,52]. Large-scale work has focused on building profiles of users and to a lesser extent map the dynamics of user behavior, and triangulating online behavior/profiles with offline (real-world) components of the user, e.g. personality.

**Other research:** Outside of game data mining, behavior-driven profiling and game development, player character names and gamer tags have formed the subject of a limited amount of investigation. These investigations generally relate to Massively Multi-Player Online Games (MMOGs) such as *WoW*, a game form that over the past decade has attracted the attention of notably sociologists, ethnographers, design researchers and researchers working with communication and behavioral economy [e.g. 2, 12, 5, 6, 3]. In this body of research, the relationship between players and characters in digital games has formed one of the lines of investigation [2, 17, 18], and the topic is also covered in a number of game development books [e.g. 27] and conference presentations [e.g. 28]. For player character names, Guitton [19] examined the

cross-modal compensation between name and visual aspect in WoW player characters, based on a non-random sample of 1261 names. The work of Guitton [19] was however limited to consider whether name properties were affected by the visual aspect (human vs. non-human differentiation only) of the character. One conclusion is that female names contained more vowels than male names. Tronstad [20] investigated the relationship between the "capacity" of WoW player characters (their skills and abilities in terms of game mechanics) and their appearance. The study is qualitative, with no sample size specified. The methodology leading to the conclusions made in the paper are not described, however, Tronstad concludes that capacity and appearance is connected and provides a means of player identification. Finally, Hagström [9] collected a non-random sample of 1366 names of WoW characters, including race, class, gender and level information at the time of observation, from one WoW server at one point in time. The work presented by Hagström [9] is based on observations, and the author does not write how the names were analyzed nor any statistical properties of the dataset. He concludes e.g. that there was no observable difference between low-level and high-level characters in terms of naming patterns. Hagström [9] posits the hypothesis that player character names between RP-servers and other server types would differ. This is also a topic addressed in the current study.

### 3 Datasets and pre-processing

The work presented here is based on several datasets from four games (*World of Warcraft*, *Battlefield 2: Bad Company 2*, *Medal of Honor* and *Crysis 2*).

#### 3.1 *Battlefield 2: Bad Company 2*, *Crysis 2* and *Medal of Honor*: Gameplay

BFBC2, Crysis 2 and MoH are all examples of the tactical shooter type, generally played in First-Person or Third-Person view. While all three are shooters there are marked differences in their design and gameplay.

***Crysis 2*** (*Crysis 2*) is a FPS (First Person Shooter) (Crytek, 2011), released across PC, Xbox 360 and PlayStation 3. It is the sequel to the 2007 game *Crysis*, and its expansion *Crysis Warhead*. The game is primarily played in single-player mode but has a multi-player component as well. In *Crysis 2*, the player assumes the role of a Force Recon Marine named *Alcatraz*, who is equipped with a "nanosuit" that provides specific extra-ordinary abilities. The game provides freedom to customize weaponry and some character abilities, and to some extent non-linear gameplay through the game's missions. The game is set in a post-alien invasion (a tentacle, squid-type alien race called the "Ceph") New York in 2023, in what has been labeled an "urban jungle" setting/atmosphere.

***Medal of Honor*** (MoH) is a series of FPS games, with currently 14 titles in the series. The first was released in 1999 (Dreamworks Interactive), the latest in 2012 (Danger Close), entitled *Medal of Honor: Warfighter*. Data used here originate from the 13<sup>th</sup> game in the series, released 2010 for Xbox 360 and PlayStation 3. This is the first of the *Medal of Honor* games to be set in a modern setting rather than World War II. The newer games in the MoH series emphasizes front-line combat and relatively open-ended mission maps. The game has a single player component but is notably known for its multi-player component, similar to the *Battlefield* series, which features teamplay, class-based gameplay and experience-driven development during gameplay.

***Battlefield 2: Bad Company 2*** (BFBC2) is a tactical battlefield shooter with a first-person shooter view (EA Dice). It was released in 2010 for PC, Xbox360 and PlayStation 3. The game puts the player in a fictional war scenario between the United States of America and the Russian Federation. The game has a single-player, campaign mode, and a multi-player mode supporting up to 24

concurrent players (32 on PC), the latter being by far the most popular version of the game, similar to the latter incarnations of the MoH series. In the multi-player mode, each player takes control of one character, selecting from different classes (or “kits”) every time a new game is started, providing different starting equipment. E.g. the Demolition class is equipped with anti-vehicle weapons and land mines. Gameplay is team-based, with different types of scenarios. Similar to MoH, players can earn ranks, awards and special equipment over the course of their multiplayer career.

### 3.2 *Battlefield 2: Bad Company 2, Crysis 2 and Medal of Honor: Dataset and pre-processing*

The dataset for BFBC2, Crysis 2 and MoH was scraped from the P-Stats Network (<http://p-stats.com/>), a service which collects telemetry data from individual game clients, aggregates the data and makes them accessible to the players. The network provides an API for fetching behavioral telemetry from these and other games from the installed client, or via the server hosting the game (as is e.g. the case for BFBC2 running in multi-player mode). A variety of aggregated behavioral features are tracked and stored under each player’s profile. For each game, each player has an account showing their in-game gamer tag. This name is used for all instances of playing a specific game, and permits post-session tracking of telemetry across different classes.

The scraping process was performed as follows: The scraper queries the P-Stats Network website and mines the returned information, initially getting player names from the first 50 profiles that updated their stats last on the website. It scrapes information from these first, then capture information from their sub-page profiles. During pre-processing all gamer tags were made lower case, but otherwise all special characters and numbers were retained. A 15 or 16 character max. length is allowed for both of these games depending on the hardware platform.

The dataset for Crysis 2 and MoH consist of 3,000 player profiles that were sampled without replacement according to discrete uniform random distribution from each of the two games. These samples were in turn randomly selected from larger datasets scraped from the P-Stats Network (Crysis 2: 4,364 players; Medal of Honor: 12,328). It is important to that note even though P-Stats Network allows the players to see their complete profile from the web, not every player may choose to use this service (by installing the ad-on required) to check their statistics. Without a way to access the behavioral data from all players of the four games, drawing inference from the sample used here to the population of players beyond the P-stats network would require the assumption that P-stats players are representative of the total player population. The dataset used for BF2BC2 here includes 10,000 players, randomly sampled (in a similar manner as above) from a larger dataset of 69,313 players. Similar to the gamer tags for Crysis 2 and MoH, during pre-processing all gamer tags were made lower case, but otherwise all special characters and numbers were retained. A 16 character max. length is allowed for BF2BC2. Apart from gamer tags, 11 behavioral features were extracted. In choosing the behavioral features to include in the analysis, the method proposed by Drachen et al. [1] was followed, i.e. selecting features that allow for evaluating of the most important gameplay mechanics in the game under evaluation. In the current case, the features selected relate to character performance (score, skill level, accuracy etc.) and game feature use (kit stats, vehicle use), and playtime.

### 3.3 *World of Warcraft*

*World of Warcraft* is a MMORPG based in the *Warcraft* fictional universe, and has dominated the MMOG-market since its release in 2004, with a 12 million peak number of active players according

1 to the developer, Blizzard Entertainment<sup>1</sup>. The gameplay is character-centric, played in Third-  
2 Person view, where the player controls one character at a time. The game emphasizes combat,  
3 questing, exploration, equipment collection/construction and character development typical of  
4 fantasy-themed RPGs. Each player character is created based on an initial choice between 13 races  
5 and 11 classes (following the *Mists of Pandaren* expansion, released 2012). Races and classes have  
6 unique abilities and skills, and creating a character is a process of making a set of foundational  
7 gameplay choices, e.g. race (e.g. Human, Orc) impacts on the appearance of the character, and class  
8 (e.g. Shaman, Mage), impacts on the abilities of the character (e.g. a Warrior is a close-combat  
9 specialist).

10 The creation of characters is a source of considerable attention from the players [6]. Apart from  
11 character development, the game poses a variety of challenges, including teamwork assignments, as  
12 well as an in-game economic system [7]. The player will at any one time control one character, but  
13 can have multiple characters. Rather than interacting within the same virtual environment, players  
14 are distributed among hundreds of copies of the game (also called realms or shards) [8].

15 Player characters in WoW (and other MMORPGs), generally have a degree of persistence, which  
16 can cover years of real-time [2, 3]. In WoW, the characters the players control have names that are  
17 unique within the context of a specific realm, each with an approximate maximal load of 10,000  
18 active characters.

### 19 20 **3.4 World of Warcraft: Dataset and pre-processing**

21 The dataset for World of Warcraft consists of all character names from the 50 largest guilds on each  
22 of the EU and US WoW realms, as per the WoW census site ([www.warcraftrealms.com](http://www.warcraftrealms.com)), as of  
23 2011. This comprises a total of 7,938,335 character names. Due to being selected from guild-active  
24 WoW players rather than completely randomly from the population of WoW players, there may be  
25 a bias in the data, however, this is not possible to evaluate given that only Blizzard Entertainment  
26 has access to the entire population of their players. Strictly speaking, this means that inference can  
27 only be drawn from the sample to players in large guilds. However, it should be noted that the  
28 sample size represent an appreciable fraction of all WoW characters.

29  
30 Of these, 4,559,746 character names belong to US player characters and 3,378,589 to EU player  
31 characters. The amount of character names is distributed almost equally across Player-vs-Player  
32 (PVP) and Player-vs-Environment (or Engine) (PVE) realm types. In comparison, the Role-Playing  
33 (RP) realms make up less than 10% of the total number of character names in the dataset. Detailed  
34 numbers on the dataset can be seen in Table 1). During pre-processing, any special characters (e.g.  
35 the acute accent: "´") used in the names were removed in order to be able to compare e.g. instances  
36 of "Gandalf" and "Gändalf".

37  
38 Following, histograms of name occurrences for each realm or realm-type was computed. In total, 35  
39 histograms were computed, based on selections in the dataset. These selections correspond to realm  
40 types (All realms, US realms, EU realms, PVP realms, US-PVE realms, . . .); the various classes in  
41 the game (Priests, Druids, . . .) and races (Trolls, Orcs, Humans, Undead, . . .). The total number of  
42 character names in each category varies, from a few hundred thousand to approx. 8 million for the  
43 "All realms" category. Dependent on the number of characters, the number of unique character  
44 names follows a logarithmic frequency distribution, i.e. a function defining that increasing the

---

1 According to the developer, Blizzard Entertainment: <http://eu.blizzard.com/en-gb/company/press/pressreleases.html?101007>



number of names in the sample provides diminishing returns with respect to increasing the number of unique names [Fig 3(b)].

## TABLE 1

## FIGURE 1

### 4. Analysis and results

In this section the analysis process for all four datasets is described, results outlined and discussed.

#### 4.1 Character Name Frequencies in *World of Warcraft*

The WoW data were drawn from approximately 300 game servers, which correspond to the same number of instances of the virtual game world. Player character names are only required to be unique within each instance, and this enables analysis of frequencies of occurrence in a similar fashion to real-life family names. These follow power law distributions in many countries [10,33,34]. Informally, this means that the second most popular name appears half as often as the most popular, the third half as often again, and so forth. Power laws are regularly found among quantities in physical, biological, and social systems, where they can also be referred to as Zipf's or Benford's law (see e.g. [22]). Popular examples include the number of citations of scientific papers, word frequencies in books, or the populations of cities. Formally, a quantity  $x$  obeys a power law if it is drawn from a probability distribution  $p(x) \propto x^{-\alpha}$ , where  $\alpha$  is the scaling parameter or exponent [21]. As a thorough review of power law distributions is beyond the scope of this work, Newman [23] is recommended for a general introduction and explanation of power laws.

Assuming a power law distribution also exists for the WoW player character names relies on two preconditions where player character names differentiate from family names: 1) A character name might only appear once per realm. Players sometimes use special characters (e.g. German umlaut) to get around this limitation, however it still influences the possible choices for a name; 2) Unlike family names or other real-life names, the choice of a name is unrestricted, only certain names that violate the *Blizzard Naming Policy* are forbidden (although there are numerous examples of character names not adhering to these restrictions [32]).

In order to validate the power law distribution assumption, a linear function was fitted to the logarithmically scaled rank-frequency plots of the 1000 most frequently occurring names in the dataset. Examples of the rank-frequency plots are provided in Figure 2, and the fitted linear function in Figure 1. The power law assumption holds for all 35 separate data selections described above.

The overall worst

accuracy for the linear predictor is achieved for the "All realms" category, which however achieved a standard error  $\sigma$  of 0.0025 (see Fig. 1(a)), with  $\sigma = \sqrt{\sum_i^n (x^i - x'^i)^2 / n}$ , where  $x'$  denotes the predicted value and  $n$  the number of considered names. Note that  $\sigma$  is computed for logarithmically scaled values. Further, note that while fitting a linear function is not the most accurate way of verifying a power law assumption [21], given the high accuracy of the fit, this issue is negligible).

The analysis shows that all character name frequencies in WoW follow a power law distribution, irrespective of how they are divided (class, race, realm type). With an increasing total number of character names, e.g. for the "PvP" realms category, it shows that the standard error  $\sigma$  is increasing. However, the approximation accuracy remains high. A possible reason for the increased

1  $\sigma$  might be the constraints imposed on character names (e.g. the *one identical name per realm*  
2 constraint).

3 Furthermore, there appears to be variety in how many times the most popular names occur  
4 as a function of race or class. For example, the characters of the mage class exhibit a higher variety  
5 of names than the other classes. For the races, the name "*Hellscream*" is the most popular among  
6 the Orc characters (almost 80 counts), but for the other races the most popular name is more  
7 frequent. For example, "*Neytiri*" is the most popular Draenai name (over 130 counts) and "*Arthas*"  
8 the most popular name for humans (over 180 counts). Why these patterns occur the present analysis  
9 cannot provide an answer to, but they indicate that the distribution of names is not random but  
10 highly affected by the environment they are used in relation to.

## 11 **FIGURE 2**

### 13 **4.2 Character name distributions across categories**

14 For further comparison of the player character name distributions across the three categories (class,  
15 race, realm type), the histogram L1-distances were embedded into a 2D space by means of Isomap  
16 projection [24] (Fig. 4(a)). The resulting projection shows categories (different realm-types or  
17 classes) that are close together also have a similar character name distributions. Conversely,  
18 categories that are distanced from each other have comparatively smaller overlaps in character  
19 names and name frequencies. Two findings can be derived from this:

20  
21 1) Names of US and EU realms have a large distance, which means that names given to player  
22 characters by US and EU players varies. This pattern is however not evident for RP realms, where  
23 the isomap projection indicates a larger overlap in character names between US and EU realms  
24 (Fig. 4(a)).

25  
26 2) There is a clear separation in player character names given to the human-like and human-  
27 proportioned races (Human, Blood Elf, Night Elf, Draenei) and the less human-like (Troll, Undead,  
28 Orc, Tauren, Dwarf, Gnome) races (note that expansion packs released following the data collection  
29 process has added additional races to the game). This pattern might have been an effect of an  
30 uneven distribution of male and female characters, and character names for certain races. However,  
31 it was found that for the less human-like races, there is a tendency towards male characters, whereas  
32 for the more human-like races a higher proportion of female characters were observed. However, in  
33 analysis of male and female characters separately showed the same pattern (results virtually  
34 identical) as for the complete dataset, and therefore character gender differences is not the cause of  
35 the patterns observed in the data.

## 36 **FIGURE 3**

37  
38  
39 In order to further investigate the separation in player character names according to race, class and  
40 realm type, a simple predictive model for character names was evaluated.

41  
42 The idea is to estimate conditional probabilities of a particular class, race, or server type given a  
43 character name, e.g. how likely is it that a character named "*Gimli*" is a "*dwarf*"? Probabilities were  
44 estimated for the 1000 most popular names in the dataset (roughly 120,000 characters).  
45 Probabilities were normalized by class/race/server-type priors so that seldom selected classes are

not punished in the analysis, etc. In particular, the uniqueness of a specific name with respect to a particular class or race is of interest within the context of player profiling as a goal, as this indicates which variables that influence player naming decisions.

Some results are summarized in Figure 4: The most influential features appear to be the player class, player race (which influences human-likeness), however, the server type and faction also give hints on player naming decisions. These relationships indicate that there are specific groups of names that players associate with specific types of characters. This was further investigated by Thureau and Drachen [32], who examined the sources of inspiration for the 1000 most popular names in the WoW dataset, finding a great variety of sources, including popular media, literature and mythology. From the perspective of game development, data mining of player character names could be useful in designing NPCs to evoke specific associations with the users.

## FIGURE 4

### 4.3 Pattern analysis for Crysis 2, MoH and BFBC2

Whereas the names of WoW player characters generally have a semantic structure and meaning [32, Figure 2], the names adopted by players of the three shooters under investigation commonly form combinations of words, letters, numbers and special characters. Examples from BFBC2 include: “*MaliciousMaulr*”, “*x6naca6x*”, “*Ankur*”, “*Daniel08*”, “*Hafling86*”, “*HackJake0025*”, “*InSaNe\_x\_ChAoZz*”, “*Craybell*”, “*ToFu651*” and “*Acid\_Snake*”. Unlike the character names in WoW, the gamer tags in the three shooter games are not dedicated to a specific character, but rather to the player’s account. Irrespective of the choice of kit/class in BFBC2 or MoH, the name displayed to the other players is the same. This means that the gamer tag is associated directly with the player, rather than a fictional character. Furthermore, within each of the three games, gamer tags must be unique, which means that the frequency-based approach towards locating patterns in naming behavior adopted for WoW above cannot be applied to these games.

As gamer tags have not – to the best knowledge of the authors – been the subject of large scale analysis before, the central purpose of the current study was explorative; to investigate if there were any kinds of patterns in the gamer tags – within specific games, across games and if there are relationships between gamer tag and player behavior (e.g. for potential application to user profiling). Towards this goal, three topics were investigated: a) If there was an underlying structure in the gamer tags, b) If gamer tags were related to in-game behavior, c) If gamer tags were related to the individual game.

## TABLE 2

### Gamer tag clustering

Initially, the 10,000 player sample from BFBC2 was subjected to cluster analysis in order to locate any underlying structure in the gamer tags.

When clustering strings, string similarity metrics are used to determine the difference between strings by checking different features of their arguments. There are numerous methods available for defining similarities between strings, including the Jaccard Similarity Coefficient, the Jaro-Winkler Distance [R3] and the Levenshtein Distance, which is a commonly applied similarity measure for string clustering, and was adopted for the current study [29]. Each similarity measure has its own way of defining similarity between strings, with associated pros and cons. Levenshtein Distance

[29] is a string distance measure that receives two strings and returns a numerical value (distance) depending on how different the strings are in terms of three edit operations: insertion of a new character, deletion of a character and replacement of a character; i.e., given two strings  $s1$  and  $s2$ , Levenshtein Distance finds the minimum cost for transforming  $s1$  to  $s2$  using the edit operations. The standard setting of the distance suggests that every such operation increases the distance by the value 1. However, this can be changed by weighting the edit operations by different numbers [30]. Given two strings the Levenshtein Distance can be calculated using a dynamic-programming algorithm, described in Manning et al. [30].

As the datasets are based on nominal attributes, the k-means variant approach k-Medoids clustering was used in combination with the Levenshtein Distance measure. K-Medoids is an object-based clustering technique, i.e. objects are used as centroids rather than hypothetical centroids as for the regular k-means approach [46]. Clustering was performed across a variety of cluster (k) numbers (max. 1000).

The larger the number of defined clusters, the more the smaller clusters break down, but the general pattern remains largely fixed (Figure 5). The resulting clusters range in size, but do not appear to follow a power law distribution as the WoW player character name frequencies does. This was also the case when the Jaccard Similarity Coefficient and the Jaro-Winkler Distance was applied to the dataset [30]. While, as outlined above, real-life names follow a power law distribution, and character names in WoW as well, the lack of a power law distribution in the BFBC2 gamer tags is possibly an artifact of the algorithms of the distance measures applied. The analysis indicates that the use of numbers and special characters held relatively minimal influence on the largest 100 clusters, indicating that there were few strings of numbers or number/word combinations that were *commonly* used in the gamer tags. Similar to classes in WoW, cluster centroid strings for the four different kits in BFBC2 were different for the largest clusters (Figure 6).

**FIGURE 5**

**FIGURE 6**

### **Gamer tags and in-game behavior**

Two different analyses were performed to investigate if there were any relationships between player behavior and gamer tags in BFBC2, as follows:

BFBC2 provides players with a choice of four classes (Assault, Engineer, Medic, Recon), or “kits” every time they start a new mission (for descriptions see Drachen et al. [36]). These classes provide different starting equipment and thus are optimal for different purposes, which in turn impacts in-game behavior. For example, the Engineer kit provides equipment useful for handling vehicles. Given the impact of class choice in WoW on player character names, investigating relationships between kits and gamer tags form an obvious first step.

The behavioral features in the BFBC2 dataset include scores players had obtained from each of the four kits (as well as number of times player was killed in that role, or was killed, and the kill/death ratio). For each player in the dataset, the kit with the highest score was selected, as an aggregate indicator of in-game playing behavior (i.e. kit preference). K-Medoid clustering using Levenshtein Distance was performed with  $k=4$  (four kits = four classes assigned), and the purity value [30] for each cluster calculated.

Purity is a measure to quantify the quality of clustering with respect to the classes assigned to data samples. Given a labeled dataset and a clustering solution produced from this dataset; purity informs how well the clusters are fitted to the defined classes[30].

$$purity(Y, C) = \frac{1}{N} \sum_i^K \max_p |c_i \cap y_p|$$

The equation formally defines the purity measure, where  $Y = \{y_j | j = 1, \dots, J\}$  is the set of classes,  $C = \{c_i | i = 1, \dots, K\}$  is the set of clusters,  $N$  is the size of the dataset and  $|c_i \cap y_p|$  is the cardinality of the intersection between the  $i$ th cluster and the  $p$ th class. The value ranges between 0 and 1, with 1 being complete overlap. Purity is a simple and transparent evaluation measure [30], which has been used in similar contexts [e.g. 35]. It is a robust measure, but is inclined towards cluster centers and thus sensitive to cluster size [30]. Alternative measures, such as the F-measure, V-measure, Entropy etc. exist and could be applied to the current dataset, but evaluation of the fitness of all of these is out of scope of this paper, and the fitness of these algorithms remains a topic of discussion [e.g. 47,48].

The result is a 37% purity between the gamer tags and kits, indicating that some non-randomness exist, i.e. that gamer tags to a limited degree varied as a function of most favored kit (i.e. kit with highest score), but the result also indicates that the precision is relatively low. In essence, the favorite kit of a player does not indicate gamer tag, and vice-versa.

As described above, the dataset for BFBC2 also included 11 measures of player behavior, related to the core mechanics of the game. Based on these 11 features, clusters of archetypical behavior were built using SIVM, a modified version of Archetype Analysis (AA) algorithm but applicable to large datasets [37,38]. The process is described in detail in Drachen et al. [36]. The end result was seven behavioral profiles within which players could be allocated, and corresponding profiles described. Examples include the “Assassin” profile, with very high kill/death ratios, high kill per minute scores, low playtime and low numbers of deaths per minute (i.e. players who are adept at killing other players without getting killed themselves). Conversely, the “Target Dummies” feature extremely low kill/death ratios, lowest skill, and low kill per minute rates.

Whereas BFBC2 kit choices form a single, indirect measure of player behavior, the SIVM-based behavioral profiles developed by Drachen et al. [36] are more detailed. These were used as the basis for a new analysis ( $k=7$ ). Purity for the SIVM clusters is 61%, indicating a stronger degree of correlation between the gamer tags and the behavioral profile of the BFBC2 players, as compared to the kits-based analysis. I.e. the detailed behavior of the players correlates better with choice of name than a high-level gameplay choice like kit selection (which still permit a variety of different playstyles). For BFBC2 at least, playstyle is a better indicator of tag choice than kit selection. Compared to WoW, where class held a greater influence, the difference is possibly related to class choice in WoW being a permanent choice for a character, which directly impacts across most aspects of gameplay. In comparison, kit choice in BFBC2 is a temporary choice with no long-term impact.

Finally, while it was observed in the WoW dataset that there is at least some relationship between name and class/race choice, it is not possible from the current analysis to evaluate if this relationship is statistically stronger or weaker than the relationship between BFBC2 behavior and

gamer tag. The methods adopted for the two cases are purposefully different, in order to evaluate different strategies towards data mining names/gamer tags in digital games. That being said, the analysis clearly indicates that in-game behavior relates to choice of name/tag, at least for a proportion of the involved words/terms. Future research could examine in more detail the relationship between the terminology used in names/tags and detailed behavioral records across many games, to evaluate if this is a universal pattern in digital games. Such research could even be extended to integrate e.g. personality profiling of the players, as discussed by Canossa [54] and Spronck et al. [53].

### **Gamer tags across BFBC2, MoH and Crysis 2**

The final test was performed on 3,000 player samples randomly selected from the parent datasets, for each of the three shooter games. The goal here was to investigate if gamer tags varied as a function of the game they were associated with. The purity measure was 34%, indicating that there was some difference in the gamer tags across the three games, but the precision is relatively low. In essence, the specific game is not a strong indicator of gamer tag, and vice-versa. A possible explanation is that the population of the players using the P-stats network client across these three games share similarities, however, further research is needed to evaluate this possibility.

## **5 Discussion and Conclusions**

There are a number of reasons for why investigating online identities is of interest, with examples including how technologies impact human behavior, cyberpsychology, cybersecurity, design evaluation, etc. [1,2,3,5]. Within the domain of digital games, at the fundamental level, investigations of game identities informs about the people who play games, and provides information for the design of games in terms of how names, appearance and other features of identities (whether avatar-based or not) are perceived by players, and thus e.g. guidelines for how to utilize these features in game design (e.g. for NPC and bot profile generation). Central to the work presented here is online profiling of players via their in-game behavior and character names/gamer tags – as well as any other information available about the users, whether for commercial or research purposes.

The analyses presented here is an early step along the path of exploring whether there are any patterns in the names chosen by players for their characters or gamer tags, and their in-game behavior. The datasets presented cover four major commercial game titles: *World of Warcraft*, *Crysis 2*, *Medal of Honor* and *Battlefield 2: Bad Company 2*. The results across these four games all show that the player character features in WoW and in-game behavior in the three shooter games is related to a greater or lesser extent with the names chosen by the players for their characters or online profiles, whereas purity measures indicate that gamer tags are only related to the specific game to a limited degree. This may be related to all of these games being tactical shooter games. This indicates that character names and gamer tags are a useful addition to player profiling, perhaps especially in sparse data situations. Especially for WoW, the analyses presented above show that the appearance of a player character and its functionality within the confines of the game mechanics is a direct influence on name selection – this to a degree where it may be possible to develop predictive models which can assist game developers to evaluate player populations based on character names and similar sparse data.

## Acknowledgements

The authors would like to extend their warmest gratitude to Dr. Menno Van Zaanen, Dr. Peter Juel Henrichsen, Dr. Thea Drachen, Dr. Kristian Kersting and Dr. Christian Bauckhage for help with data access, discussions, advice and feedback.

## References

- [1] J. Bohannon, Game-Miners Grapple With Massive Data. *Science* 330(6000) (2010) 30–31.
- [2] C. Steinkuehler, D. Williams, Where Everybody Knows Your (Screen) Name: Online Games as Third Places. *Journal of Computer-Mediated Communication* 11(4) (2006) 885–909.
- [3] C. Pearce, A. Artemesia, *Communities of Play: Emergent Cultures in Multiplayer Games and Virtual Worlds*. The MIT Press, 2009.
- [4] M. Madden, A. Smith, J. Vitak, Digital footprints: online identity management and search in the age of transparency. Pew Research Center, 2007.
- [5] E. Castranova, *Exodus to the VirtualWorld: How Online Fun Is Changing Reality*. Palgrave Macmillan, 2007.
- [6] S. Caplan, D. Williams, N. Yee, Problematic internet use and psychosocial well-being among mmo players. *Computers in Human Behavior* 25(6) (2009) 1312–1319
- [7] J. Bohannon, Slaying monsters for science. *Science* 320(5883) (2008) 1592
- [8] H. G. Corneliussen and J. W. Rettberg, *Digital Culture, Play, and Identity*. The MIT Press, 2008.
- [9] C. Hagström, Playing with names: Gaming and naming in World of Warcraft, in: H. G. Corneliussen and J. W. Rettberg (Eds.), *Digital Culture, Play, and Identity*. The MIT Press (2008) 265–286
- [10] D. H. Zanette and S. C. Manrubia, Vertical transmission of culture and the distribution of family names. *Physica A: Statistical Mechanics and its Applications* 295(1-2) (2001) 1–8
- [11] D. Wiszniewski, R. Coyne, R., Mask and identity: The hermeneutics of self-construction in the information age, in: K. A. Renninger and W. Shumar, W. (Eds.), *Building Virtual Communities*, Cambridge University Press (2002) 191–214
- [12] W. S. Bainbridge, The scientific research potential of virtual worlds. *Science* 317(5837) (2007) 472–476
- [13] N. Ducheneaut, and R. J. Moore, 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* (2004)
- [14] D. Williams, M. Consalvo, S. Caplan and N., Looking for gender (lfg): Gender roles and behaviors among online gamers, *Journal of Communication* 59 (2009) 700–725
- [15] A. Drachen, A. Canossa, G. N. Yannakakis, Player modeling using self-organization in tomb raider: Underworld, in: *Proceedings of IEEE Computational Intelligence in Games* (2009) 1–8
- [16] C. Thureau, K. Kersting and C. Bauckhage, Convex non-negative matrix factorization in the wild, in: *Proceedings of the IEEE International Conference on Data Mining* (2009).
- [17] A. Tychsen, M. Hitchens and T. Brolund, Character play: the use of game characters in multiplayer role-playing games across platforms. *Computers in Entertainment* 6(2) (2008)
- [18] P. Vorderer and J. Bryant, *Playing Video Games*. Lawrence Erlbaum Associates (2006)
- [19] M. J. Guitton, Cross-modal compensation between name and visual aspect in socially active avatars. *Computers in Human Behavior* 26(6) (2010) 1772–1776



- [20] R. Tronstad, Character identification in world of warcraft: The relationship between capacity and appearance, in: H. G. Corneliusen and J. W. Rettberg (Eds.), *Digital Culture, Play, and Identity*. The MIT Press (2008) 249–264
- [21] A. Clauset, C. R. Shalizi, C.R. and M. E. J. Newman, Power-law distributions in empirical data, *SIAM Review* 51(4) (2009) 661–703
- [22] L. Pietronero, E. Tosatti and A. Vespignani, Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. *Physica A* 293(1-2) (2001) 297–304
- [23] M. E. J. Newman, Power laws, pareto distributions and Zipf’s law. *Contemporary Physics* 46(5) (2005) 323–351
- [24] J. B. Tenenbaum, V. de Silva and J. C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science* 290(5500) (2000) 2319–2323
- [25] G. Gabriel and M. Maher, Coding and modeling communication in architectural collaborative design. *Automation in Construction* 11 (2002) 199–211
- [26] T. Wright, E. Boria and P. Breidenbach. Creative player actions in FPS online video games. *Game Studies* 2(2) (2002).
- [27] R. Bartle. *Designing Virtual Worlds*. New Riders, Indianapolis, 2003.
- [28] H. Smith 2007. The imago effect: Avatar psychology. Presentation at the 2007 Game Developers Conference (San Francisco, California) (2007).
- [29] V. I. Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission*, 1 (1965), 8-17.
- [30] C. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- [31] W. E. Winkler. String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage. In *Proceedings of the Section on Survey Research*, US Bureau of the Census (Washington DC, USA) (1990), 354-359.
- [32] C. Thureau and A. Drachen. Naming Virtual Identities: Patterns and Inspirations for Character Names in World of Warcraft. In *Proceedings of the 10th Entertainment Computing Conference* (Vancouver, Canada) (2011).
- [33] A. Clauset, C. R. Shalizi, M. E. J. Newman. Power-Law Distributions in Empirical Data. *Society for Industrial and Applied Mathematics Review*, 51(4) (2009) 661-703.
- [34] Y. E. Maruvka, N M. Shnerb, D. A. Kessler. Universal features of surname distribution in a subsample of a growing population. *Journal of Theoretical Biology*, 262(2) (2010), 245-256.
- [35] S. Aranganayagi and K. Thangavel. Improved K-Modes for Categorical Clustering Using Weighted Dissimilarity Measure. *International Journal of Information and Mathematical Sciences*, 5(2) (2009), 182-188.
- [36] A. Drachen, R. Sifa, C. Bauckhage and C. Thureau. Guns, Swords and Data: Clustering of Player Behavior in Computer Games in the Wild. In *proceedings of IEEE Computational Intelligence in Games*, 2012 (Granada, Spain), IEEE Publishers, 163-170.
- [37] C. Thureau, K. Kersting, and C. Bauckhage. Convex Non-Negative Matrix Factorization in the Wild. In *Proceedings of the IEEE International Conference on Data Mining* (2009).
- [38] K. Kersting, M. Wahabzada, C. Thureau, and C. Bauckhage. Hierarchical Convex NMF for Clustering Massive Data, *Proceedings of ACML*, 2010.
- [39] T. Fields and B. Cotton. *Social Game Design: Monetization Methods and Mechanics*. Morgan Kauffman Publishers, 2011.
- [40] K. Orland, Ubisoft boss: free-to-play a natural reaction to high piracy rates. *Ars Technica*. August 22<sup>nd</sup> (2012). Available from: <http://arstechnica.com/gaming/2012/08/ubisoft-boss-free-to-play-a-natural-reaction-to-high-piracy-rates/?comments=1#comments-bar>

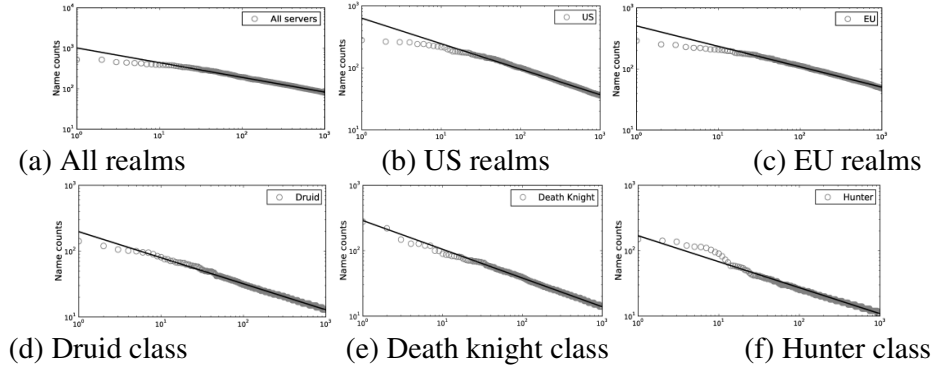


- [41] G. Zoeller. Game Development Telemetry in Production, in: M. S. El-Nasr, A. Drachen and A. Canossa. Game Analytics: Maximizing the Value of Player Data. Springer Publishers, 2013.
- [42] S. Santhosh and M. Vaden. Telemetry and Analytics Best Practices and Lessons Learned, in: M. S. El-Nasr, A. Drachen and A. Canossa. Game Analytics: Maximizing the Value of Player Data. Springer Publishers, 2013.
- [43] A. Drachen, M. S. El-Nasr and A. Canossa. Game Analytics: The Basics, in: M. S. El-Nasr, A. Drachen and A. Canossa. Game Analytics: Maximizing the Value of Player Data. Springer Publishers, 2013.
- [44] K. Bauer, D. Grundwald and D. Sicker, The Arms Race in P2P. In Proceedings of the 37th Research Conference on Communication, Information and Internet Policy (2009) 1-18.
- [45] M. S. El-Nasr, A. Drachen and A. Canossa. Game Analytics: Maximizing the Value of Player Data. Springer Publishers, 2013.
- [46] J. Han J. and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2006.
- [47] A. Rosenberg and J. Hirschberg. V-Measure: A conditional entropy-based external cluster evaluation measure. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (Prague, Czech Republic, 2007), 410–420.
- [48] Y. Zhao and G. Karypis. Criterion Functions for Document Clustering: Experiments and Analysis. Technical Report TR 01-40, Department of Computer Science, University of Minnesota (2001).
- [49] N. Ducheneaut, N. Yee, E. Nickell and R. J. Moore. 2006. Alone Together? Exploring the Social Dynamics of Massively Multiplayer Games. Proceedings of CHI 2006. New York: ACM. (2006) 407-416.
- [50] N. Ducheneaut, N. Yee, E. Nickell and R. J. Moore. 2006. Alone Together? Exploring the Social Dynamics of Massively Multiplayer Games. Proceedings of CHI 2006. New York: ACM. 92006) 407-416.
- [51] J. Blascovich and J. Bailenson. Infinite reality : avatars, eternal life, new worlds, and the dawn of the virtual revolution (1st ed.). New York: William Morrow. ISBN 0061809500.
- [52] R. Feizy, I. Wakeman and D. Chalmers. Distinguishing Fact and Fiction: Data Mining online Identities. In Proceedings of 5th International Workshop on Security and Trust Management (STM) (2009).
- [53] P. Spronck, I. Balemans and van Lankveld, G. Player Profiling with Fallout 3. In Proceedings of the 8<sup>th</sup> AAA Conference on Artificial Intelligence and Interactive Digital Entertainment (2012).
- [54] A. Canossa. Play-Persona: Modeling Player Behavior in Computer Games. PhD-thesis. Danish Design School, Copenhagen (2009).

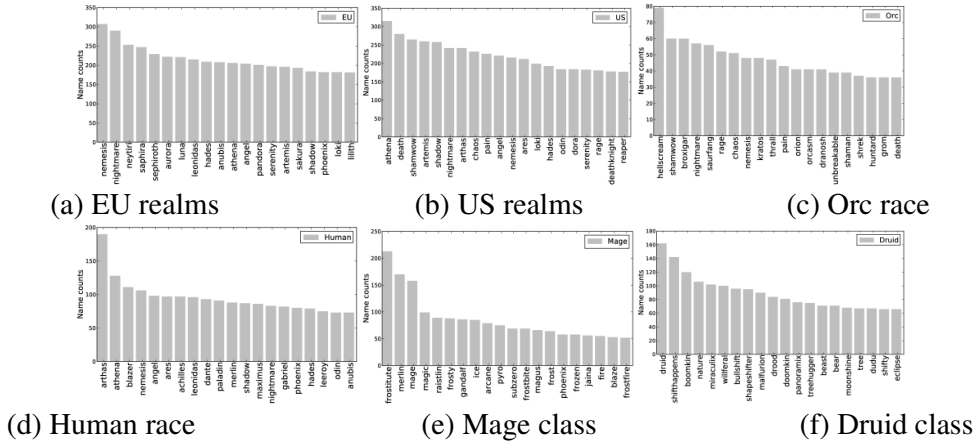
Location	(a) Location			(b) Realm type	
	# characters	# unique names	Realm type	# character	# unique names
All	7,938,335	3,803,819	PVP	3,128,464	1,869,481
Europe (EU)	3,378,589	1,820,269	PvE	3,884,205	2,207,478
United States	4,559,746	2,495,960	RP	619,892	499,481

(US)

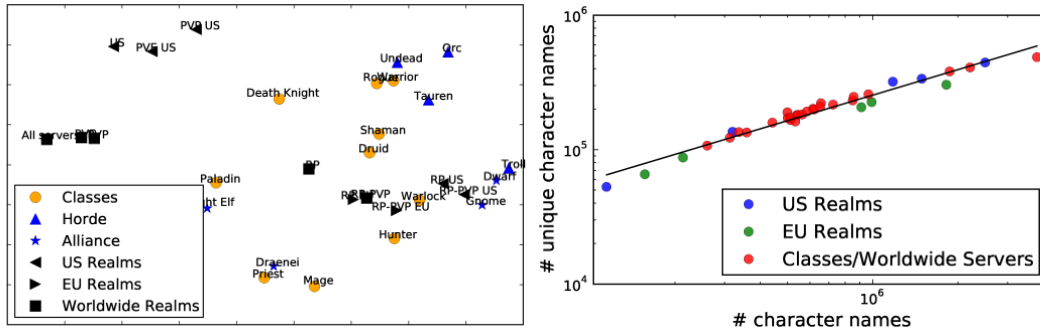
**Table 1:** Basic statistics of the character name dataset. We collected a total of 7,938,335 character names. The realm-types specify if the particular world is hostile (PVP), friendly (PVE), or targeted towards role playing (RP) which prohibits non-medieval character names. [Source: Thureau and Drachen, 2011]



**Fig. 1:** Rank-frequency plots for different selections of the character name dataset. Both axes are logarithmically scaled. The y-axis denotes the total count for a particular name, the x-axis denotes the rank index (the left most ranked name has the highest number of occurrences). [Source: Thureau and Drachen, 2011].

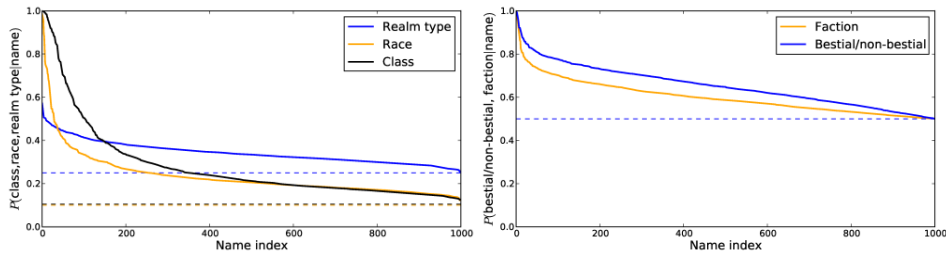


**Fig. 2:** Player character histograms for selected categories of characters (realm, class and race). Histograms show the total number of name occurrences on the y-axis and on the x-axis the corresponding name. The x-axis is sorted so the most frequently occurring name comes first (left side of the histograms). [Source: Thureau and Drachen, 2011].



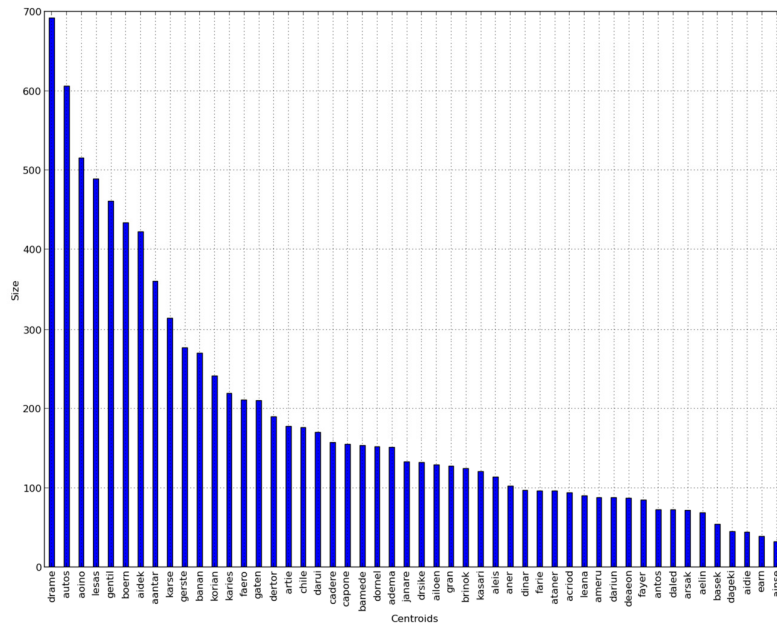
**Fig. 3:** Fig. 4(a, left) shows 2D-Isomap projection of histograms of character names for different categories. Categories which are closer together (e.g. "Undead" and "Orc" in the upper right part corner) have a larger overlap in character names. Fig. 4(b, right) shows the ratio of unique names and total number of names for different categories: the x-axis shows the number of unique names per category, the y-axis shows the number of players. Both axes have a log-scaling [Source: Thureau and Drachen, 2011]

[FOR COLOR REPRODUCTION ON WEB]

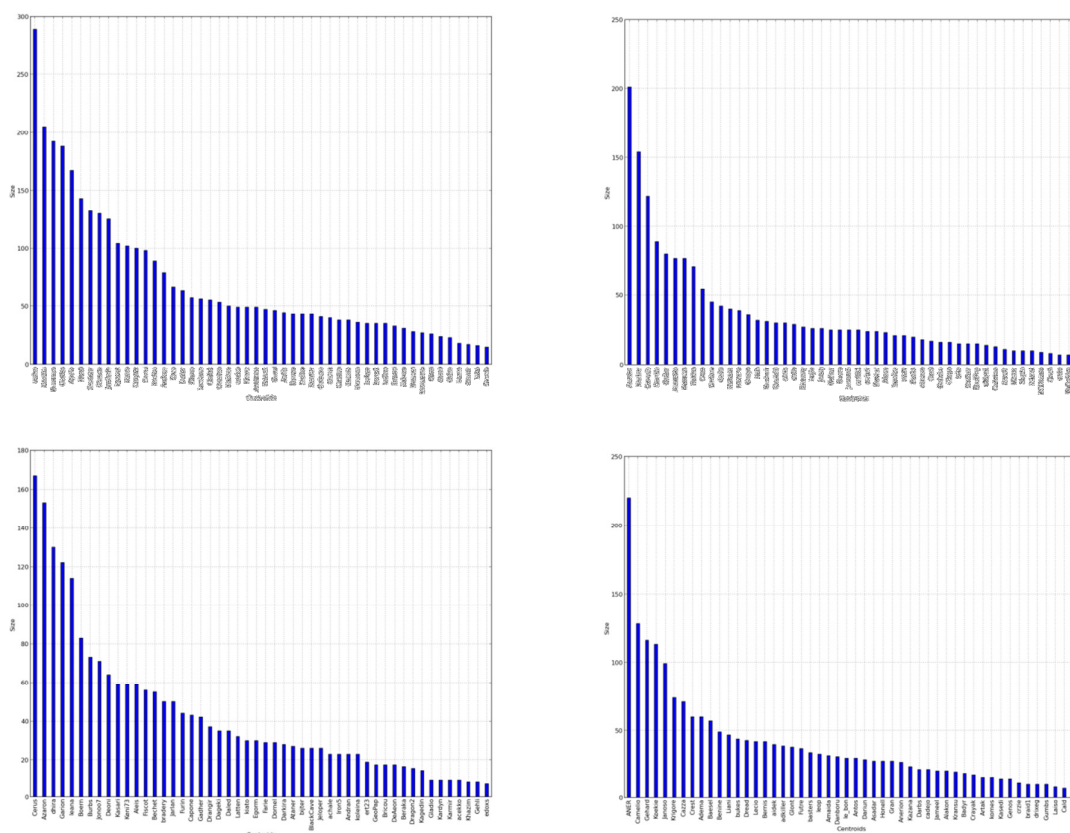


**Fig. 4:** Probabilities for a specific class, race, server type, faction (Horde or Alliance), and human-like/less human-like race given the character name. The probabilities are evaluated for the 1000 most popular names and the highest probability is selected. Afterwards, the names are sorted according to their highest probability. It can be seen that, for example, the player class (left figure) and faction (right figure) influence character names to a certain extent, i.e. some names are strong predictors of class/race/realm, others less strong. The dashed lines denote probabilities for random selection for the corresponding feature [Source: Thureau and Drachen, 2011].

[FOR COLOR REPRODUCTION ON WEB]



**Fig. 5:** *K-Medoids clustering on BFBC2 gamer tags (10k players) with 50 clusters. Strings along the X-axis are K-Medoids centroids (gamer tag strings central to each cluster) and Y-axis the cardinality of the cluster.*



**Fig. 6:** *K*-Medoids clustering on BFBC2 gamer tags, separated into four groups according to kit with the highest total score (“favorite kit”). 10,000 players total, with 50 clusters for each kit. Strings along the X-axis are *K*-Medoids centroids (gamer tag strings central to each cluster) and Y-axis the cardinality of the cluster.

Assault-Recon	Medic-Engineer	Assault-specialists	Driver-Engineer
JASON BOURNE BR	F(R)leND	Kriegerman	lleluja
Aggressivewar	GALKIN1976	Killer Dave_73	JimmyGER
EnsignDaluz	dogbolta	Eismann 43	Bird76[GER]
Gangstel	CaSpErTbH	Casios	bloodbath
LilShay	c4rmo	Cicawa	Imihunter
Gangstel	Kovvalsky	Capt J Hardnuts	DestroyerMarv
[007]kun	Cecil	Capn Crunch 9	knochenjager74
65Days__Epic	Kot Da Vinci	Destreure	Jarhead1988
Ganibal	Albatross777	APEDIGGER	FSB TERMINATOR
Da\$1PER	King Baldwin	Juan1972	IROC_87
Assasins	Veterans	Target Dummies	
Darkm0o0n	FrankieSuperRed	CaptainKirk	
.GOD..	LaughingMan	CarterCarter	

I ShOoT YoU Die	HansRS	fred-frog-man
5niperNo5niping	DarklyNoon	^i^AngelOfDeath
f0rge	George85	Happy-Camper
JOK3R_v2	ArnieF4440	Killerbraut666
Evil Popsicle	FirePin	HSH DeadShooter
Coookie	dr_musik	kiko213
DrunkenKnight	Agecanonixxxx	JaCeK2010.PL
CELTIC_GHOST	Kostya82	ICEMAN NOVOSIB

**Table 2:** The table lists 10 centroid names found by k-medoids from the seven behavioral BFBC2 clusters. Each list contains the 10 names that represent (or summarize) the common names occurring in the behavioral clusters according to Levenstein distance.