Looking for archetypes: Applying game data mining to hearthstone decks

Antonio M Mora, Alberto Tonda, Antonio J Fernández-Ares, Pablo García-Sánchez

Looking for Archetypes: Applying Game Data Mining to Hearthstone Decks

Antonio M. Mora[a], Alberto Tonda[b], Antonio J. Fernández-Ares[c], Pablo García-Sánchez[d]

[a] *Dept. of Signal Theory, Telematics and Communicat., University of Granada, Spain. amorag@ugr.es*
[b] *UMR 518 MIA-Pairs, INRAE, Université Paris-Saclay, Thiverval-Grignon, France. alberto.tonda@inrae.fr*
[c] *Dept. of Languages and Computer Systems, University of Granada, Spain. antares@ugr.es*
[d] *Dept. of Computer Architecture and Computer Technology, University of Granada, Spain. pablogarcia@ugr.es*

Abstract

Digital Collectible Cards Games such as Hearthstone have become a very prolific test-bed for Artificial Intelligence algorithms. The main researches have focused on the implementation of autonomous agents (bots) able to effectively play the game. However, this environment is also very attractive for the use of Data Mining (DM) and Machine Learning (ML) techniques, for analysing and extracting useful knowledge from game data. The objective of this work is to apply existing Game Mining techniques in order to study more than 600,000 real decks (groups of cards) created by players with many different skill levels. Data visualisation and analysis tools have been applied, namely, Graph representations and Clustering techniques. Then, an expert player has conducted a deep analysis of the results yielded by these methods, aiming to identify the use of standard - and well-known - archetypes defined by the players. The used methods will also make it possible for the expert to discover hidden relationships between cards that could lead to finding better combinations of them, enhancing players' decks or, otherwise, identify unbalanced cards that could lead to a disappointing game experience. Moreover, although this work is mostly focused on data analysis and visualization, the obtained results can be applied to improve Hearthstone Bots' behaviour, e.g. predicting opponent's actions after identifying a specific archetype in his/her deck.

Keywords: Video games, Hearthstone, Archetypes, Collectible Card Games, Artificial Intelligence, Game Data Mining, Data visualisation, Clustering Techniques

1. Introduction

In the field of AI research on video games, a large part of the effort is focused on the study of the intelligent agents that can play the game, while another considerable part of the literature is concerned with procedural content generation. However, a task of great interest for game designers and developers is to understand how players interact with the game to gain relevant knowledge that can help them take better decisions [1].

In complex games like Collectible Card Games (CCGs), such as *Magic: The Gathering*, it is particularly interesting to analyse how players interact with each other due to their dynamic, complex and changing nature. In such games, players collect cards with special rules and create their own decks to play against other players. Different Internet communities of players show and discuss the decks they have composed, whereas game developers create and launch new cards over time in the so-called *expansions*, so players increase their collections and improve their decks in order to keep up with other players.

Hearthstone, Heroes of Warcraft (HS) is one of the most popular Digital CCGs (DCCGs) nowadays, with over 100 million players. In addition, this game is starting to be seen as a testbed for different branches of AI. This is due to the large variety of different cards and rules, which together with the hidden information and randomness, allow for rich and varied combinations of effects, very challenging to predict [2].

Before starting a game in HS, the two players (called *Heroes*) facing each other have to prepare a deck of 30 cards from those available in their personal collection. This collection can be expanded by buying random card packs using in-game currency, earned upon completion of daily quests and other means, or real money. The objective of the game is to reduce the opponent hero's health to 0 using various types of cards by playing them, in turns, over a common board. The available types are: spells, minions, hero cards, and weapons. *Spells* are cards that produce an immediate effect and then they are discarded (e.g. reduce the enemy hero's health by 2 points); *Minions* are cards that are placed on the board and are used to attack heroes or other minions, having their own special characteristics and moves. *Hero cards*, when played, will replace the player's hero, providing them with a small amount of Armor and a different (usually more effective) Hero Power. Finally, *weapons* allow the Hero to attack the other characters for several turns. To play any card, players spend *mana crystals*. At the beginning of each turn, the number of crystals is increased by one unit (starting with 1, up to a maximum of 10) and recharged to be used again during that turn. Figure 1 shows a screen capture of the game.

Figure 1: Example of a Hearthstone game. The players hero is on the bottom part of the screen (near the visible cards) and the enemy's near the top part (with the hidden cards representing their current hand). Minions on the upper part of the board belongs to the enemy and those on the bottom to our player.

As mentioned above, one of the most interesting features of CCGs is the deckbuilding phase, i.e. creating decks before the game. In HS, players can choose from a pool of neutral cards, and those that belong to the Hero class they have chosen for the game: Mage, Druid, Paladin, Hunter, Rogue, Priest, Warlock, Shaman, or Warrior[1]. Each of these Heroes has a specific Hero Power that costs 2 mana crystals to activate, can be only used once per turn, and has a different effect depending on the Hero being used.

HS players are used to the concept of an *archetype*: a particular deck type to be used in a specific way or situation. For instance, the *Ice Mage* archetype is a deck in which the Mage uses spells to protect themselves until they are able to get a specific combination of cards in their hand, which combined cause extreme damage to the opponent. Expert players are perfectly familiar with the most popular archetypes, and they often design new archetypes to counter them.

There are many different archetypes, since even a slight variation in deck design could lead to a different game plan, but the most popular or the 'core' ones (from which others are derived) are:

- *Aggro*: a clearly offensive behaviour in which the main objective is to reduce the opponent hero's health attacking it directly, ignoring the enemy minions. The aim is to win the game as soon as possible, using low mana cost cards. Normally, if the opponent has not been defeated after several turns (7 or 8) it is very difficult for the Aggro player to win.

- *Control*: a defensive conduct, in which the player tries to eliminate the enemy minions in the first turns, controlling the board, to get to the last turns alive and then to use high-power cards, that are also expensive in mana cost, to finish the game.

- *Midrange*: an intermediate profile between the two previously described ones. The aim is to control the board at the very beginning and become aggressive later, during the mid-game. It uses cards with a good balance between attack, defense and cost.

- *Combo*: the goal of players of this archetype is to hold on without dying until they draw a specific combination of cards from their deck, which then allows them to kill the enemy hero in a single turn using the combined effect produced by those cards.

[1]There is a tenth Hero class launched in May 2020: the Demon Hunter. However, when this study was conducted, there were no available data/decks for this class.

Due to the huge popularity of the game, several platforms have emerged for users to share the card lists of the decks they use. One of the largest is *Hearthpwn*[2], where players vote, copy and discuss the most popular decks. Currently, this database has more than 600,000 decks. These data have been obtained using *crowdsourcing*, which therefore allows us to study organic, extensive and dynamic data produced by the users [3].

This paper aims to extract information from a large database created by players of a DCCGs, conducting a *Game Data Mining* study [3]. Specifically, we will perform a clustering process to automatically detect groups of decks that share features, and check, with the help of an expert, if they fit known archetypes. This might be of interest to developers of such games, who can understand how players use the resources made available to them, and how to adapt the game over time, for example by modifying cards or rules, or creating new cards. Furthermore, it can be of relevance to researchers in the creation of intelligent agents playing CCGs, for example to adapt the behaviour of the agents if they detect that the opposing player is using a particular archetype.

The steps followed in this study are: First, we downloaded all the decks available on the *Hearthpwn* website on September 1st, 2019. After pre-processing the raw data, we then performed a descriptive analysis of the dataset to get a first overview of the data. For each class a co-appearance network is created, where each node is a card, and each edge is the number of decks that share that card. Clustering methods such as K-Means [4] and Agglomerative Hierarchical Clustering [5] are used to obtain different groups in each class. Then a visualisation of every network based on the Spanning tree computation is done, and finally the obtained results are analysed trying to find archetypes in the clusters as well as representative cards to study.

This article continues the preliminary research initiated in [6], in which a simpler clustering study limited to two classes was conducted. Given the promising results and interesting conclusions we reached, we decided to extend our study applying more advanced methods for data analysis and visualization.

Although the clustering methods used in this paper have been widely studied in the past in other fields, we have not found a deep expert analysis based on their application to a very large amount of data (both concerning

[2] `https://www.hearthpwn.com/`

the number of decks, as well as in the number of dimensions considered in the dataset). We believe that the proposed methodology and the conclusions reached may be of interest to a large number of researchers. For example, those who want to understand the behaviour of players (especially CCG players) from a large amount of data. It can also be of interest to researchers in artificial intelligence who focus on this type of video games (for example, to predict what cards an enemy will play in future turns). In addition, the used dataset, which has been composed for this study (through web scraping and pre-processing), has been made publicly available to the community at the following URL: `https://bit.ly/HSdataset`.

The rest of the paper is structured as follows. First, the state of the art in CCGs and Game Data Mining and visualisation is described in Section 2. Then the methodology used to download the dataset, generate the networks, and apply the algorithms is presented in Section 3. The next part (Section 4) shows different graphs and clustering results and discusses them. Finally, conclusions and future work are presented in Section 5.

2. State of the art

Game Data Mining [7] is one of the multiple research lines on video games. This is understood as the application of Data Mining techniques to datasets related to any video game, such as telemetry measures, user-monitoring data, player-generated information, play recordings, etc. The most common aim is the extraction of knowledge, mainly focused on obtaining conclusions on any of the game factors related with player experience [8]: enjoyment, playability, engagement or balance. These conclusions could help designers improve the game mechanics. Other approaches are centered on modelling the player's behaviour itself [9], which can be used for the creation of non-player characters, for instance.

Obtaining the dataset is the main bottleneck in these studies. So, even if there are several papers about Game Data Mining, the different video games analysed are just a few (those for which there are available data). For instance, Thurau and Bauckhage [10] analysed more than 190 million records (from 4 years) of *World of Warcraft*, and found different tendencies in the evolution of guilds. Weber and Mateas [11] applied classification techniques in order to forecast enemy behaviour in *StarCraft*. Also *Madden NFL* [12] and (Infinite) *Super Mario* [8] have been studied from this perspective.

However, regarding this line, the most prolific game so far has been *Tomb Raider: Underworld*, which has been deeply analysed in a considerable number of contributions. Drachen et al. have several works applying different data mining and machine learning techniques to more than 1,300 records of players that have finished the game, such as [9], where the authors applied Self-Organising Maps to identify player models (archetypes), or [13] in which the researchers used classification methods in order to predict players' behaviour with respect to their game finishing time (or their potential withdraw).

Also, Reanudie et al. [14] worked on player modelling, and conducted a categorical clustering study on the game *The Division 2*, in order to identify behavioural patterns related to 9 different builds (analogous to our archetypes).

The objective of the present paper is to analyse data to find deck archetypes in the DCCG *Hearthstone*, which has been very prolific in the literature as the environment for many different studies.

Previous research has been mainly focused on the creation of competitive agents to play the game autonomously [15, 16, 2], with other works centered on the design part, such as the game mechanics analysis or game balance testing [17], using agents.

Data mining has also been applied to HS. Indeed there have been two Data Mining Challenges (AAIA'17[3] and AAIA'18[4]) using this game as a test-bed. However, the 2017 Challenge and the derived papers [18, 19] were devoted to help the AI win the game, whereas the 2018 edition and related papers [20, 21] had the aim to predict the win-rates for specific decks. Regarding this, [22] used decks archetypes for the prediction of battle outcomes, creating clusters, and predicting the win-rates comparing their similarity with standard archetype decks.

The work by Zuparic et al. [23] also analyses a large amount of Hearthstone data (3 years), associating the decks archetypes with their reached win-rates. However, the aim is to study, from the information theory point of view, the evolution and entropy on the players' deck building with respect to the different expansions and game updates launched along the years. In a similar way Fontaine et al. [24] conduct an analysis of decks in Hearthstone,

[3]https://knowledgepit.ml/aaia17-data-mining-challenge/
[4]https://knowledgepit.ml/aaia18-data-mining-challenge/

but with the aim of studying the game balance. Mesentier et al. [25] also worked on balancing HS considering standard decks, but from the point of view of the optimisation of possible changes on cards features by means of an Evolutionary Algorithm.

Deck archetypes have been considered in some other papers, such as [26], where the authors try to predict, using a Recurrent Neural Network, the deck archetype of a player just considering the first card he/she plays (in the first turn). They are able to predict with a 78% accuracy specific archetypes for some heroes.

In this study we will apply clustering methods to a large HS dataset (extracted from a website), but instead of trying to model player behaviour as in [27], we aim to discover key features (cards in this case) in predefined decks which could lead us to identify a cluster or set of decks as belonging to an archetype. This would help to (automatically) identify game 'profiles' in those decks belonging to the same cluster as an already known archetype, which could be useful for developers (to evaluate game mechanics or the impact of an expansion) and also for autonomous agents (to decide the best strategy to face an opponent, or to predict the following cards he/she will play [28]).

Another research line related to Game Data Mining is *Game Data Visualization*, i.e. the use of graphical tools to project information related to a game, in order to make it easier to interpret that information, as well as to detect or find hidden information, such as relationships between different game features or variables. For example, the work by Wallner and Kriglstein [29] revised many different gameplay visualisation approaches, i.e. player behaviour during the game. Or [30], where Drachen and Schubert described different tools and algorithms to show spatial information of games such as trajectories, heat maps or players' behaviour.

However, our aim here is to follow the steps of studies such as [31], where the authors analysed, using clustering and visualisation methods, data gathered from the game Overwatch. They projected the resulting clusters in several graphs focusing on different features. Thus, visualisation tools have been also applied in the present work, namely, the output of the Agglomerative Hierarchical Clustering algorithm [5], the *dendrogram*; or a specific graph-based visualisation, such as the presentation of Spanning Trees by means of the Davidson and Harel layout algorithm [32]. The latter method has been used for the presentation of cards relationships in all the studied decks, with the aim of detecting the most representative or key cards for every deck, as well as their pairings, which could be of interest for game designers in order to find unbalanced (too powerful) cards, for instance.

So, this work presents a complete analysis on a large dataset (with a considerable number of dimensions) related with HS archetypes, extracted from a website where players publish their decks. The analysis has been conducted from different perspectives, i.e. we have first performed a descriptive study, then applied machine learning methods for clustering and visualization, and finally, complemented it with a deep revision by an expert player. The expert has extracted interesting insights that could be relevant for the scientific community studying Hearthstone, game designers, and also for other players.

The methodology followed in the study is described in the next section.

3. Methodology

3.1. Obtaining the dataset

As the objective of this work is to analyse the decks that players create, it is necessary to obtain a large amount of data. In our case, we have used the data available on a repository: the HearthPwn website (`https://www.hearthpwn.com/`). This database contains information about all the cards available in the game, and offers to its users the possibility to create and share decks built from those cards. Currently there are more than 600,000 decks created, that can be filtered by expansion (set of cards launched by the game developers from time to time), hero class, or type of game, among others. Users can view other users' decks and copy them into the game to use against other players.

To download the data we have developed a script in Python that iterates over `deck ids` to obtain the corresponding URL and downloads the webpage with the deck list. The information obtained in HTML format is parsed using the *BeautifulSoup* [5] library to obtain the list of cards, the date, the class, and the game type of deck (the HS game types we considered are: Ranked, Tavern Brawl, Arena and Adventures). With the name of the single cards in each deck, it is also possible to access to more information, such as the cost of making the complete deck with Arcane Dust (the virtual currency of the game), the *mana* cost of each card, or the card type: Spell, Minion

[5]`https://pypi.org/project/beautifulsoup4/`

or Weapon. Other information, such as the Rarity of cards, can also be extracted.

We have limited our analysis to decks belonging to the "Ranked" category, a popular game mode where players compete in a ranked ladder. It is also the most common game type in the whole dataset, representing 62% of all decks.

We encoded the dataset so that each sample (row) corresponds to a deck, and each feature (column) corresponds to a card from the entire collection. A '1' in a cell position indicates that the deck has that card, a '2' indicates that it has 2 copies (the maximum allowed for non-Legendary cards) and a '0' indicates that the card is not included in the deck.

Our dataset can be found and downloaded at `https://bit.ly/HSdataset`.

3.2. Method of analysis

Initially, we perform a descriptive analysis of the dataset, to assess the number of decks per Hero Class, the date of creation, or the most common cards of each class. This is useful as an initial overview of the whole dataset, and will help to understand further analyses.

Then, a clustering analysis is performed, resorting to two techniques :

- **K-Means** [4], a classic method that starts from a set of patterns and tries to separate them into k different groups, according to their features.

- **Agglomerative Hierarchical Clustering (AHC)** analysis [5], an algorithm which, starting from samples, pairs two by two similar clusters and builds a binary tree, called *dendrogram*, representing their similarity.

The first technique has been applied because it is fast and effective, as it has been proved in a large number of studies with different kinds of data [33]. On the other side, Hierarchical clustering offers a simple visual output, that can be interpreted easily by a human expert, as is the case in this work.

Since each hero has a subset of specific cards that only that class can use, it does not make sense to conduct the clustering analysis with all the cards/decks, as the clusters obtained would be the classes themselves, considering they have disjoint features - their exclusive cards. Thus, the input of both algorithms is the dataset, but split by hero class.

While in K-Means we want to detect if we can extract archetypes (clustering decks), in the AHC we want to extract information about how cards are related (clustering cards). That is the reason why in the AHC the input is the transpose of the array: now each card is a row, and each feature (column name) is the ID of the deck to which the card belongs.

In order to complement the data analysis, we have applied a visualisation algorithm to the dataset. Thus, we have composed a graph per hero where the nodes are cards and the links are their relationships, i.e. the appearance of both cards together in one of the studied decks belonging to that hero class. So, the links have an associated arity.

Each of these graphs can be seen as a Complex Network [34] indeed, since there exist patterns of connection between their nodes that are neither purely regular nor purely random; and it also presents a *long tail* effect (few cards with many connections and many cards with few connections), for instance.

Thus, given the high complexity of every graph, we have computed the **Minimum Spanning Tree** [35] for each of them. A Spanning Tree is a graph representation in which all the nodes are reachable/connected, i.e. there is at least one connection with every node, but there are no cycles or loops. Once this tree is built, it is plotted by means of the **Davidson-Harel layout algorithm** [32], which is based on the application of a Simulated Annealing metaheuristic [36] to deploy the nodes in an aesthetic (and readable) way. The aim of this graphical representation of data is to demonstrate its value as a tool for the identification of unknown card relationships, as well as a detector of potential critical cards which might be analysed, for example, to check if they are unbalanced, i.e. too powerful when compared to other similar cards.

The obtained results of the analysis and visualisation are presented and discussed in the following section.

4. Data Analysis and Visualisation

In this section, we first describe the whole dataset visually. Then, different Game Data Mining and Visualisation methods are applied, and their outputs are presented and commented, mainly from the expert player and developer point of view, since we argue that these results are of interest mainly for those two actors.

4.1. Descriptive Analysis of the Dataset

Once the data are extracted and preprocessed to correct errors, 540,000 decks will be considered in this study. Figure 2 shows the distribution of dataset decks by hero class. Although they have a similar number, there is a 32% difference between the class with the highest number of decks (Priest) and the one with the least (Warrior). The most common classes (Priest, Mage and Druid) are also more oriented towards control archetypes and long-term strategy, so this can explain the variability of user-created decks.

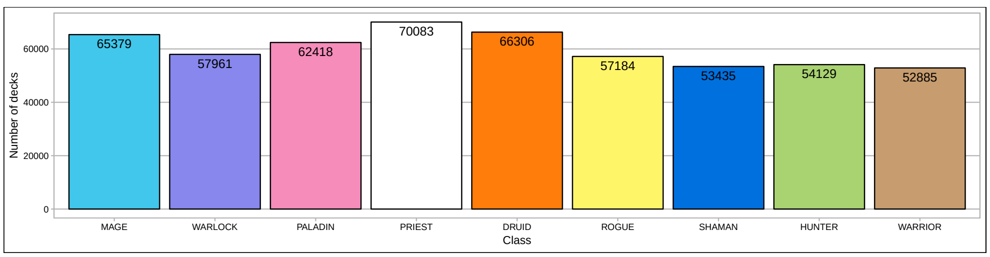


Figure 2: Number of decks by hero class.

Figure 3 shows the creation of decks over time. The spikes that occur immediately after a new expansion emerges can be seen clearly. Therefore, even though many of these new decks may not be suited to the meta-game that will develop during the season of that expansion, a considerable number of them will be the basis for more refined decks.

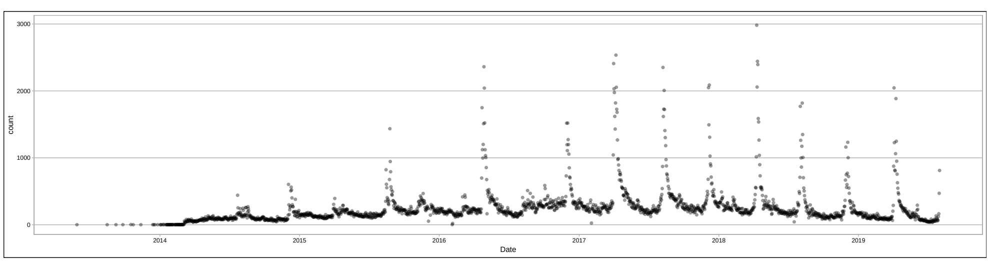


Figure 3: Number of decks introduced in the data bases analysed, over time. It is interesting to observe how the spikes in the number of decks are in correspondence to the release of a new expansion, that added more cards to the available pool while at the same time often removing (or modifying) some of the previously popular cards.

The spread of the number of minions, spells and weapons does not follow a normal distribution, as it can be seen in Figure 4. In fact, following the community (of players) recommendations, ideally every deck should have a third of minions and a third of spells, which matches the last graph. More than half of the decks do not use weapons, as not all classes have cards of that type, but when they are used, the most common number is 2. Hero cards are not shown, as decks that employ them rarely use more than one.

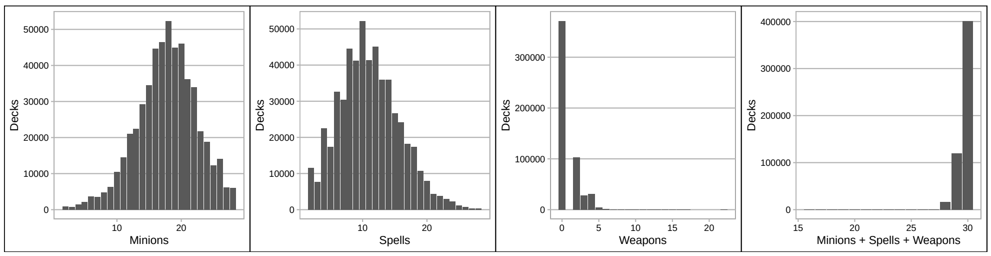


Figure 4: Distribution of Minions, Spells, Weapons.

Figure 5 is particularly interesting, as it shows that, despite having more than 3,000 cards available in the pool, all decks from one class have one particular card with more than 50% chance. For some combinations of class and card, the probability is even higher: for example, the card *Backstab* appears in Rogue decks with 80% frequency. Classes like Mage or Priest have up to 3 cards with a percentage of appearance higher than 70%. The Warlock class is perhaps the least predictable with respect to its top ten, however, there is a minimum of a 30% chance of getting at least one of the top 10 cards right.

After this descriptive analysis, which shows the main characteristics of the considered dataset, we conduct an analysis using Machine Learning methods, namely, clustering algorithms.

4.2. Clustering Analysis

The K-Means algorithm has been applied to the decks in order to assess how they are related. As mentioned above, the dataset has been divided into 9 parts, one per hero, and the clustering method has been used on each of them. We have set to 10 the number of clusters for each class, a value expected to produce enough variety of archetypes, while delivering a reasonable amount of data to be analysed. This number corresponds to the 4 archetypes: Aggro, Control, Midrange, Combo; plus their 6 possible combinations: Aggro-Control, Aggro-Midrange, Aggro-Combo, Control-Midrange, Control-Combo, Midrange-Combo.

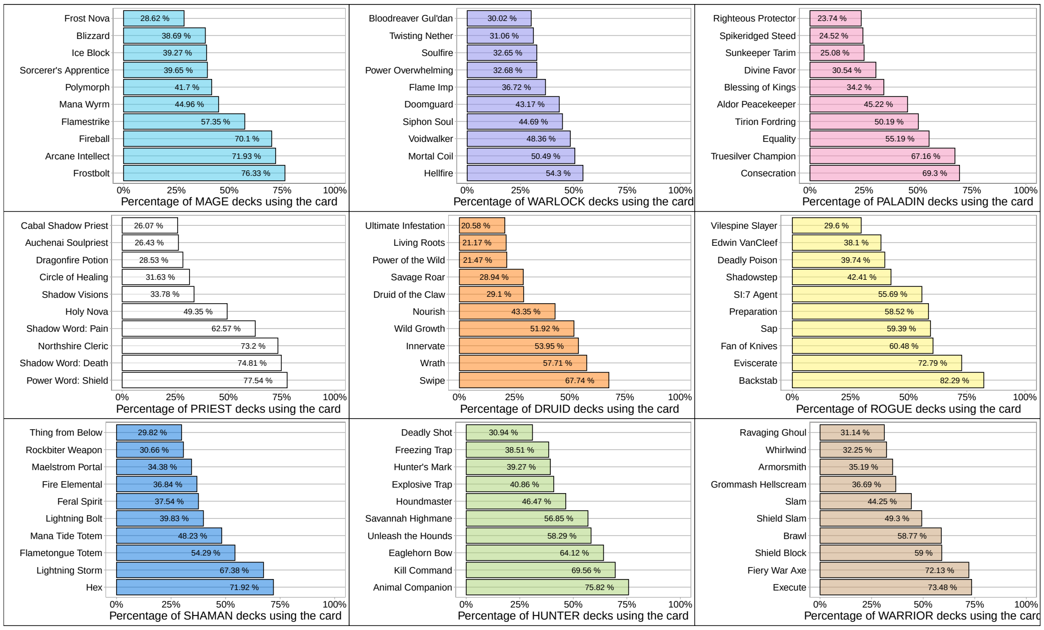
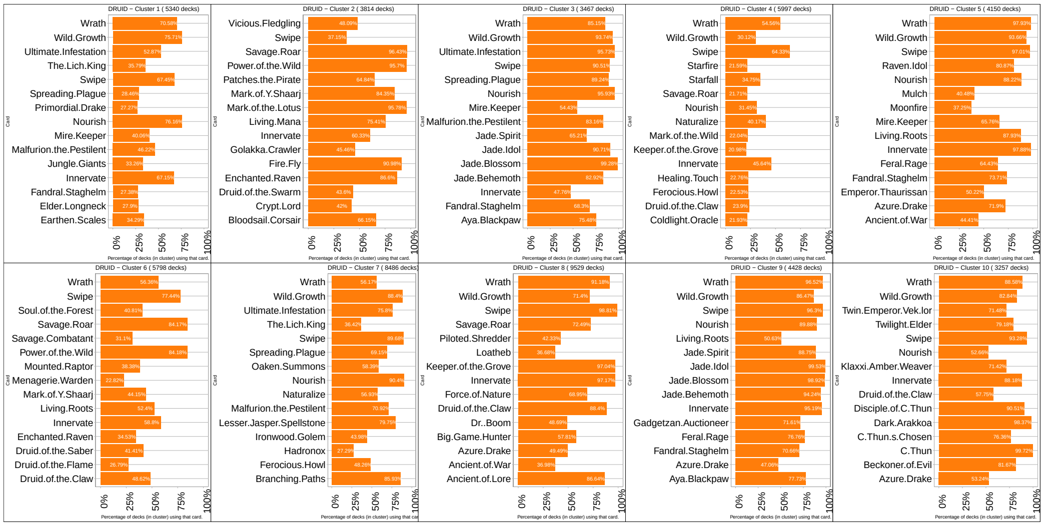


Figure 5: Most commons cards in decks, by hero class. Interestingly, some of the cards that appear most frequently had been banned or 'nerfed' (cost increased and/or effectiveness reduced) over time by the game developers. Notable examples are *Power word: Shield* for Priest, *Innervate* for Druid, *Hex* for Shaman, and *Fiery War Axe* for Warrior.

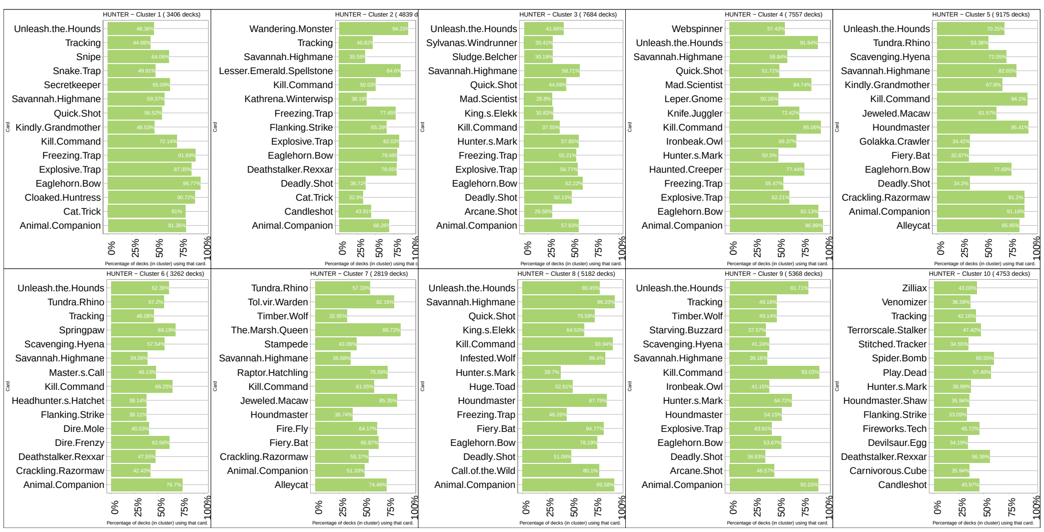
After applying K-Means, we extracted the 15 most common cards from the decks of each cluster, and plotted their frequency in several figures, divided by hero, see Figures 6, 7, 8, 9, and 10.

One of the authors, an experienced Hearthstone player that reached the highest rank (Legend) in the competitive ladder, manually inspected the clusters and provided an expert analysis for all the classes. In the following, the notation used for clusters is the initial/s of the hero class, plus the cluster id (e.g. M2 indicates the second cluster for the Mage class).

Druid. (Figure 6.a) Clusters D1, D5, D7 all present cards that provide advantages in the late game (such as *Wild Growth* and *Nourish*); but while D1 and D7 have control cards (such as *Starfall*), D5 exploits the late-game advantage to close combos, using potentially one-turn-kills like *Malygos* or *Aviana*. Clusters D2 and D6, on the contrary, have none of these cards, but feature weak, cheap creatures such as *Arcane Raven* and *Fire Fly*, plus cards that enhance all friendly creatures on the board, such as *Savage Roar*, thus grouping decisively Aggressive archetypes. D3 and D9 show a preponder-



(a) Druid



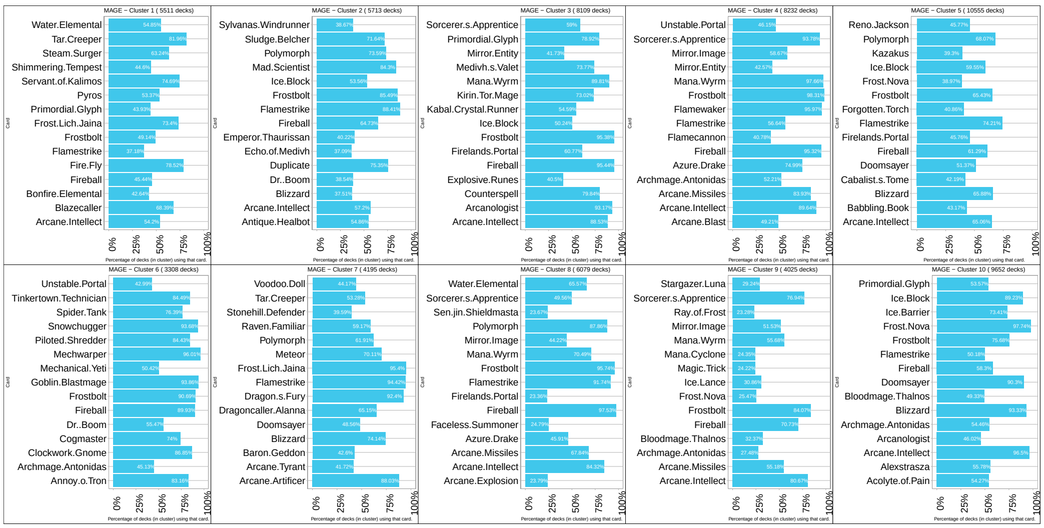
(b) Hunter

Figure 6: Top 15 most common cards of each cluster for Druid (a) and Hunter (b).

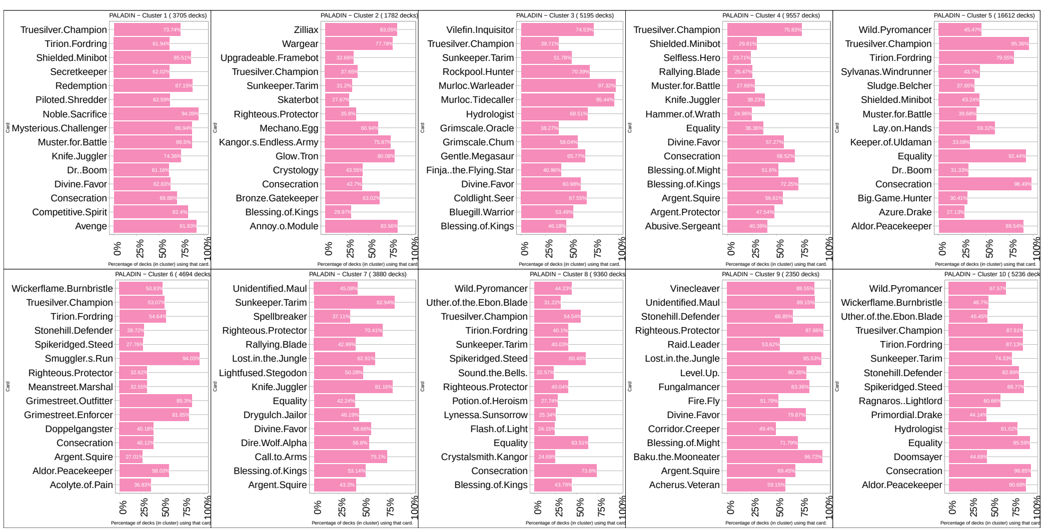
ance of *Jade* cards (*Jade Idol*, *Jade Spirit*, *Jade Behemoth*), thus placing these decks in the category of *Jade Druid*, a specialised midrange archetype. Cluster D10 presents mostly cards with the *C'Thun* keyword, identifying the decks belonging to this cluster as variants of the combo *C'Thun Druid* archetype. Clusters D4 and D8 are harder to categorise, as they seem to either be mid-range variations of aggressive decks, or present poor cohesion, possibly representing outliers.

Hunter. (Figure 6.b) Hunter is a Hero class featuring cards particularly suited to aggressive plays, so it is not surprising to see a prevalence of clusters corresponding to fast decks, featuring cheap cards. Clusters H4 and H9 have a strong prevalence of cards able to deal large amounts of damage (*Eaglehorn Bow*, *Kill Command*). H5, H6, and H8 represent equally aggressive decks, with cards exploiting synergies between minions of type *Beast*, such as *Houndmaster*, plus Beast-type minions that were popular during different seasons of the game (*Crackling Razormaw*, *Infested Wolf*, *Springpaw*). Cluster H10 is again an aggressive deck, but this time featuring minions of type *Mech*, such as *Spider Bomb* and *Zilliax*. H1, H2, and H3 are more midrange decks, based around Secrets and synergistic cards, such as *Cloaked Huntress* and *Lesser Emerald Spellstone*. Finally, cluster H7 describes decks built around the Quest *The Marsh Queen*, that requires the player to put on the battlefield 1-cost minions such as *Jeweled Macaw* and *Alleycat*; for that reason also features cards that fetch 1-cost minions, such as *Tol'vir Warden*. Among all Hunter clusters, H3 and H10 show the least amount of coherence, with the most common card appearing only slightly more than 60% of the times.

Mage. (Figure 7.a) Clusters M3, M4, M6, and M9 all represent aggressive archetypes, featuring cards such as *Fireball* and *Frostbolt*. M3 exploits synergies with secrets (*Arcanologist*, *Counterspell*, *Medivh's Vallet*), M4 relies upon *Flamewaker* and cheap spells to damage the opponent, M6 shows a strong presence of *Mech* minions (*Mechwarper*, *Snowchugger*), and M9 is based around cheap spells (*Ice Lance*, *Magic Trick*) plus the minion *Mana Cyclone* to generate new damaging spells. Clusters M2, M5, M7, and M8 all fall under different control archetypes, using either secrets in the case of M2, a large number of board resets (*Doomsayer*, *Flamestrike*) for M5, or a unique late-game finisher (*Dragoncaller Alanna*, *C'Thun*) for clusters M7 and M8, respectively. Notably, cluster M7 also includes decks built using



(a) Mage



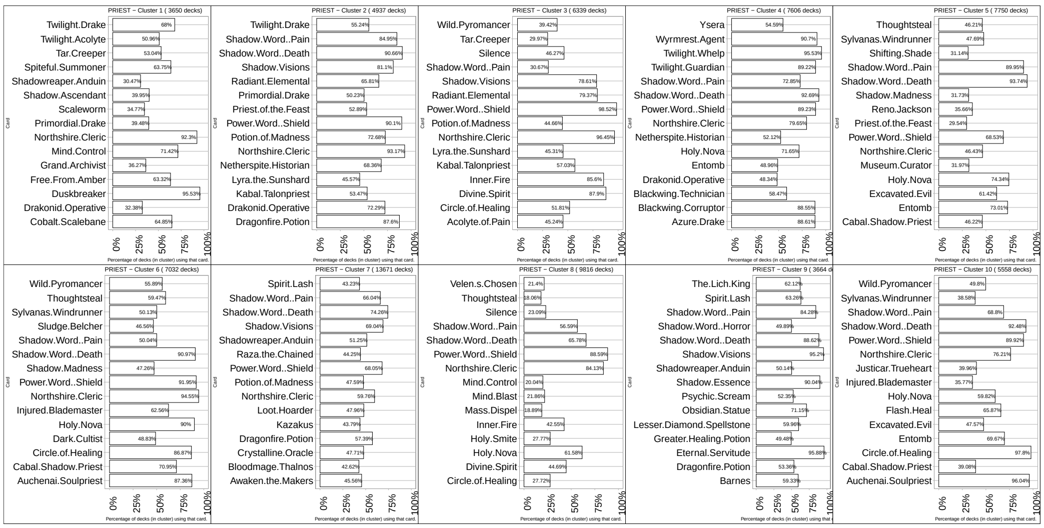
(b) Paladin

Figure 7: 15 most common cards of each cluster for Mage (a) and Paladin (b).

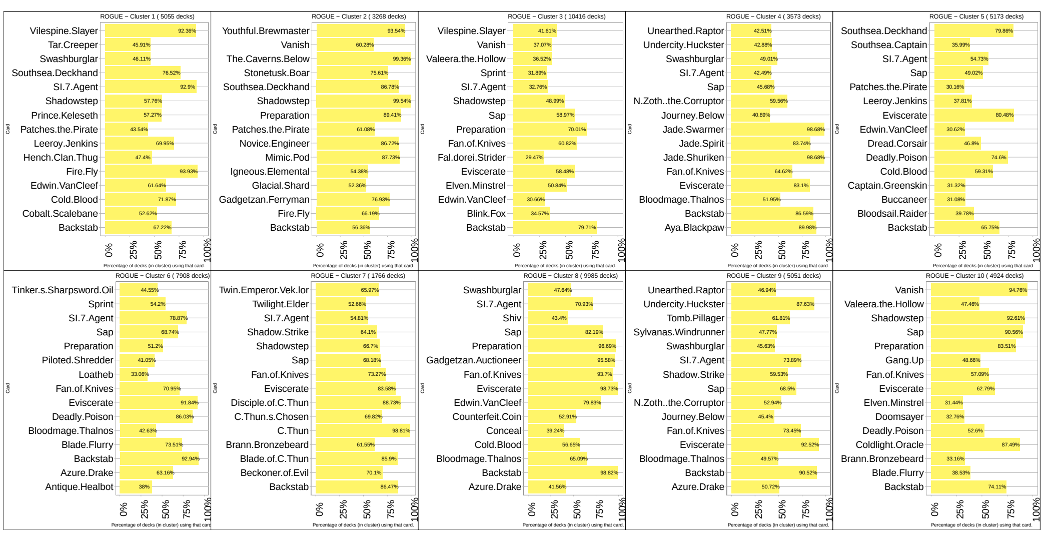
only odd-cost cards, exploiting the synergy with *Baku, the Mooneater*, that provides a powerful effect in exchange for limiting the possibilities of deck construction. Cluster M1 encompasses decks using the synergy between *Elemental* minions and *Jaina, Frost Lich*, thus positioning these archetypes in a mid-range position. Finally, cluster M10 includes combo decks, based on the quest *Open the Waygate*, *Archimage Antonidas*, and *Sorcerer's Apprentice*.

Paladin. (Figure 7.b) Depending on the season, Paladin players favoured decks ranging from aggressive, to midrange, to control. Clusters PA2, PA3, PA4, and PA7 all represent different flavours of aggro Paladin, aiming to overcome the opponent with vast numbers of cheap creatures. PA2 exploits Mech-type minions such as *Annoy-o-Module*, *Zilliax*, and *Glow Tron*. PA3 uses Murlocs, such as *Murloc Warleader*, *Murloc Tidecaller*, and *Vilefin Inquisitor*. PA7 uses smaller creatures that can be put into play directly from the deck using the *Call to Arms* spell, plus synergistic cards like *Knife Juggler*. Among this group, PA4 is the slowest, maybe leaning on midrange, using more expensive cards such as *Truesilver Champion* and *Blessing of Kings*. Clusters PA5, PA6, PA8, and PA10 all belong to the control Paladin archetype, with small variations that are season-dependant. All decks feature cards that can be used to remove large number of creatures, such as *Consacration*, *Equality*, and *Wild Pyromancer*. PA6 adds cards that buff minions in the player's hand, like *Smuggler's Run* and *Grimestreet Outfitter*. PA8 groups control decks that are focused around a big finishing card, such as *Lynessa Sunsorrow* or *Shirvallah the Tiger*. Cluster PA9 represent a set of midrange decks, featuring only odd-cost cards and *Baku the Mooneater*, a card that powers up the base Paladin's hero power if your deck contains only odd-cost cards. Finally, PA1 groups decks based on Secrets such as *Noble Sacrifice* and *Avenge*, plus *Mysterious Challenger*, a minion that is able to put a large number of Secrets in play directly from the deck.

Priest. (Figure 8.a) Most Priest decks fall under the control or combo archetypes, as the class is not very suited to aggression. Interestingly, all clusters present Power Word: Shield as one of the most frequently used cards. The widespread adoption of this card is likely one of the reasons why it was later completely changed by the Hearthstone team. Clusters PR1, PR2, and PR4 all feature control spells able to destroy large number of minions, such as *Dragonfire Potion*, plus expensive Dragon-type minions, like *Duskbreaker*, *Twilight Guardian*, *Drakonid Operative*, and *Cobalt Scalebane*, all of which can be



(a) Priest



(b) Rogue

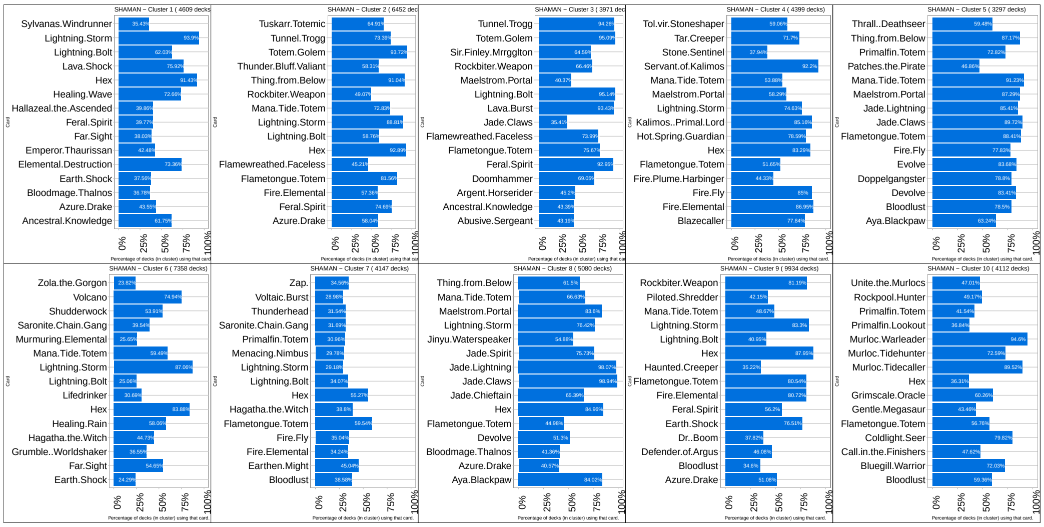
Figure 8: Top 15 most common cards of each cluster for Priest (a) and Rogue (b).

either finishers in the end game, or provide advantages in the mid-game. PR6 and PR10 group again control decks, this time based on *Auchenai Soulpriest*, a minion able to turn healing into direct damage, and healing cards such as *Circle of Healing* and *Flash Heal* that can then be used either to heal friendly minions or kill enemy ones. PR3 and PR8 group combo decks based on *Divine Spirit*, that doubles a minion's health, and *Inner Fire*, that gives a minion attack equal to its health. Usually, these combo decks close the game with a single attack that clears completely the opponent's health. PR7 is another combo deck, this time based around the interaction between *Shadowreaper Anduin* and *Raza the Chained*; the cluster also features prominently *Awaken the Makers*, a Quest that can help the Priest player survive until the combo is complete, that might also be played independently from the Anduin-Raza combination, and that is probably why the cluster itself is not very coherent. PR5 represents a type of control decks that only use one card of each type, to then exploit *Reno Jackson* to heal. PR9 is another version of control decks, that uses spells like *Shadow Essence* to play powerful creatures in the early game, and then keeps resurrecting them with spells such as *Eternal Servitude* and *Lesser Diamond Spellstone*.

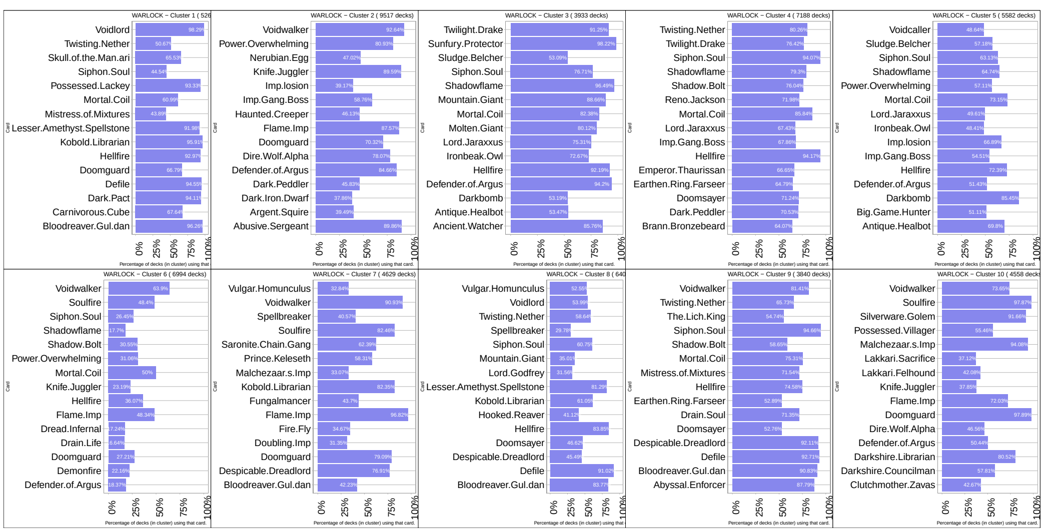
Rogue. (Figure 8.b) This hero class features cards well-suited to combo decks, such as *Preparation*, that reduces the cost of other spells, and several ways of drawing large amount of cards. The ability of Rogue of going from almost losing to winning in one round by drawing several cards in a row prompted players to name several archetypes *Miracle* Rogue. Cluster R8, for example is focused around *Gadgetzan Auctioneer*, a minion that lets the player draw a card each time they play a spell, plus cheap spells: the idea is to draw the entire decks and release a massive amount of damage in one turn. R10 groups decks that aim to force the opponent to draw until their deck is depleted, using *Coldlight Oracle* and *Gang Up*. Cluster R2, for example, is based around the quest *The Caverns Below*, which empowers all minions but requires to play the same minions several times, which is possible through cards like *Shadowstep* or *Youthful Brewmaster*. R9 is based on the idea of using Deathrattle minions such as *Tomb Pillager* and *Undercity Huckster*, and then resurrect them multiple times by playing *N'Zoth, the Corruptor* and *Shadowstep*. R3, R6, and R7 are all variations of control decks, using cards like *Sap* and *Fan of Knives*, based either on finishing the game attacking with a weapon (*Tinker's Sharpsword Oil*, *Blade Flurry*) or with C'Thun, a creature that deals damage when it comes into play if it was powered up by synergistic cards like *Disciple of C'Thun* or *Blade of C'Thun*. Among these three clusters, R3 is the least coherent, as its most frequent cards (*Backstab* and *Preparation*) tend to appear in a variety of different decks. R1 and R5 group aggressive decks, based on cheap creatures and either *Prince Keleseth*, or synergy between Pirate-type minions like *Southsea Deckhand* and *Dread Corsair*. Finally, R4 is a midrange deck based on *Jade* cards like *Jade Shuriken*, *Jade Swarmer*, and *Aya Blackpaw*.

Shaman. (Figure 9.a) The game mechanic unique to Shaman is *Overload*, that makes it possible to play extremely cheap cards, by locking some mana crystals in the next turn, trading an immediate advantage for a future drawback. Clusters S2 and S3 both represent aggressive decks built around this mechanic, with cheap Overload minions like *Totem Golem* and *Flamewreathed Faceless*, plus others that gain advantages if the player Overloads, like *Tunnel Trogg*. Other aggressive decks are highlighted by cluster S10, based around Murloc-type minions with cards like *Murloc Warleader*, *Murloc Tidecaller*, and *Coldlight Seer*, and S5, that aims at having large amounts of minions in play through effects like *Doppelgangster* and then use *Evolve* and *Bloodlust* to power them up. Midrange decks are grouped in clusters S4, S8, and S9, with minor variations: S4 is focused on Elemental-type minions (*Servant of Kalimos*, *Fire Elemental*, *Kalimos, Primal Lord*), S8 is using *Jade* cards (*Jade Claws*, *Jade Lightning*, *Aya Blackpaw*), while S9 is not exploiting synergies between minions of the same type. S1, S6, and S7 all describe control archetypes, with spells that can remove enemy minions like *Lightning Storm*, *Hex*, and *Elemental Destruction*. S1's goal is to kill the opponent with spells like *Crackle* and *Lava Burst*, S6 aims to finish the game using *Shudderwock*, a minion that repeats all Battlecries (come-into-play effects) played so far, while S7 has the objective of gaining late-game advantage with *Hagatha the Witch* and cheap minions like *Fire Fly*; this last cluster, however, is also the least coherent, with the most frequent card only appearing in about 60% of the decks.

Warlock. (Figure 9.b) The Warlock's hero power allows a player to draw cards, in exchange for reducing its life points. This makes the class suitable for both aggressive decks, that tend to run out of cards in their hand in midgame, and control decks, that benefit from having multiple options to choose from in their hand when answering threats. Clusters WK2, WK6, WK7, and WK10 all describe aggressive archetypes, with cheap minions like *Flame Imp*

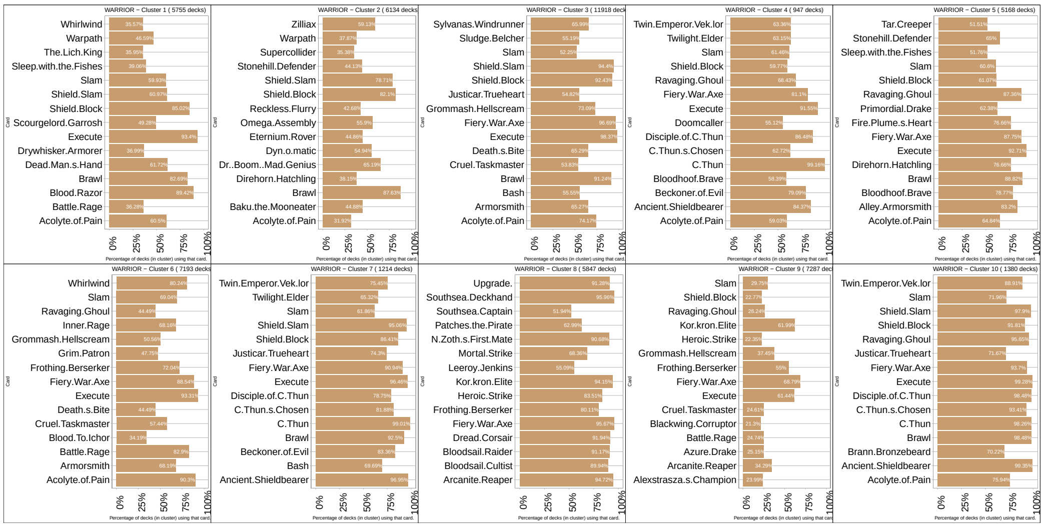


(a) Shaman



(b) Warlock

Figure 9: Top 15 most common cards of each cluster for Priest (a) and Rogue (b).



(a) Warrior

Figure 10: Top 15 most common cards of each cluster for Warrior (a).

and direct damage spells like *Soulfire*. While the clusters are similar, they still show some interesting differences: in particular, WK2 is even more aggressive than the rest, with cheap minions like *Abusive Sergeant* and synergy cards like *Knife Juggler*; while WK10 employs cards that require to discard other cards like *Doomguard*, and synergies that create positive effects when a card is discarded, like *Malchezaar's Imp* and *Silverware Golem*. A large part of decks in the WK7 cluster use *Prince Keleseth*, a card that buffs all minions in the deck, if certain conditions are fulfilled. Cluster WK6 is the least coherent, with the most frequent card, *Voidwalker*, appearing only slightly more than 60% of the times. All other clusters describe control decks. WK1 is based on large Demon-type minions such as *Voidlord*, ways of putting them in play in the early game like *Possessed Lackey*, and ways to resuscitate them like *Bloodreaver Gul'dan*. WK3, on the contrary, groups decks exploiting large minions that have reduction in cost if the player has more cards in their hand (*Mountain Giant*) or if the player is low on hit points (*Molten Giant*), both common occurrences for Warlocks. WK4 describes decks that play one copy of each card to exploit *Reno Jackson*, while WK5 uses cards that were more common in specific seasons, like *Darkbomb*. WK8 and WK9 are very similar, as they both use the removal *Defile*, and employ *Bloodreaver Gul'dan* as a late-game swing, without all synergies included in WK1.

Warrior. (Figure 10) Clusters W1, W2, W3, and W4 all represent variations of Warrior Control archetypes. Decks in W1 rely upon *Dead Man's Hand* to try and finish the game through fatigue damage, W2 groups both Mech synergy (*Dr. Boom, Mad Genius*, *Zilliax*) and Odd Warrior (*Baku, the Mooneater*), W3 decks seem to exploit older cards (*Sylvanas*, *Justicar Trueheart*) possibly representing Wild decks, W4 is a Control version of *C'Thun Warrior*, with the *C'Thun* cards and several other synergies. W8 is a set of decisively aggressive decks, with cards such as *Leeroy Jenkins*, *Patches the Pirate*, *Southsea Deckhand*. W6, W7, and W10 all represent combo decks: W6 includes cards that can damage all minions on the board (*Whirlwind*, *Death's Bite*) plus minions that benefit from being damaged (*Grim Patron*, *Frothing Berserker*); W7 and W10 are variations of *C'Thun Warrior*, with fewer control elements with respect to W3, and cards such as *Brann Bronzebeard* to try and finish the game using a colossal amount of damage from *C'Thun*. Cluster W5 groups together Quest Warrior archetypes based on *Fire Plume's Heart*, and more generic mid-range decks still based on Taunt minions (*Stonehill Defender*, *Direhorn Hatchling*). Finally, cluster W9 shows relatively few points in common between its decks, with the most common card being *Fiery War Axe* appearing in only 68% of cases, and might thus represent a collection of outliers, or very different mid-range decks.

Table 1 summarises all the findings, so the reader can easily check the archetypes detected by the expert in the different clusters.

Table 1: Summary of the expert analysis of all clusters found in the experiment, with clusters organised by predominant archetype. Clusters highlighted in red are the least coherent, and harder to characterise.

Hero Class	**Aggro**	**Midrange**	**Control**	**Combo**
Druid	D2, D6	D3, D4, D8, D9	D1, D7	D5, D10
Hunter	H4, H5, H6, H8, H9, H10	H1, H2, H3, H7	-	-
Mage	M3, M4, M6, M9	M1	M2, M5, M7, M8	M10
Paladin	PA2, PA3, PA4, PA7	PA1	PA5, PA6, PA8, PA10	-
Priest	-	-	PR1, PR2, PR4, PR5, PR6, PR9, PR10	PR3, PR7, PR8
Rogue	R1, R5	R2, R4, R9	R3, R6, R7	R8, R10
Shaman	S2, S3, S5, S10	S4, S8, S9	S1, S6, S7	-
Warlock	WK2, WK6, WK7, WK10	-	WK1, WK3, WK4, WK5, WK8, WK9	-
Warrior	W8	W5, W9	W1, W2, W3, W4	W6, W7, W10

4.3. Cluster Cost Analysis

This section presents a study on the card costs distribution per cluster. We aim to test cost distribution as an additional way to potentially detect archetypes in deck builds, given the assumption that the cost distribution is highly correlated with the type of playing profile, as well as with the game pace for which a deck is designed. While this assumption could seem intuitively correct (Aggro decks for example would on average feature cards with lower costs than Control decks), this could not be true for all classes and all clusters.
Then, Figure 11 summarizes the card costs distribution for the 10 clusters previously identified for each hero, using boxplots.

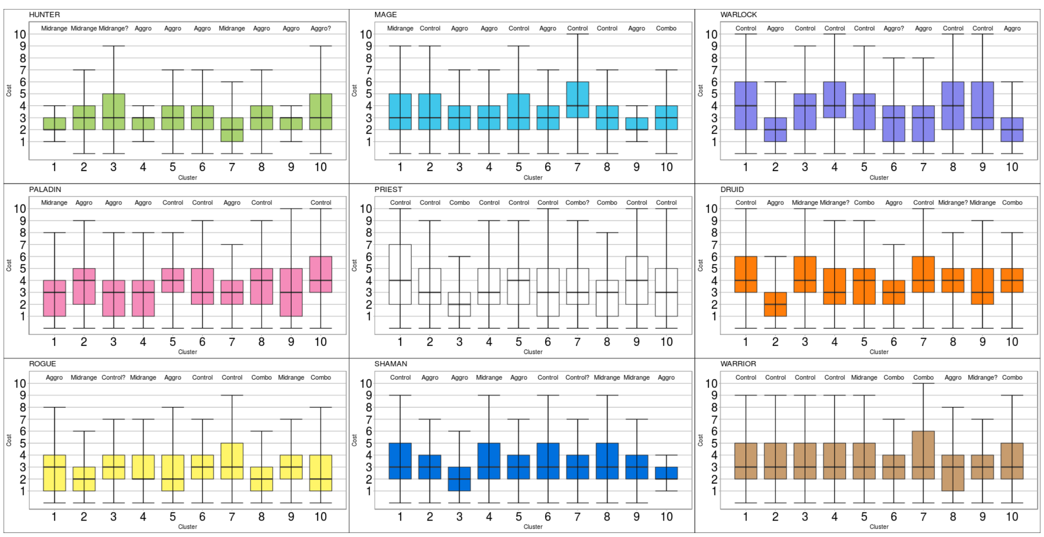


Figure 11: Boxplots showing the cost distribution for the cards on each one of the 10 identified clusters per hero. The clusters have been labelled with the archetype assigned in previous analysis by the expert (See Table 1)

Interestingly, the clusters for Hunter (in green) are hard to tell apart just from their card cost distribution. The expectation is that Aggro decks would feature cards with lower average cost, but since the class is Aggro-oriented, even the midrange decks (for example, H1, H2, H3) would prominently feature low-cost cards, for example the 2-cost Secrets.

Clusters for Mage (in light blue), on the other hand, can be easily distinguished: clusters featuring Aggro decks, like M3, M4, M6 and M9 clearly have a cost distribution skewed towards lower values; Control and Midrange have higher average cost and a larger variety, see for example M2, M5, M7. In some cases, however, it is still hard to figure out differences in archetype just from the cost distribution, as clusters M8 and M10, for example, while being practically identical, refer to Control and Combo, respectively.

Warlock clusters (in dark purple) are also easy to separate just evaluating the cost distribution of their cards: the ones skewed towards lower costs, like W2, W6, W7, and W10, are all Aggro; while the rest, with a clearly higher mean cost and larger variance, are all Control decks.

Paladin is another class for which the clusters (in pink) cannot be straightforwardly classified just by the cost distribution of their cards. The reason is that even aggressive Paladin decks use cards with relatively higher costs.

Just as Hunter is more focused towards aggression, Priest (in white) is a class more focused on controlling the board, so all clusters refer to either Control or Combo decks. Consequently, just considering the cost distribution is not enough information to classify the clusters into archetypes.

Druid (in orange) is another class where separating Aggro decks from the rest is easy: D2 and D4, the two Aggro clusters, clearly have cards with lower mean cost, and less variance. Then again, the cost distribution profiles do not seem enough to tell apart Control from Midrange or Combo. While the two Control clusters D1 and D7 have a wider range of minimum and maximum cost, for example, the same is true for the Midrange cluster D3.

Clusters in the Rogue class (yellow) are extremely hard to differentiate, as this class presents a considerable number of 0-cost cards, such as *Backstab* or *Preparation*, that are shared between a considerable variety of decks, and skew all costs down. Cluster R7 is the only one that stands out, featuring Control decks with high-cost cards.

Shaman (in dark blue) seems to have clearly identifiable clusters. The Aggro clusters S3 and S10 have lower mean costs and a lower variety of card values. The same is true for S2 and S5, but in their case, they are harder to separate from a Midrange S9. While cluster S7 also has a profile very close to those, it must be noted that it was the least coherent in the previous analysis.

Clusters from Warrior (in light brown) are an interesting case study, because all those classified as Control, like W1, W2, W3, W4, have the exact identical profile; and the only Aggro cluster is clearly more skewed towards 1-cost cards than all the rest. Still, the Midrange W5 and the Combo W10 have the same profile as W1-4.

Generally speaking, just considering card cost distribution is not enough to clearly separate the clusters. While Aggro decks tend to feature distributions skewed towards lower costs, this is not always true for all classes; and depending again on class characteristics, distinguishing Combo, Control, and Midrange is often impractical. Thus, in order to properly analyse the clusters, human expertise seems indeed necessary.

4.4. Agglomerative Clustering

Once the clusters generated by K-Means have been analysed, Agglomerative Hierarhical Clustering has been applied. AHC can show also interesting information about the influence of the cards. Due to visibility reasons, just the 100 most popular cards in all the decks per hero have been considered, otherwise the binary tree output, named *dendrogram*, would be completely unreadable.

We show here only the results of the Warrior class as an example, because they are sufficiently representative for all the classes. Thus, Figure 12 shows the complete generated dendrogram of the Warrior cards, and, aiming for utility, the subtrees with height=4 are displayed in more detail in Figure 13.

The height of the fusion, provided on the vertical axis, indicates the similarity/distance between two cards. The higher the height of the fusion, the less similar the cards are. This height is known as the *cophenetic distance* between the two cards. Most of the cards are in a big cluster (subtree 4), but there exist several relevant cards (single cards) that have enough weight to appear in their own subtree, even at level 1. Several pairs of cards shown are usually used in combos, have some kind of synergies or belong to the same expansion. For example: *N'Zoth* and *Bloodsail Cultist*.

Even if this kind of clustering method is normally useful in many domains, it is not the best tool to identify relationships between cards, which we think could be interesting to report. Thus, in the following section we present a visualisation approach more suited to this aim.

4.5. Graph-based Visualisation

Aiming to represent the relationships among cards in all the decks for a hero, we create a graph in which the cards are nodes, and the links model the appearance of two connected cards in any deck. These links will have an associated weight which is the number of appearances of the two cards together in decks.

Figures 14, 15, and 16 show the results of the Minimum Spanning Tree [35] for each class, using the Davidson-Harel layout algorithm [32], on neutral plus hero class cards. Nodes represent cards, and their size depends on their frequency of appearance. The thickness of the links represents their arity, regarding the number of appearances of the two cards connected in the analysed decks.

For the following considerations, it is important to remind that in Hearthstone cards from the Classic set are always available for deckbuilding, while cards from the expansions periodically are phased out of the pool for the Standard decks.

It is easy to notice how the node corresponding to *Azure Drake* is quite sizeable in all classes: it is interesting to connect this strong presence with the fact that the card was removed from the Classic set in April 2017; as the card was too versatile and fit too many different decks, Blizzard took the decision of retiring it. In the same way, *Sylvanas Windrunner*, prominently featured in Paladin, Priest, Shaman, and Warlock, has been removed from the Classic set in the same date as *Azure Drake*. *Defender of Argus*, another popular card for almost all classes, was instead just modified to be less powerful in January 2014. *Acolyte of Pain*, also frequently appearing as a large node in the networks, was removed from the Classic set in March 2020, not because it was considered unbalanced, but because it was a very popular option for card drawing, often becoming an obligatory pick for archetypes focused on drawing, like Combo or Control. Other cards that are popular for multiple classes, like *Emperor Thaurissan* or *Fire Fly*, are part of an expansion, and thus naturally exited the pool of cards after a few months.

Figures 17, 18, and 19, display networks that include only cards exclusive to a Hero class. We can see that sometimes when one card is shared between a large variety of decks, that card was considered a problem and changed over the course of the existence of the game. For example, Druid saw the power of both *Wrath* and *Nourished* reduced over time, with small modifications to cost and effect. The same is true for *Unleash the Hounds*, a Hunter card that was modified multiple times (once in 2013, and twice in 2014). Rogue's *Preparation* was also reduced in power in May 2019. *Hex*, a Shaman card, had its cost increased along with *Fiery War Axe*, a Warrior card, in the same patch (September 2017). Priest's *Northshire Cleric* was instead removed from the Classic set in March 2020, after a large modification of the class cards that happened in the same patch. Warlock's *Doomguard* followed the same fate in April 2019, as Blizzard believed it to not be aligned with the

Warlock's class identity.

While it seems that the representation we presented can immediately highlight potentially problematic cards, it is worth noting that other prominent class cards were not targeted by patches. Usually, cards that can fit just a single archetype (Aggro, Midrange, Control, Combo) are not considered problematic, even if almost all decks of that archetype choose to include the card. One good example for neutral cards is *Doomsayer*, pretty popular for Control decks, but never used in Aggro or Midrange, and rarely seen in Combo decks.

5. Conclusions

Understanding how players play a game is a major concern for developers, as they can adapt elements of the game, such as rules and content, to adjust the balance or fun it can provide. In this paper we propose to use Game Data Mining [7] and Data Visualisation techniques, to obtain information about how players create *Hearthstone* decks. The goal is to demonstrate whether it is possible to identify deck archetypes starting from a large set of user-created card lists. To this aim, we have extracted a dataset with more than 500,000 players' defined decks from the *HearthPwn* website, and performed a descriptive analysis, together with clustering and graph-based visualisation algorithms.

After expert analysis of the results, we have provided information on how the cards are related to each other, and how it is possible to detect different archetypes from the data created by the users. However, the proposed automatic clustering approach also showed a few limitations: 3 out of the 30 clusters analysed seems to be composed of mostly outlier decks, identifying no clear archetype (D4, D8, W9); moreover, distinct clusters in the same hero class seem to present very similar archetypes (W7, W10); and finally, it is sometimes possible to detect two distinct archetypes inside the same cluster (M7, W2). These issues are typical of clustering, an unsupervised machine learning methodology for which there is no ground truth, and parameters such as the expected number of clusters or their expected density have to be defined *a priori* by the user.

Relationships between cards can be seen visually using the dendrogram generated by the Agglomerative Hierarchical Clustering algorithm, however, it has also some limitations regarding its visibility. Thus, Spanning Tree +

Automatic Layout display methods have been applied on the datasets, representing in an interpretable graphical way the relationships between cards.

The results obtained are not only relevant, for example, for the understanding of the current state of the game based on information provided by the most enthusiastic players. Artificial Intelligence researchers who want to improve algorithms that play the game can use the generated clusters to predict which cards are most likely to be drawn in future turns from those that the enemy has already used [28].

As future work, a card co-appearance matrix can be created, in which each cell is the number of decks that share that particular pair of cards. Using other clustering algorithms, such as the Leiden Algorithm, card communities can be detected [37], and from using their centrality and density measures, these communities can be plotted in a strategic diagram to see what decks belong to motor, transversal, specialised or emerging/disappearing categories. Other visualisation techniques, such as plotting and studying the changes in decks over time and different expansion releases may help to understand how users play the game.

A feature extraction method could be also applied, in order to 'generate' features related to the decks, such as summarising the number of minions in the set, or the amount of beast cards, weapon cards, or combo cards, to cite some examples. Other relevant information, such as the ladder level achieved by the player who created the deck, could also be used. All this information could potentially better describe the decks for their analysis.

Moreover, other clustering algorithms such as *Density-Based Spatial Clustering of Applications with Noise* [38] can partially solve the issue of deciding *a priori* the number of clusters; nevertheless, they feature different parameters to be tuned.

Finally, we will also explore the use of categorical clustering approaches, as proposed for example the by the authors of [14].

Acknowledgements

This work has been supported by the Ministerio español de Economía y Competitividad under projects PID2020-113462RB-I00 (ANIMALICOS) and PID2020-115570GB-C22 (DemocratAI::UGR). It has been also funded by projects P18-RT-4830 and A-TIC-608-UGR20, granted by Junta de Andalucía, and project B-TIC-402-UGR18 (FEDER and Junta de Andalucía)

References

[1] G. N. Yannakakis, J. Togelius, Modeling Players, Springer International Publishing, Cham, 2018, pp. 203–255.

[2] P. García-Sánchez, A. P. Tonda, A. J. F. Leiva, C. Cotta, Optimizing Hearthstone agents using an Evolutionary Algorithm, Knowl. Based Syst. 188.

[3] C. Rughini, Citizen science, galaxies and tropes: Knowledge creation in impromptu crowd science movements, in: 2016 15th RoEduNet Conference: Networking in Education and Research, 2016, pp. 1–6.

[4] J. A. Hartigan, M. A. Wong, Algorithm AS 136: A K-Means clustering algorithm, Journal of the royal statistical society. series c (applied statistics) 28 (1) (1979) 100–108.

[5] B. S. Everitt, S. Landau, M. Leese, D. Stahl, Hierarchical clustering, Cluster analysis 5 (2011) 71–110.

[6] P. García-Sánchez, A. Fernández-Ares, A. M. Mora, Data mining of deck archetypes in Hearthstone, in: R. Lara-Cabrera, A. J. F. Leiva (Eds.), Proceedings of the VI Congreso de la Sociedad Española para las Ciencias del Videojuego, On-line, October 7-8, 2020, Vol. 2719 of CEUR Workshop Proceedings, CEUR-WS.org, 2020, pp. 132–144. URL `http://ceur-ws.org/Vol-2719/paper14.pdf`

[7] A. Drachen, C. Thurau, J. Togelius, G. N. Yannakakis, C. Bauckhage, Game data mining, in: M. S. El-Nasr, A. Drachen, A. Canossa (Eds.), Game Analytics, Maximizing the Value of Player Data, Springer, 2013, pp. 205–253. `doi:10.1007/978-1-4471-4769-5\_12`.

[8] B. G. Weber, M. Mateas, A. Jhala, Using data mining to model player experience, in: FDG Workshop on Evaluating Player Experience in Games, ACM Press, 2011.

[9] A. Drachen, A. Canossa, G. N. Yannakakis, Player modeling using self-organization in Tomb Raider: Underworld, in: 2009 IEEE symposium on computational intelligence and games, IEEE, 2009, pp. 1–8.

[10] C. Thurau, C. Bauckhage, Analyzing the evolution of social groups in World of Warcraft®, in: Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, IEEE, 2010, pp. 170–177.

[11] B. G. Weber, M. Mateas, A data mining approach to strategy prediction, in: 2009 IEEE Symposium on Computational Intelligence and Games, IEEE, 2009, pp. 140–147.

[12] B. G. Weber, M. John, M. Mateas, A. Jhala, Modeling player retention in Madden NFL 11, in: D. G. Shapiro, M. P. J. Fromherz (Eds.), Proceedings of the Twenty-Third Conference on Innovative Applications of Artificial Intelligence, August 9-11, 2011, San Francisco, California, USA, AAAI, 2011, pp. 1701–1706.

[13] T. Mahlmann, A. Drachen, J. Togelius, A. Canossa, G. N. Yannakakis, Predicting player behavior in Tomb Raider: Underworld, in: Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games, IEEE, 2010, pp. 178–185.

[14] D. Renaudie, R. Lizatovic, A. Azadvar, R. Elmqvist, K. Hofmeister, B. Kristmannsson, Categorical clustering applied to the discovery of character builds in TCTD2: the BaT approach, in: 2020 IEEE Conference on Games (CoG), 2020, pp. 152–159.

[15] A. R. Da Silva, L. F. W. Goes, Hearthbot: An autonomous agent based on fuzzy art adaptive Neural Networks for the digital collectible card game Hearthstone, IEEE Transactions on Games 10 (2) (2017) 170–181.

[16] M. Świechowski, T. Tajmajer, A. Janusz, Improving Hearthstone AI by combining MCTS and supervised learning algorithms, in: 2018 IEEE Conference on Computational Intelligence and Games (CIG), IEEE, 2018, pp. 1–8.

[17] P. García-Sánchez, A. Tonda, A. M. Mora, G. Squillero, J. J. Merelo, Automated playtesting in collectible card games using Evolutionary Algorithms: A case study in Hearthstone, Knowledge-Based Systems 153 (2018) 133–146.

[18] A. Janusz, T. Tajmajer, M. Świechowski, Helping AI to play Hearthstone: AAIA'17 data mining challenge, in: 2017 Federated Conference

on Computer Science and Information Systems (FedCSIS), IEEE, 2017, pp. 121–125.

[19] Ł. Grad, Helping AI to play Hearthstone using Neural Networks, in: 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2017, pp. 131–134.

[20] A. Janusz, T. Tajmajer, M. Świechowski, Ł. Grad, J. Puczniewski, D. Ślkezak, Toward an intelligent HS deck advisor: Lessons learned from AAIA'18 Data Mining competition, in: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2018, pp. 189–192.

[21] J. Jakubik, A Neural Network approach to Hearthstone win rate prediction, in: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2018, pp. 185–188.

[22] J. Betley, A. Sztyber, A. Witkowski, Predicting winrate of Hearthstone decks using their archetypes, in: 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), 2018, pp. 193–196.

[23] M. Zuparic, D. Khuu, T. Zach, Information theory and player archetype choice in Hearthstone, Information Sciences 559 (2021) 236–250.

[24] M. C. Fontaine, S. Lee, L. B. Soros, F. De Mesentier Silva, J. Togelius, A. K. Hoover, Mapping Hearthstone deck spaces through MAP-Elites with sliding boundaries, in: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 161169.

[25] F. De Mesentier Silva, R. Canaan, S. Lee, M. C. Fontaine, J. Togelius, A. K. Hoover, Evolving the Hearthstone meta, in: 2019 IEEE Conference on Games (CoG), 2019, pp. 1–8.

[26] M. Eger, P. Sauma Chacón, Deck archetype prediction in Hearthstone, in: International Conference on the Foundations of Digital Games, FDG '20, Association for Computing Machinery, New York, NY, USA, 2020.

[27] A. Drachen, R. Sifa, C. Bauckhage, C. Thurau, Guns, swords and data: Clustering of player behavior in computer games in the wild, in:

2012 IEEE conference on Computational Intelligence and Games (CIG), IEEE, 2012, pp. 163–170.

[28] A. Dockhorn, R. Kruse, Predicting cards using a fuzzy multiset clustering of decks, International Journal of Computational Intelligence Systems 13 (2020) 1207–1217.

[29] G. Wallner, S. Kriglstein, Visualization-based analysis of gameplay data - a review of literature, Entertainment Computing 4 (3) (2013) 143 – 155.

[30] A. Drachen, M. Schubert, Spatial game analytics and visualization, in: 2013 IEEE Conference on Computational Intelligence in Games (CIG), 2013, pp. 1–8.

[31] P. Braun, A. Cuzzocrea, T. D. Keding, C. K. Leung, A. G. M. Pazdor, D. Sayson, Game data mining: Clustering and visualization of online game data in cyber-physical worlds, in: C. Z. et al. (Ed.), Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference KES-2017, Marseille, France, 6-8 September 2017, Vol. 112 of Procedia Computer Science, Elsevier, 2017, pp. 2259–2268.

[32] R. Davidson, D. Harel, Drawing graphs nicely using Simulated Annealing, ACM Transactions on Graphics (TOG) 15 (4) (1996) 301–331.

[33] K. P. Sinaga, M.-S. Yang, Unsupervised K-Means clustering algorithm, IEEE Access 8 (2020) 80716–80727. `doi:10.1109/ACCESS.2020.2988796`.

[34] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Physics reports 424 (4-5) (2006) 175–308.

[35] R. C. Prim, Shortest connection networks and some generalizations, The Bell System Technical Journal 36 (6) (1957) 1389–1401.

[36] P. J. Van Laarhoven, E. H. Aarts, Simulated Annealing, in: Simulated Annealing: Theory and applications, Springer, 1987, pp. 7–15.

[37] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, F. Herrera, SciMAT: A new science mapping analysis software tool, Journal of the American Society for Information Science and Technology 63 (8) (2012) 1609–1630.

[38] M. Ester, H. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: E. Simoudis, J. Han, U. M. Fayyad (Eds.), Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA, AAAI Press, 1996, pp. 226–231.
URL `http://www.aaai.org/Library/KDD/1996/kdd96-037.php`

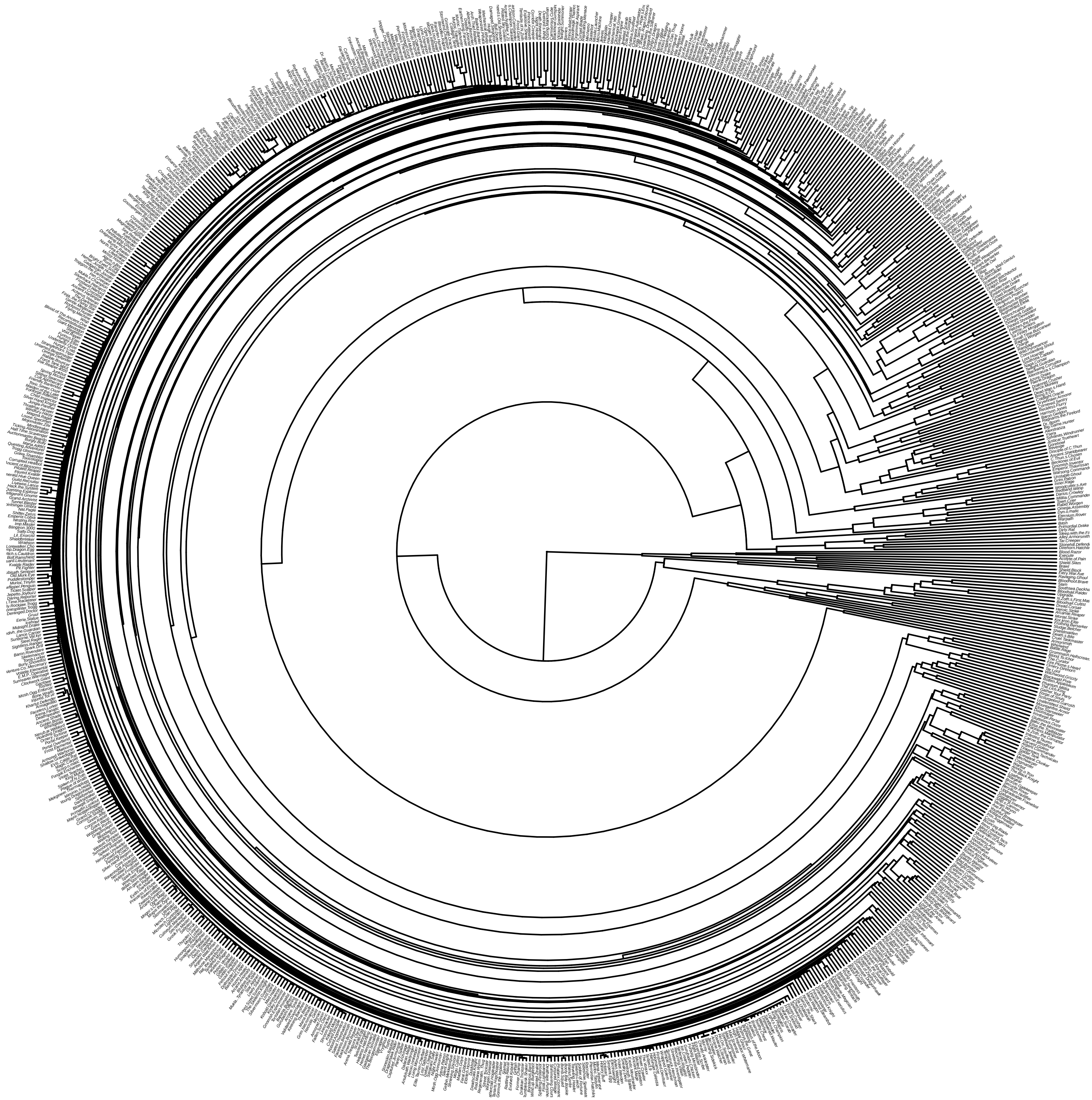


Figure 12: Global AHC for the most used 100 cards by decks belonging to the Warrior class.

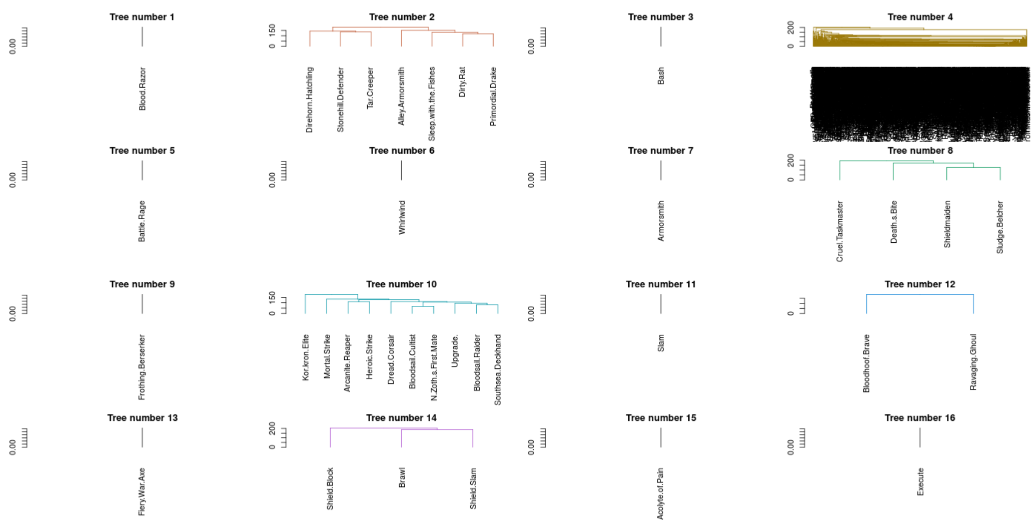
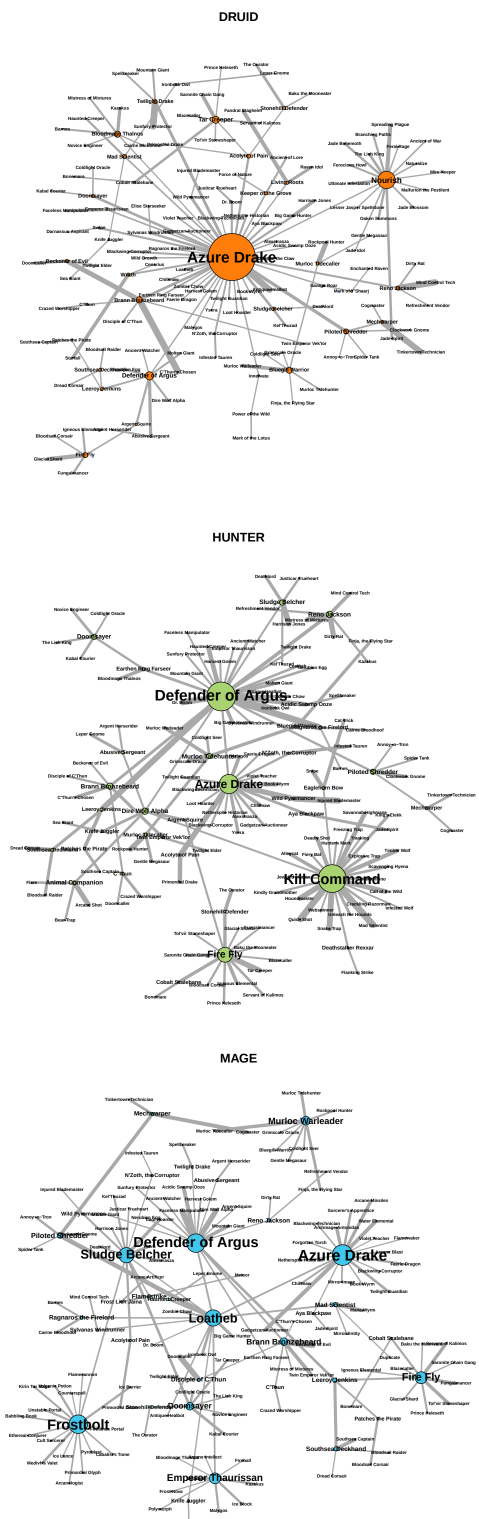


Figure 13: Subtrees of the AHC cluster. Each id correspond to a leaf (from left to right) of a binary tree of 4 levels. So 1 is more related to 2, and 3 to 4, and therefore 1-2 and 3-4 are also related.



38

Figure 14: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Druid, Hunter, and Mage hero classes. All the cards are considered (neutral and hero ones).

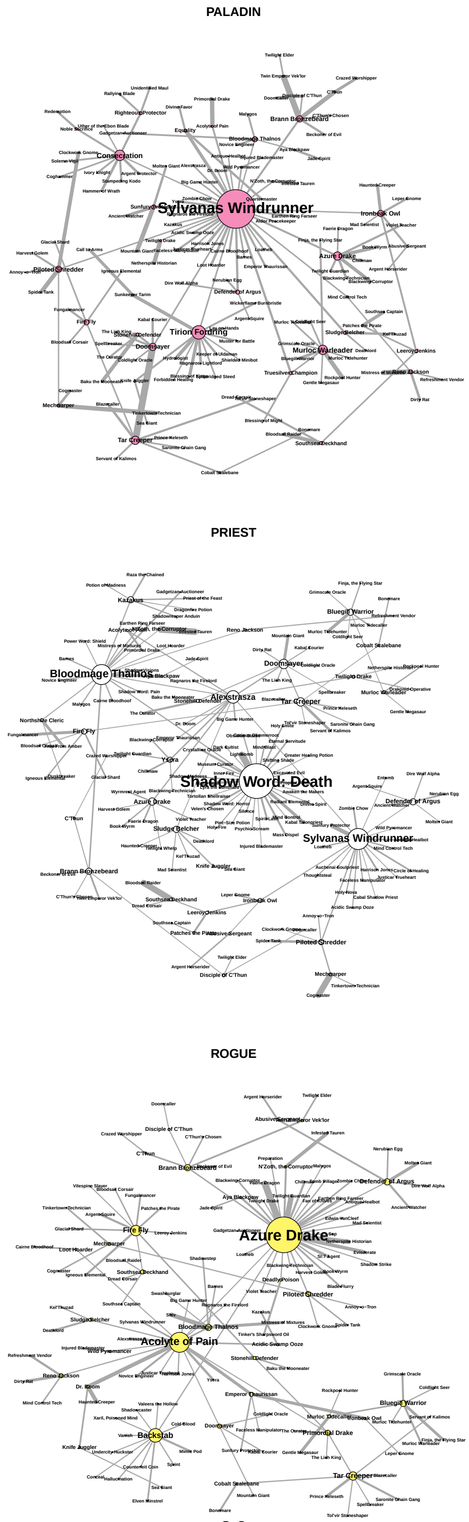


Figure 15: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Paladin, Priest, and Rogue hero classes. All the cards are considered (neutral and hero ones).

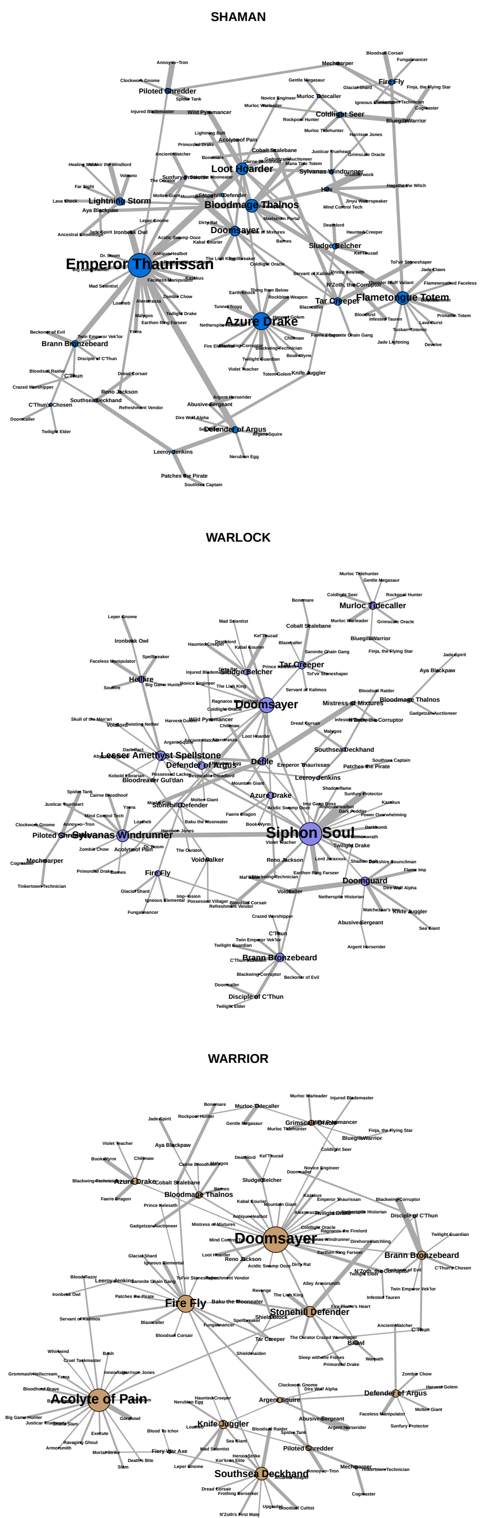


Figure 16: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Shaman, Warlock, and Warrior hero classes. All the cards are considered (neutral and hero ones).

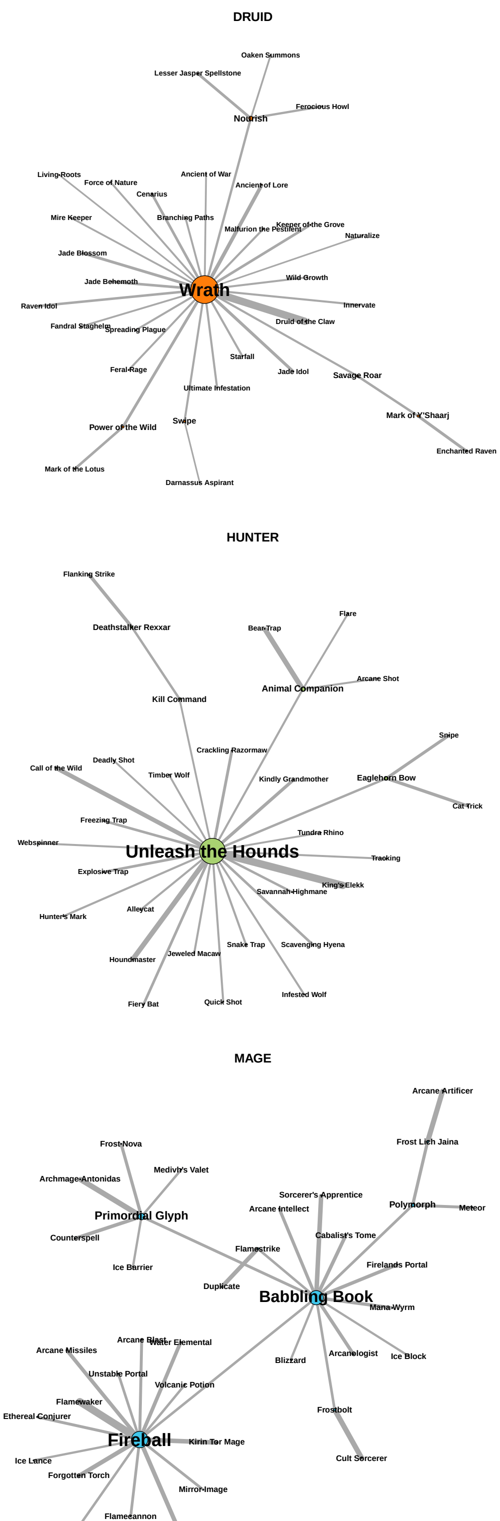


Figure 17: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Druid, Hunter, and Mage hero classes. Just exclusive hero cards are considered.

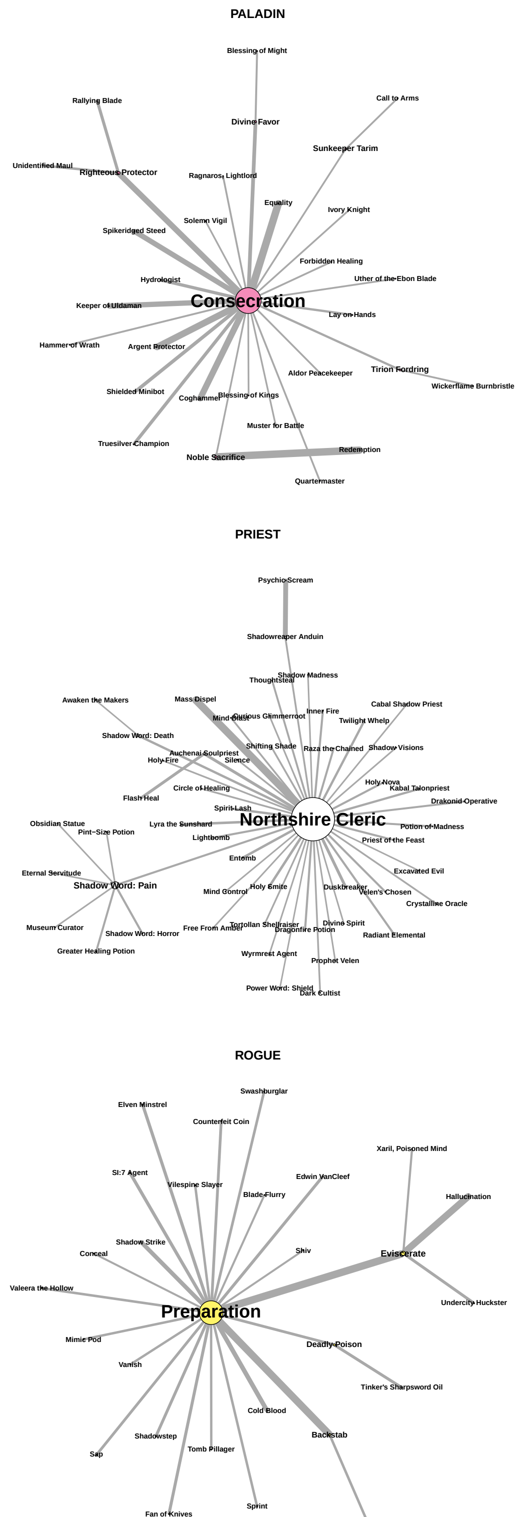


Figure 18: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Paladin, Priest, and Rogue hero classes. Just exclusive hero cards are considered.

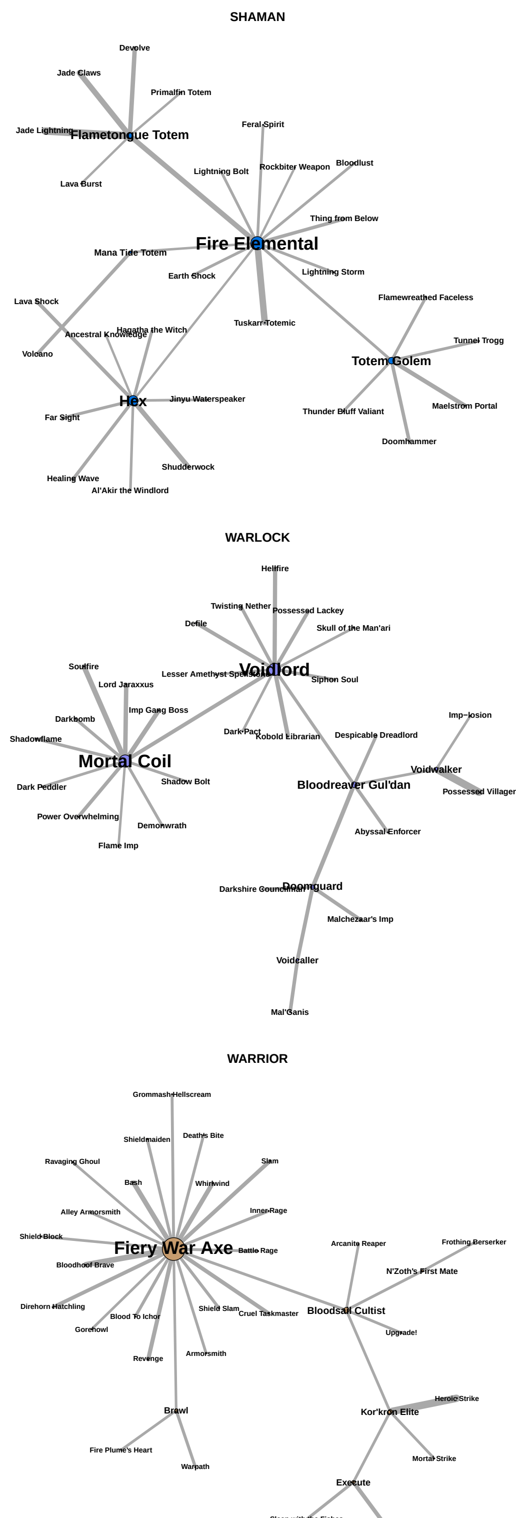


Figure 19: Minimum Spanning Tree with The Davidson-Harel layout algorithm for Shaman, Warlock, and Warrior hero classes. Just exclusive hero cards are considered.