Portfolio: The Social Web

Nizar El Mahjoubi Steven Raaijmakers Josh Willekes

March 2019

Abstract

This document introduces web concepts and the social web in particular. The portfolio firstly contains weekly summaries of lectures given by Davide Ceolin at the Vrije Universiteit. For this course, we imagined an application to tackle a social web problem. Online video game players get randomly assigned to other players when playing online. This presents an opportunity to link players from similar social circles together. We built this application and the app report treats the motivations and different steps in development. Finally, we conclude by stating individual contributions.

Contents

1 Introduction 3

I The Social Web 4

2 Introduction to the Social Web 4

3 What do people do on the social web? 6

4 What Data looks like on the social web 7

5 How do we mine, analyze and visualize the social web? 8

6 Personalization on the Social Web 10

7 Studying the social web 12

II Cats and Captions vs. Creators and the Clock 14

III App report: GameFriends 30

8 Theoretical background 30

9 Context 31

10 Methodology 31

11 Limitations 32

12 Conclusion 33

IV Individual Contributions To GameFriends 34

13 Nizar's contribution: background research 34

14 Steven's contribution: building the application 35

15 Josh's contribution: Quality Assurance 35

1 Introduction

In this portfolio we present the things we have learned during the course "The Social Web". The first few chapters will summarize the lectures of the course, the second chapter contains the slides we have used for our presentation of the paper "Cats and Captions vs. Creators and the Clock", a paper about comparing multimodal content to predict popularity, based on captioned images found on Reddit. In the third chapter we present an App Report about the application we have developed for this course, called GameFriends. The goal of this application is bringing together players of Fortnite using the Facebook and Fortnite APIs, which we think will improve gameplay experiences and expand social circles. Finally, in chapter IV you can find the individual contributions of each group member to this project.

Part I
The Social Web

2 Introduction to the Social Web

In the words of Steve Lohr [16], "digital technology is changing both how words and ideas are created and proliferate, and how they are studied." Digital technology is redefining the manner in which humans relate to each other. The set of social relations that link people through the World Wide Web are referred to as the Social Web. [11] Floridi et al. (2015) [18] name the following transformations which follow from the pervasiveness from ICT in our everyday lives: the blurring of the distinction between reality and virtuality, as well as between human, machine and nature. Furthermore, information has become abundant instead of traditionally scarce. Finally, Floridi et al. (2015) also mention the shift from primacy of entities to primacy of interactions.

The internet has spread across the world in impressive figures over the past years and decades. In January 2019, the total penetration of Internet Users around the globe was 57% which means 4.388 billion of the 7.676 billion people in the world, had access to internet. Of these 4.388 billion an astounding 3.484 billion were active social media users according to Statista [19]. The amount of internet users and active social media users went up by respectively 9% and 9.1% between January 2018 and January 2019. This means that if the rise of internet users in the world continues to grow by such rates, internet penetration will concern the entire globe sooner than later.

To understand the practices, implications, culture, meaning of the sites, as well as users' engagement with them, we study the Social Web.

2.1 Community based systems

Since the beginning of the World Wide Web, there have been stepping stones which led to the internet as we know it today. This subsection will first briefly explore and highlight this history. Then, there will also be an introduction to some characteristics and examples of community-based systems.

2.2 A brief walk through and highlights

The World Wide Web was created in 1989 [27] and was published publicly in 1991, date by which there were just above 2.6 million internet users (which was reached by 1990). Microsoft Hotmail, one of the most popular e-mailing services in the early age of internet, was launched in 1997. Google, the most popular search engine at present day, reached this status in 1998 (albeit in lesser form than the current day). By 1999, the number of connected computers to internet reached 70 million. Two years later, an online encyclopedia which we all know today was launched, namely Wikipedia. Social Networking websites were starting to prosper in 2003, year in which Linkedin, MySpace, WordPress and Skype were launched. A year later, Facebook was launched for Harvard University. The next year, Youtube was launched and another year later in 2006, Twitter was online. By 2011, Facebook had 550 million users over the world. Google made an attempt at a social network by launching Google+ in that same year.

2.3 The Big Ones

Characteristics of Community-Based Systems The community-based systems which dominate the web in many fields today rely on creators not necessarily hired by the company behind the website to create content for the the other users to consume, with synthesizers compiling the created content for the other Consumers. The ratio according to P. Bursilovsky [14] is of 1 Creator for 100 Consumers and 10 Synthesizers. The creators and synthesizers are referred to as Participators and the Consumers as Lurkers. Users are distinguished by their trust and reputation, which constitute their social capital in their social network.

Examples of Community-Based Systems: The Big Ones Wikipedia was launched in 2001 as a side-project of Nupedia. The latter dated from a year earlier and was a platform for free-content articles written by experts and founded by J. Wales and L. Sanger. To facilitate collaboration before the peer-reviewing process, they launched Wikipedia. By January 2019, Wikipedia counted 17,21 billion page views in a month [1]. Facebook was first launched for Harvard in 2004. The next year, Facebook included other universities, then high school students, after which professionals inside corporate networks, and finally everyone. This year also saw Facebook granting the ability to outside developers to build applications. This was formalized three years later in 2007 with the Facebook API, a platform consisting

of Facebook Markup Language (FBML) and Facebook Query Language (FQL). Three years later in 2010, Facebook went further by providing the Facebook Open Graph and in 2013 with the Facebook Graph Search. Jack Dorsey launched Twitter in the summer of 2006. The platform gathered 500 million users in 6 years. In fact, in 2012, 340 million tweets were posted daily. By the present day, the twitter users went down to 326 million users but the number of tweets went up to 500 million tweets daily. To give an other idea of the number of tweets, the Summer Olympics of London generated 150 million tweets in the span of 16 days. Google launched its social network on June 28, 2011 under the name Google+. A year later, the platform had 500 million users from which half (235 million) were active users. However, in 2018, a major security breach led to a data leak of its users. This was revealed by the guardian after which the platform closed in 2019.

2.4 Social Computing

Social computing presents an interdisciplinary study of the social structure in which technology puts power in communities, and emerging social the internet is set to be a platform for emerging social structures. The manifestos of social computing are diverse and Forrester Research (2008) [29] provides many examples among which social networks, blogs, wikis and user-generated-content. Social computing also leads to new tenents and implications which are briefly described below.

2.4.1 Tenents and implications of Social Computing

The main tenents of Social Computing C. Li (2006) [4] cites tree main tenets which arise from social computing. Innovation will shift from top-down to bottom-up, which places innovation in the hand of the users. Moreover, value will shift from ownership to experience. And finally, the power which the institutions currently have will transfer to the communities.

New means of communication Social computing has strengthened asynchronous messaging, beyond email and text-messaging. Between this messaging, a lot of communication is irrelevant and trivial, but some can be helpful and interesting. an example of this, are celebrities who use social networks to communicate with their audience. However, this means of communication is revealing itself to be addictive to some people, and especially teenagers.

New forms of communities Online communities have formed around a number of natural attributes of the members or any type of interest, hobby, or cause, where people can help one another with information and advice.

New forms of collaboration As a university student, code has to be shared in a collaborative environment and the articles we write have to be able to be accessed and edited by all. Github and Overleaf have made this possible for us.

New sources of knowledge Beyond what search-engines can provide, people can use their direct connections to answer questions (folklore knowledge) and tap into the knowledge of their friends by their news updates for example.

New sources of Entertainment Youtube, Google's video platform, is probably the most known new form of community based entertainment. However, people have accounts and multiple online platforms and visit them regularly.

New venue for Self-Expression People experience a large sense of self-satisfaction from being able to express themselves and help others on the social networks. Wikipedia is one of these examples.

New Economic models New collaborative economic models have been proposed in order to tap into this new form of community enabled by social computing. One of these economic models is the Collaborative Economy Honeycomb by the Web Strategist [20].

New Ethical issues However, social computing has also brought about new ethical issues of which privacy is probably the most talked about. Fake news and clickbaiting are also often cited. We are getting an idea of what we call echo chambers in which people get fed with reinforcing information about often biased views.

2.4.2 Categorizing Social Web Sites

Not all Social Websites are the same and it is possible to distinguish between them in different ways. This section will explore the two main categories: social Networking Sites and social Media Sites. Then, the social websites will be split according to culture, country and activity.

Categories Social Websites can be split into primarily Social Networking Sites which can be open or closed. These social networking sites can be general-purpose (such as Facebook or MySpace) or vertical (such as Couchsurfing or StackOverflow). The other type is the Social Media Sites which also can be open or closed. These can be further split into different websites according to their media types.

Differences through culture Especially in the past, Social Networking Sites were strongly split across cultures. The Netherlands used Hyves while France used Skyrock for example.

Differences between countries Nowadays, Facebook dominates the western world while V Kontakte is most widely spread in Russia and QZone in China.

Differences in activities Social Websites are also further distinguishable. BeautifulPeople, a dating website, employs restricted access to appear selective. UseNet and public discussion forms are structured by topics. Social Networking Sites are structured as personal networks with the individual at the center of their own community.

Social Sites features Social Websites often employ personal profiles to establish online connections and participate in online groups by communicating with its online connections. This way, the user can also share content and express his or her opinion.

Different ways of expressing opinions Members are able to leave comments on the content, voting by ranking (starts or upvote/downvote), or marking as 'favorite'. Sites use different ways to present and organize those comments (hierarchical, timestamping, counting, etc.). Useful but tricky: popularity web vs. quality web.

The social web plays a big part of society nowadays, since people spend a lot of time using the social web applications. One could ask what it is that keeps the attention of the people concerning the social web, which will be answered in the next section.

3 What do people do on the social web?

The social web consists of social interactions between people. The people we meet online are considered to be friends. However, there are some significant differences between the interactions we make with our friends online or offline. For example, the number of interactions you can have online is not limited by distance nor time, in contrast to real life interactions. Thereby, an online relationship can be "managed" by blocking the other person.

3.1 Multiple persona

Since there are multiple social systems available, people often have multiple accounts, each containing a digital representation of the user. A user's representation on Facebook (interactions with friends) most likely differs from its Tinder profile (dating focused) and its Linkedin profile (work related).

Instead of having multiple different accounts, it is also possible to store your entire profile with a locally trusted party. This happens for example with DigiD [1], which store your entire profile (as known by the government) in a database. Other parties may request parts of this profile for their specific uses.

Another example of this is OpenID, a decentralized authentication mechanism that allows users to log on to different services using the same digital identity.

OAuth enables users to grant access to their resources without having to share their password. We make use of this technique in the application we created and describe in part III of this portfolio. In the OAuth the app requests a key to the user, after which the user forwards this request to its provider, the provider grants access.

In the context of the social web, we can give multiple definitions to "sharing". We can share a link, or a photo, while we can also share our opinion in the form of a rating or a review.

3.2 Meme

An (internet) meme is a piece of digital content sharing common characteristics of content, form, and/or stance [26]. Some memes go viral, in which they are shared with a large audience. Memes are interesting to study since they reflect the current social landscape. Therefore memes can be used to spot issues, because satire is often linked to post-truth.

3.3 Social web issues

Within the social web there exists some issues, especially concerning privacy. However, the General Data

[1] DigiD: `www.digid.nl`
[2] GDPR: `gdpr-info.eu/`

Protection Regulation (GDPR) [2] is a big step ahead in solving these privacy issues.

Another subject of debate concerning privacy on the social web is the truthfulness of online profiles, which can be hard to verify. Thereafter, privacy is hard to understand since the terms and conditions can be easily misinterpreted if read at all.

There is a lack in awareness among users about the public nature the internet was initially designed for.

3.3.1 Bill of data rights

The Bill of (Data) Rights is published in 2007 asserts that all users of the social web have ownership of their own personal data, like the list of people they are connected to and the activity stream of content they produce. Also, sites that are supporting the Bill of Rights [3] have to gain control to their users of whether and how their personal information is shared with others.

3.3.2 General Data Protection Regulation

The GDPR is a regulation in EU law on data protection for all individuals within the EU. Its key points are, among other points, increased territorial scope, data portability, privacy by design and the right to be forgotten.

3.3.3 Right to be Forgotten

"The right to be forgotten arises from the desire of people to determine the development of their own lifes in an autonomous way without being stigmatized as a consequence of a specific action performed in the post" [17]. This is now a part of the Human Rights Law, however media companies are exempted from the law but search engines are not.

Again, this brings some issues. Most request to be "forgotten" come from citizens wanting to protect their personal information. However, it might be the cases that criminals abuse this law in their advantage.

3.3.4 Net neutrality

"Network neutrality is the principle that Internet service providers should treat all transmission of data over the Internet equally and not discriminate based on user, content, website, platform, application, type of equipment, or method of communication." [9]. Net neutrality is important to protect privacy and avoid discrimination.

3.3.5 Search neutrality

Search neutrality consists of the though that search engines (SE) should have no editorial policies other than that their results are solely based on relevance. SE should not exclude websites based on their own policy, penalizing for example competitors. However some exceptions are allowed, for example when adding in personalization to the system, which we will discuss in section 6.

People using the social web create a big stream of data which can be used in benign and malicious ways. The data created by users consists of all shapes and sizes and will be discussed in the next section.

4 What Data looks like on the social web

4.1 Blogs

The first type of data discussed in this lecture was Blogs. Back in the 1980's users used to post online diaries. Later users would use the same method to keep "diaries" on what results they found when they searched the web. Later users used this as a way not to only keep tabs on what they searched, but on other topics as well that kept said person busy, such as an educational blog of professors detailing their research or journalists reporting their findings. Dedicated software was developed with the intent of making it easier for more people to blog such as Wordpress. Blogs could be maintained by a single author or multiple and could have different types of media content such as photos, video (later called vlogs) or audio.

4.2 Wikis

Wikis were originally an online database meant for large scale editing by large groups of people. The most used softwarepackage for this is Mediawiki which is still used today in most notably Wikipedia. It is owned by the Wikimedia foundation which is a non-profit foundation with the goal of bringing information to as many people as possible for free, as they believe that information should be free and shared.

4.3 Crowdsourcing

One factor that makes wikis work so well is that a crowd is collectively invested in improving the available data. In the example of wikipedia this is done through the improving of wikipedia articles by adding more knowledge as well as media, such as images or

[3] Bill of (Data) Rights: `https://domainmarketresearch.com/?page_id=359`

sometimes sound. Crowdsourcing is a type of platform where you make use of the large number of people all willing to contribute a little in order to get things done. One of the first platforms to take advantage of this on the web was AmazonMechanicalTurk which allowed people to oursource their processes to other companies/people. Crowdsourcing should not be confused with Crowdfunding, which is very similar although Crowdfunding is purely about monetary contributions whereas Crowdsourcing can also be done through means other than money.

4.4 Structuring web

Over the years of the web growing people have seen the necessity to bring some order in to the web. Thus came the development of the so called "Semantic Web". The Semantic web is quite literally "Adding semantics to the web" by specifying terms within a domain in so called Ontologies. Similarly in order for interfaces or applications to interact with data on the web they need to know the semantics of the data. This is done by defining vocabularies within that domain that give information about the data and how it interacts with other pieces of data within that domain. An example of this is FOAF.

4.4.1 FOAF

FOAF or Friend of a Friend is an ontology that describes people, their activities and their relations to other people. It links people based on the information. In some ways it is similar to what Facebook does nowadays, only the relation within Facebook is only defined as "Friend" where in FOAF it could be any type of relation. FOAF has it's own RDF URI namespace. As an example we could have a foaf:Person with a foaf:givenName John and a foaf:familyName Doe. He could have a foaf:age 42.

4.4.2 SIOC

SIOC or Semantically-Interlinked Online Communities is an ontology for representing data from the Social Web using the Resource Description Framework (RDF). SIOC is mainly used to represent data found in message boards, forums and mailing lists. SIOC is often used in combination with FOAF to represent data found in profiles and in social networks.

4.4.3 Schema.org

Schema.org is a website which maintains vocabularies across the web. The vocabularies are built and maintained through community sites such as github and mailing lists. The schemas are used in multiple aspects of the semantic web such as email and web pages.

4.4.4 Activity Streams

Activity streams display data of Activities in a format of an Actor, followed by a verb, then an object and optionally a target. It displays information about someone or something interacting with an object in a certain way. E.g. John Doe returns a book to the library. John Doe being the actor, the book the object, returning the verb and the library the target. The activity stream vocabulary is made and maintained by the W3C consortium, an organization that has created and maintained many standards for the world wide web, examples including HTML and CSS. The combination of different verbs with different objects allows for mapping of activities based on the website and the combination.

4.4.5 Open Graph

OGP or the Open Graph protocol is a protocol originally developed by Facebook in order to give other pages on the web the same functionality as Facebook pages. This is mainly seen through their like and share button you see on many pages throughout the web.

4.4.6 Knowledge Graphs

Knowledge graphs are one of Google's main fields of interest as Google search is the most used search engine on the web. To create a knowledge graph they take objects, which could be anything from people to animals to buildings to objects and gather information about it, such as how old he/she/it is, what type of object it is, They can then connect these objects to other objects that are closely related due to the object, to show you these objects too when you search for the first object. In Google Search this often appears as a panel next to the results of your search query.

5 How do we mine, analyze and visualize the social web?

Data has become abundant since the advent of internet. Global data volume has increased substantially over the years and is increasing even more rapidly today. The global data volume which comprised 4.4 zettabytes in 2013 has been forecasted by Reinsel, Gantz and Rydning [24] to encompass 163 zettabytes

of data in 2025. This has lead to the rise of Big Data which allows for unique and precise insights into many different subjects through the use of data analysis and data science. This lecture aimed to introduce the concepts of big data and data science and their implications.

5.1 Big Data

Big data refers to earlier mentioned large volumes of data which can be analyzed and structured in order to derive different types of insights or to build products which use this data dynamically.

Data is mainly retrieved automatically or semi-automatically which allows for the large expansion of data. Because of this, data becomes overwhelming and there is too much volume to look at. Also, the number of dimensions increases which make big data difficult to look at from a variety perspective. Our increasing computational capabilities also allow for faster retrieval of data which account for its velocity. Finally, data is not necessarily valid and not all data can be trusted. This is referred to as its veracity.

5.1.1 Data abundance

This much data everywhere which can provide for insights across many different areas of our societies and users. In fact, insights can be gained across different domains which will be highlighted shortly.

5.1.2 Big Data applications

Because of the abundance of data, it is important to be critical of the information which is available on the internet. This is not only true for texts but also for data visualizations which have become widespread on news websites. The same criteria as for academic papers can be applied to big data (source critiscm), such as looking at the credentials of the author, the data of the publication, the bias of the source, the sources of the source, and so on.

Not only information but also websites have to be viewed critically. In fact, because of the ease of creation and hosting of a website, some websites have bad intentions with their visitors. There are software solutions which can help with this task such as WOT (Web Of Trust).

5.1.3 Users

Big data has allowed for many participants of society to use big data for various reasons. Everyone is able to look up a certain trend with Google Trends or google.org. Furthermore, specialists such as historians and media scholars also have much to gain from analyzing big data. Data visualizations of often complex matters have become widespread in today's largest journals online. Large sets of data can be instantly understood by a large number of people if the visualization is appealing and informative enough. Science as well has had a lot of gain from Big Data. Commercial companies have also been one of the leading users of Big Data because of the value of the available information. Big Data allows companies to gain unique insights into processes and motivations of their or all consumers. A last example of users of Big Data is governments and politicians who can gain unique insights from the citizens and what they value. A recent example of this is the Trump 2016 campaign during which Big Data was employed by a company named Cambridge Analytica to exploit the private social media activity of a huge swath of the American electorate, developing techniques that underpinned its work on President Trump's campaign in 2016 according to the New York Times [25].

This has lead to a new industry which profits from analyzing available data for different purposes (such as those described above).

5.2 Data Science

Shah, Horne en Capellà [3] coined the term "big judgement" to describe the expertise skills necessary to complement statistical and programming skills, in order to make sense of the large amount of data available. In fact, data science uses scientific methods, programming skills and domain expertise to gain insights from data.

5.2.1 Data scientists

A data scientist is uses data science to extract value from data. This person mines large volumes of structured or semi-structured data and makes sense of the data using statistical analysis and artificial intelligence techniques.

5.2.2 Data products

Many companies have revolved their business model around providing a seemingly free product in exchange for data from their users. They then employ different techniques such as advertising to profit from this data. Some of these companies (e.g. Facebook) are amongst the most valuable companies in the world. Users are here in a feedback loop in which they provide information about the use of a certain

data product, which then is applied to improve the data product.

Owning a large data set is not enough to extract the valuable information necessary to attain certain goals. The data has to be structured such that any analysis is possible. Furthermore, it is not necessary to own such a platform to have access to sets of data on the internet. Data scientists employ data mining to extract valuable information from available data.

5.2.3 Data mining

Data mining refers to the organizing and analyzing of large volumes of data in order to extract concrete insights about the data. Data mining can be done on all types of data, of which most is not publicly available. However, there are options which allow everyone to mine data.

5.2.4 Data mining software: LikeMine

Jin et al. [12] describe a tool which allowed them to mine social media data. Huge amounts of data from social websites are available online and can be mined to extract valuable insights. In their study, the authors propose LikeMiner which allows to mine for the power of likes to understand the interests of users at a given time for example.

Data mining algorithms Different algorithms exist to analyze data. Some of these algorithms allow for the clustering of the data into distinct groups (k-means algorithm or support vector machines). Data algorithms such as Apriori learn association rules between variables in a database. Google developed and built its website on PageRank which allows for the classification of the links which they have in their database.

Data mining and journalism Journalists also have access to all this data and some journalists specialize in the mining of data and putting it together in a coherent story which may appeal to the masses. An example of data journalism is an infographic which is a very simplified and understandable graphical visualization of data which allows for quick presentation of information.

5.2.5 Mining the social web

The social web allows for unique understanding of users and society both at the individual as at population levels. The social web can be mined to obtain big data from user-generated content with the aim to derive insights and form conclusions about users.

Twitter for example allows for live brand sentiment display for brands which are enough talked about on Twitter through the classification of texts to certain sentiments. Furthermore, Sentiment analysis refers to the analysis of text to extract data about how a text is perceived (positive/negative).

We have described how big data evolves and its applications on the web. In fact, the social web is a goldmine for data miners. We will discuss the characteristics and applications of social web data in the next section.

6 Personalization on the Social Web

Nowadays systems are using recommendations to personalize the environment for their users. In order to create these, systems need to build user models to provide user specific recommendations. User models are created using user data, which can be obtained in multiple ways. For example, the system could ask the user explicitly for this data, while another way could be the system deriving the date itself. Once sufficient data is gathered, the system has to interpreted the data in such a way that it can be exploited.

6.1 Data gathering

We distinguish different types of data:

- Usage data: logs about clicks on links or time spent on certain pages
- Social data: (hash)tags and comments
- Documents: pictures or videos
- Personal data: affiliations and locations

It is often the case that once a user is new to a system the available user data is sparse. There exists multiple ways to solve this sparsity problem, for example semantic enrichment.

Data in combination with a user model creates a user profile.

6.2 User profile

A user profile is typically a data structure in which a specific characterization of the user is stored. The profile represents the user in a certain context which is relevant to the system: a user profile from one system can be irrelevant to other systems.

6.2.1 User model

A user model (UM) is classified as the recipe for obtaining and interpreting user profiles. This boils down to the UM being a set of definitions of the interpretations of certain observations concerning a user. These observations must be translated into characteristics to create user profiles. The process of representing a user is called user modeling.

6.2.2 Overlay User Modeling

Overlay user models were originally used to classify the knowledge of students, based on domain concepts and hypotheses. These hypotheses concern the user's knowledge about different concepts in relation to an expert's knowledge.

6.2.3 Customizing

When we talk customizing user models, the user itself explicitly provides the elements for its own user model and is able to adjust these elements.

6.2.4 User Model Elicitation

User model elicitation is a form of interactive user modeling, in which the system asks questions and observe the user's answers. Based on the observations the system can create a user model. The data of the answers can be extracted in different ways, for example by using Natural Language Processing [2]. It is also possible to observe the user via logs. These logs concern the behavior of the user on which the system can draw conclusions, based on machine learning and clustering.

6.2.5 User stereotypes

A user model can also be build by using stereotypes. Using this technique we use a set of characteristics to describe a group of users. A user can be assigned to multiple stereotypes.

Based on certain user data the system assigns a probability of user being part of a stereotype. Each stereotype contains a set of attributes and a importance value for this attributes.

For example, a users IP address can be located at a university, which can give this user a probability of 0.6 being part of a stereotype: professor. Now we can make some assumptions about professors: they are interested in economy, culture and politics.

6.3 User Modeling Building Blocks

In this section we will discuss how a user model can be created by analyzing a tweets of a Twitter user.

1. **Temporal constraints**. First off, we must define which set of the tweets should be analyzed and therewith take into account the time period of the tweets to recognize temporal patterns. For example, someone with a nine to five job might be tweeting about work related issues during his time at work, but in the weekend may tweet a lot about sports.

2. **Profile type**. Next we need to determine what type of concepts should represent the interest of a user. When analyzing a tweet, we can look into multiple things that can represent the users interest. For example, we can take into account the entity, the topic or the hashtags of a tweet.

3. **Semantic enrichment**. The tweets can be enriched in the semantic thus creating more data that can be analyzed. This can be achieved by either enriching the links or the context of the tweet.

4. **Weighting scheme**. Not all concepts are as important, thus we must define the weight of each concept. A way to determine these weights is using TF-IDF [22].

6.4 User adaption

User adaption is adapting the system to a user to improve the application functionality and the user experience.

Within user adaption there two perspectives concerning its success:

1. Consumer perspective: in this perspective the adaption is focused on maximizing the satisfaction of the user. However, this is hard to measure.

2. Provider perspective: in this perspective the focus is on adapting the system in such a way that profit is maximized. Yet, the influence of the user model (and thus the personalization) on the profit is hard to measure.

There are multiple strategies to evaluate your adaptions:

1. By taking user studies, in which you ask a group of users whether you did a good job.

2. Log analysis, in which you analyze the click-data and infer your success.

3. Evaluation of user modelling:

    - By measure the quality of the UM directly by comparing your new model with an existing *correct* model. You can also let people judge the UM you generated for them.
    - Applying the user modeling strategies in a recommender system.

6.4.1 Issues in User-Adaptive Systems

Overfitting can be an issue in user-adaptive system. In this case the system may adapt too strong to a user in which, for example, a user only sees things that are relevant to him, instead of also being able to see novel things. The key is to find the right balance between relevance and novelty.

The "Lost in hyperspace"-problem is a recurring problem in user adaption. In this case, for example, the re-ordering of certain menu items may lead to confusion for a user.

6.4.2 Recommendation Systems

Recommendation system (RecSys) predict relevant, useful and interesting items for a user in a given context.

Within recommendation systems we have different categories, one being *collaborative filtering*. This can be achieved in different ways:

- Memory based (user - item): in which the preferences of user A is stored, whereafter the system shows the preference of user A to user B, which is supposed to be a similar user.
- Model based (item - item): based on the items user A likes, the system shows similar items to user A.
- Model based (clustering): the users are clustered based on their preferences. Then recommend items to these clusters.
- Model based (bayesian): determine probabilities that user will like item B given that it likes item A. The probabilities can be based on other users preferences.

We can collaborative filter people based on their similarity of interests

We now discussed almost all aspects of the social web. The social web is also interesting for (behavioral) scientists to study, since the social web is reflection of todays society. Therefore in the next section we will discuss the study of the social web.

7 Studying the social web

The World Wide web is one of the most influential pieces of internet technology in history. Over 10 billion people use the web with more than 100 billion web documents produced on it. Developing the web is not just a single task process. In order to develop the web not only must we construct new uses for the web, but we also need to study and understand what happens on the web in order for new developments to be made. But what is the web? Many people would call the web a thing, when really it's more of a construct of societal activity in creating interlinked content. Data Analysis The web consists of large, very large amounts of data. Extracting knowledge from said data is important for creating understanding of the web. For this we use the 3 Vs of Big Data. First we have Volume which refers to the amounts of data, the amount of objects the size of the data. Next Variety, so what different types of data are there, where do these different data types come from, what is it being used for? Finally Velocity, how often is that data being generated? A social media website such as Twitter or Facebook generates millions of Tweets or photo's each day, whilst a search engine such as Google produces Billions of searches per day. In the age of fake news a new V has also popped up: Veracity. Can the data that has been generated be trusted or is it all fake? All these Vs emphasize the importance of Data Analysis on the web.

7.1 Web Development

When developing new software applications for the web, software is usually designed by having a specific use in mind using the appropriate technology. This is then tested on a small scale to test for microscale properties. However, users have different ways of interacting with the developed applications and these interactions are of much larger interest to analyse, since these macrosystem level challenges do not occur at microscale level.

7.1.1 Search Engines

An example of the above can be seen in search engines. At a microscale level we designed software that allowed us to rank documents based on our search queries. At some point this resulted in people gaming these search algorithms in order to have their pages/documents found first at a Macro level. At some point we found out that the algorithm was being gamed, so we had to research this behaviour in order to create new technologies that no longer allowed for the gaming of the system. In the end we

settled with a middle ground where SEO (Search Engine Optimization) is legal and people can game the system in order for their files to be found when it would want to be found.

7.2 The Web Graph

The World Wide Web can be interpreted as a huge graph consisting of nodes and edges. The nodes in this case are the different webpages whereas the edges between the nodes are the hyperlinks on each webpage that link one page to the other and vice-versa. After analysing this graph it gave insight into how fast the web was growing whenever you add a new page or new link on the web (similar to adding a node or edge to a graph). This same graph is also used by search algorithms to determine which pages contain the most useful information. Google Pagerank (named after one of the founders of Google Larry Page) assume that whenever a hyperlink links to a page that that page holds some authority over the information that is being searched for. According to Google 20-25% of all searches have not been seen before, meaning that every search generates a new HTML page on Google and therefore a new node in the Web graph.

7.3 Collective Intelligence

Like said before the Web does not only contain a technological aspect but also a social aspect. A website that exemplifies this aspect is Wikipedia. Wikipedia was built on Mediawiki software but it was not the first or only application to do so. Wikipedia however became the fifth most used website on the internet, way more than any other website using the same software. Thus part of Wikipedia's success must lie in the social aspect of the website. When asking why people use Wikipedia the reasons can vary. Some users enjoy writing for Wikipedia whereas other users contribute to Wikipedia from the ideology that information should be free or that they feel importance in helping others.

7.4 Social Machines

The applications on the web are an early example of social machines as we know them in the real world as well. The problem with the web Social machines is that they are largely isolated from each other, therefore not being able to interact with each other. Society has already shown that social machines can interlink in the real world, so there is no reason why this should not be able on the web. For this technologies need to be developed that allow users to create, share and adapt social machines, after which interlinking between different social machines could become a possibility.

7.5 Networks

As mentioned earlier the web is one big Network consisting of linked smaller networks of different types and sizes. Studying these networks gives us new insights on how the web connects the world. There are multiple network types that emphasize a different type of knowledge provided by said network. Social networks focus not on what you know but who you know, Cognitive Social networks not on who you know but who they think you know and knowledge networks not on what you know but what they think you know.

7.6 Web Science

To conclude on Web Science. Web Science is not just computer science but rather an intersection of Computer Science with other scientific disciplines such as, but not limited to, law, media, sociology and economics. There are many different parts in society, each with their own roles and objectives within society. Therefore the web needs to be able to be used by these different groups for different objectives. To further develop the Social Web we must understand the different ways these groups use the web to achieve their objectives.

Part II
Cats and Captions vs. Creators and the Clock

In this section we summarize the paper "Cats and Captions vs. Creators and the Clock: Multimodal Content to Context in Predicting Relative Popularity". In this paper the authors propose a model that can predict out of two post inputs which one will become more popular.

Cats and Captions vs. Creators and the Clock:

Comparing Multimodal Content to Context in Predicting Relative Popularity

Group 22: Josh, Nizar & Steven

Introduction

"Cats and Captions vs. Creators and the Clock:
Comparing Multimodal Content to Context in Predicting Relative Popularity"

Hessel, Lee and Minmo - Cornell University - March 2017

- Multimodal content = mix of text, audio, image or video.
- Focus on content factors only.

Data

- Dataset:
  - Tan & Lee dataset: 106 million **Reddit** posts (2007 - 2014).
  - Focus on posts with image + caption.
  - Image-centric subreddits:
    i. r/pics
    ii. r/aww
    iii. r/cats
    iv. r/MakeupAddiction
    v. r/FoodPorn
    vi. r/RedditLaquiristas
- Popularity = # upvotes - # downvotes.

Data - Pairs

Pair-wise sampling:

- Posted to same community (place)
- *Posted within approx. 30 seconds of each other (time)*
- Popularity(post x) + 20 >= Popularity(post y) (noise)
- Popularity(post x) = 2 x Popularity(post y) (noise)

	# Pairs
pics	44K
aww	33K
cats	15K
MakeupAddiction	10K
FoodPorn	9K
RedditLaquiristas	8K

Data - Time

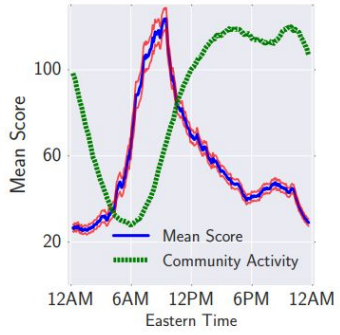
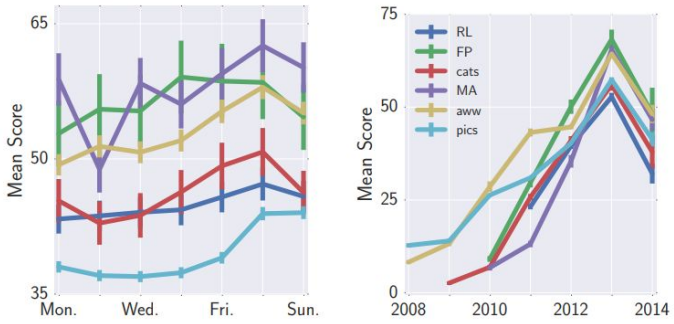


Figure 3: Average score versus time of day (eastern) on aww with 95% CI (red) and activity levels.



(a) Day of week versus score

(b) Year versus score

Figure 4: Relationship between various measures of time and eventual submission score with 95% confidence intervals.

Data - Validation

Human validation:

	aww	pics	cats	MA	FP	RL
Humans	60.0	63.6	59.6	62.2	72.7	67.2

Table 3: Human annotation accuracy results.

Model Design

- Data structure: {x1i, x2i, yi} n i=1

$$\hat{y}_i = w^T (f(x_{1i}) - f(x_{2i}))$$

Cats and Captions (Visual & Textual)

Visual Models

- Color
- Histogram of oriented Gradients
- GIST descriptors
- VGG-19
- ResNet50

Textual models

- Structural
- Unigram
- Topic based
- Deep Averaging Network (DAN)
- Long Term Short Memory (LTSM)
- Bi-LTSM

Creators and the Clock (Users and Timing)

User

- Type
- Activity
- Quality

Timing

- Random
- Earlier
- Time

Results (Unimodal)

		aww	pics	cats	MA	FP	RL
Timing	Random	50.0	50.0	50.0	50.0	50.0	50.0
	Earlier	*51.7*	*51.1*	*49.9*	48.9	*48.6*	48.7
	Time	50.2	50.2	*50.7*	*50.4*	*49.7*	*50.6*
User	Type	50.6	51.2	50.7	52.8	51.8	56.1
	Activity	51.1	53.6	*52.8*	55.0	53.9	60.6
	Quality	*54.7*	*55.5*	*52.9*	*60.7*	*55.5*	**67.3**
Textual	Struct	56.2	54.8	56.5	50.9	52.3	52.5
	Topic	55.2	55.8	56.8	60.4	55.2	55.5
	DAN	58.6	*58.3*	58.5	*62.2*	*57.6*	59.8
	LSTM	*59.4*	*58.8*	*58.7*	61.0	*57.0*	59.1
	Bi-LSTM	*59.7*	*58.9*	*59.3*	*61.8*	*57.8*	*59.6*
	Unigram	*59.7*	*58.6*	*59.5*	*63.0*	*57.6*	*60.8*
Visual	HOG	51.7	52.8	51.9	53.5	53.5	53.5
	GIST	52.7	53.0	53.5	55.9	56.5	56.3
	ColorHist	55.3	53.7	55.6	55.0	56.5	54.5
	VGG-19	63.4	58.9	61.1	62.4	62.8	62.1
	ResNet50	**64.8**	**60.0**	**62.6**	**64.9**	**65.2**	*64.2*

Results analysis:

- All models have higher accuracy than Random
- Timing: Pairwise ranking controls for time
- User: Previous quality predicts current quality
- Textual: simpler is better
- Visual: Convolutional Neural Networks have highest accuracy.

Results (Multimodal)

	aww	pics	cats	MA	FP	RL
Time + User	54.1	54.7	52.1	58.8	54.2	64.8
All User	56.3	55.3	54.6	60.9	56.0	**68.4**
ResNet50	64.8	60.0	62.6	64.9	*65.2*	64.2
Text + Image	**67.1**	**62.7**	**65.9**	**67.7**	**65.8**	66.4

Multimodal accuracy results

- Content features outperform user features
- Performance increase when combining text + image

	aww	pics	cats	MA	FP	RL
Time + User	55.5	51.7	52.6	56.9	52.8	60.5
All User	60.4	51.0	54.3	**63.1**	57.9	**66.0**
Text + Image	**65.5**	**66.0**	**67.3**	62.7	**62.6**	65.4

Out of domain accuracy results

Analysis of Aww (Image only)

- Lighting is important
- Animal taking up the entire image
- Top 10 images all dogs

Analysis of Aww (Text only)

Found this 3 day old baby under a car in...
This is Dexter. A year 1/2 ago my friend...
My very first dog and best friend is...
This sweet girl was found on the side of...
Found this little guy getting rocks...
Every time we start the boat...
Henry is quite the lady killer
Reddit, meet Sutton! My new kitty baby!
Cloudy forgets to close his mouth...
First post! My begging English Mastiff.
Soft kitty, warm kitty, little ball of...
My sleepy kitty enjoying the sun
Happy kitty, sleepy kitty and man she...
Shih Tzu + Beagle = Adorable
GO! GO! GO! GO! GO! GO! GO!...

- Longer captions
- Have a story
- Most popular unigrams:
  - Saved *(β = .50)*
  - Wife *(β = .45)*
  - Roommate *(β = .42)*
  - Cancer *(β = .41)*

Analysis of Aww (Multimodal)

Earlier mentioned conclusions are confirmed.

Conclusion

“Multimodal content generally improves performance”

Part III
App report: GameFriends

Introduction

Video games have allowed for multi-player gaming since their conception in the early eighties. What started as an awkward experience with bad visuals has become an immersive experience in which players from the entire world play together. Video game platforms draw millions of players each day to compete or team up together. More than ever, video games have also become a social activity.

Because of the worldwide component of the most popular video games, players come across players from all over the world. In a team context, this can be frustrating to players since they do not necessarily speak the same language. Thus leading to difficulty in communicating. Furthermore, the random allocation of team members in some online games lead to players getting set up with players for below their skill level which can lead to more frustration.

We imagined an application which allows for teaming up with friends from near social circles in order to improve game-play experience as well as avoid social isolation in everyday life as a result of playing games. Because of the academic value of our application, we sought to answer the following research question:

Is there a need to and can we connect players from similar social circles to team up for multiplayer games?

We first sought to understand whether the need existed to connect players through their social circles. Then, we built an application which allows for players to connect with players from their social circle. The use case we imagined was Fortnite players connecting with eachother through Facebook, thus leading to the sub-question:

Can we use Fortnite and Facebook to connect players from similar social circles to team up for multiplayer games?

In this report we try to find an answer to our research question using literature while also creating an application that fills the demands proposed. The app will show your Facebook friends who also play Fortnite and are willing to team up. Finally, we conclude that there is indeed a need to connect players and think that our application can fill this need.

8 Theoretical background

To put the application in academic perspective, we sought to understand whether there is a benefit for the gamers to play with gamers from their offline social circles. So we looked first at online video gaming effects on social circles and health. Second, the application has to be attractive for gamers to use it so we looked for literature which highlight the effects of playing with offline contacts in an online environment on the overall gameplay experience.

8.1 Improving social circles

Mental and physical health is influenced with by offline social circles. This effect is greater for adolescents [28]. There is thus an important health incentive to improve offline social circles. In fact, [5] found that online video gaming led to a decline in the size of a player's social circles as well as the player's offline social circles. The players were found to substitute their contacts in real life with their virtual friends. Adolescents and adults were both reported to make great sacrifices to play online video games, sometimes privileging gaming to meeting with friends, family and/or their partner [10]. This effect was strengthened for adolescent gamers who could privilege gaming to sleep and/or work and education.

Social online gameplay was found by Kowert, Domahidia, Festl, and Quandta [15] may lead to smaller offline circles. Moreover, the social circles were also found to be of lesser quality. The most reliable predictor of social support was found to be social proximity. In this sense, playing with friends was found to lead to increased social support. Furthermore, since players sought out players to be friends with, this increases the need for social contact online as well as offline. In the context of the application which we imagined, this confirms that players which know each other offline are more inclined to engage in social support - strengthening the mental health aspect.

These findings brought us to consider our application as being beneficial to social circles, and thus the physical and mental health of the players. Then, we sought to understand whether the gameplay improved whilst playing with users which were not randomly assigned.

8.1.1 Improving gameplay experience

Most multiplayer online games engage in some form of random allocation. If a player has no friends online, they are assigned to random players. This can lead to many frustrations, of which two are presented next.

First, because of the global aspect of the most popular online video gaming, players sometimes get assigned to players from other countries. This can lead to communication problems if both players don't speak the same language. Second, the skill level of the players may not be of the same skill-level, leading to frustration for the more skilled players.

Knowledge of the skills of other players is a predictor of positive group experience, as demonstrated by Kaye and Bryce [13]. de Kort, Ijsselstein and Gajadhar [7] found that gameplay experience was improved when a social aspect was added to the game.

This confirms our hunches and sets an academic context to our application. Thus, we now discuss how the application was build and how we applied our findings to the application.

9 Context

The year 2018 has been the year of online-gaming, with one game in particular: Fortnite. This game has gather over 200 million players worldwide since November 2018 [4], and is still growing. Fortnite consist of multiple game-modes, with Battle Royale being it most popular one.

In the Battle Royale mode 100 players are dropped on an island. The reachable part of the island is constantly shrinking due to a deadly storm, therefore narrowing the playing field. A victory can be achieved by remaining the last player on this island, which can be achieved by eliminating other people on this island with weapons and items you can find on the ground.

In Fortnite's Battle Royale one has the possibility to play either by yourself, or with one or multiple partners (up to a maximum of four). Fortnite is considered the most fun when playing with a partner, however the situation may arise that you are not able to find a *suitable partner*.

Fortnite offers you to play with any other random player that is also willing to play Battle Royale at the same moment. Often this match is a poor decision since the probability of the other random player being able to communicate is not high. Besides, the other random player simply might be on another skill level, making it more difficult to play together.

The most *suitable partner* therefore would be a partner that is able to communicate with you through voice, and is on the same skill level.

Our project aims to find this best case scenario partner using a combination of the Facebook API and the Fortnite API.

10 Methodology

The complete source code of the application can be found on `https://github.com/sjraaijmakers/gamefriends`. *This repository also contains a screencast of the application.*

We choose to build our application as a standalone application on the web, since we feel this is the most accessible. We created the website using NodeJS [5] and used Mongoose [6] to access our MongoDB [7].

What we had in mind for our application was the user to login using his/her Facebook account. For the first login the user has to manually register its appropriate Fortnite usernames. Thereafter, the application shows the friends and the friends-of-friends who also granted access to the application, together with their Fortnite statistics. However, during the build process of the application we experienced a lot of limitations, which we will touch upon in section 11.

10.1 APIs

In our application we use the Facebook API and a Fortnite API. In the application we send GET-requests to the respective server thus using their API. The servers replies with a JSON object, containing the information we requested for.

10.1.1 Facebook API

We choose to make use of the Facebook API since this is a rather simple way to obtain the digital representation of someones social circle. Using the Facebook API we can show the friends and the friends-of-friends of a person, provided that they granted access to our application. We use the Facebook API as a user login, using OAuth [8].

The login process consists of the user requesting Facebook for an access token which is sent back to our

[4] `www.statista.com/statistics/746230/fortnite-players`
[5] NodeJS: `https://nodejs.org/en/`
[6] Mongoose: `https://mongoosejs.com/`
[7] MongoDB `https://www.mongodb.com/`
[8] OAuth: `https://oauth.net/`

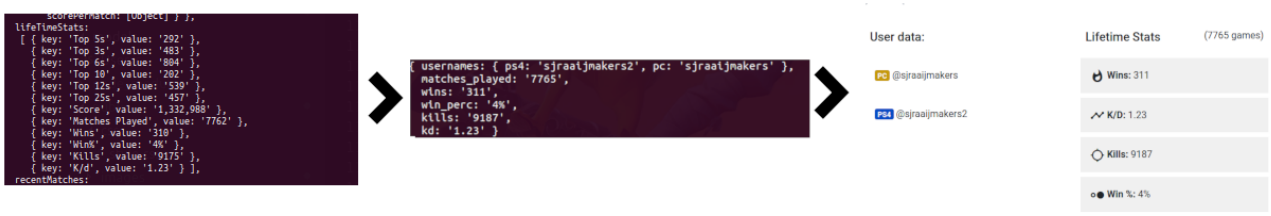


Figure 1: Data flow: on the left side the response of the FortniteTracker API. In the middle the cleaned JSON object. On the right side the final output on the website.

website. This access token has to be send in the headers of every GET-request we make to the Facebook API, and allows for extra security.

10.1.2 Fortnite API

Since Fortnite has no official API, we make use of the API that is provided by FortniteTracker [9]. This API however brings some limitations which are discussed in section 11. Using the FortniteTracker API, we determine the skill level of users by showing multiple statistics. We choose to display the so called K/D-ratio, which stands for the "kill-death"-ratio. This simply is the number of total kills against the total times you have lost. Next to the K/D, we show the amount a player has won a game, the total amount of kills and the win percentage. The JSON object received by the FortniteTracker API contains a large amount of keys, of which a lot is irrelevant to our application. The part that we are interested in is stored under the key "lifetimestats". We narrow down the values to generate a JSON object that contains all relevant information. Finally, this object is send to the view to generate the user profile. This flow is illustrated in Fig. 1.

10.2 Database

We wire together the information of both APIs in our database. In the database we save the Facebook user ID and the corresponding Fortnite usernames. An example can be found in Fig. 2.

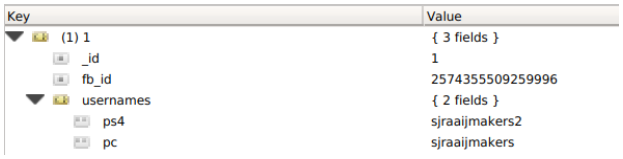


Figure 2: Database document for a user.

Whenever a user logs in we can retrieve the other information based on either the *fb_id* or the *usernames* using the appropriate GET-request. For example, if we want to know the full name of the user we can send a GET-request to the Facebook API. If we want to know the Fortnite stats we send a GET-request to the FortniteTracker API.

10.3 Front-end

The front end of this application was created using a combination of Bootstrap [10] for the CSS and PUG [11] for the HTML. We use Express [12] to create the web server and the routes of the website.

11 Limitations

11.1 Facebook API

During the process of creating the application we experienced some major downfalls, most of them concerning the Facebook API. The main goal of using the Facebook application is because it contains a digital representation of the social circle of its users. However, with API 2.0+ it is only possible to retrieve friends that give permission to our application. To retrieve the someones Facebook friends (*user_friends* [13]) we also need Facebook to verify our application. Once Facebook has verified the application it is possible to ask the users of our app for permission to re-

[9] Fortnitetracker API: `https://fortnitetracker.com/site-api`
[10] Bootstrap: `getbootstrap.com/`
[11] Pug: `pugjs.org/api/getting-started.html`
[12] Express: `https://expressjs.com/`
[13] Facebook user_friends: `https://developers.facebook.com/docs/graph-api/reference/user/friends/`

trieve their friends. The application however did not get verified, thus it was not possible to actually show the friends or friends of friends within the application. A solution to this would be Facebook verifying our application. For now, we created static object in our database.

Another problem with the Facebook API is the callback mechanism which is restricted to a URL using the HTTPS protocol. In order to use the HTTPS protocol on our own server we would need a SSL certificate, which is something we could not manage within the scope of this project. However, for testing purposes the API allows us to use a regular HTTP url, provided that it is hosted on the localhost [14].

11.2 Fortnite API

The FortniteTracker API is limited to one GET-request per 2 seconds. Thus meaning the application would not work well once a lot of users are using the site simultaneously. However, we can fix this by making the data shown semi-dynamic. We can achieve this making a queue system in which we store all the usernames assigned to our application. We would then order it in such a way that all

However, this can be fixed by storing the response of the requests in our database instead of sending a request every time a user views player data. Meanwhile we would create a queue system that send a request every two seconds to keep the data in the database up-to-date. Again, this would not work on a very large scale.

Finally, it is not possible to verify if a certain user is indeed the Fortnite player he/she claims to be. This can be fixed by letting the user send a friend-request to a account we created with a token in the message.

12 Conclusion

To conclude we strongly believe that bringing people of a similar skill level and those who are already friends together provides tangible benefits to players of online games. By allowing players to sign up to Gamefriends they can match and play with their friends they know from. Not only that but online players would be improving their social circles. There were however some limitations to our implementation of this application, mainly due to the APIs that we used. The Facebook API only allows us to retrieve information after they have reviewed our application and the user using the application has to give permission for our application to gather information about friends and friends of friends. In addition the Fortnite API had a limit of one request per 2 seconds, thus if a large group of people were to use the application at the same time, it would not work. Overall we think this application could enhance online gaming quality and build and/or improve social circles.

[14] Localhost: `https://en.wikipedia.org/wiki/Localhost`

Part IV
Individual Contributions To GameFriends

13 Nizar's contribution: background research

13.1 Introduction

My task within the application project was to find sufficient academic support and context for the application which we imagined. It was thus important to find studies which supported our hunches about online video gaming. The project that we imagined is an application which brings players together from similar social circles. This to achieve two main goals: improving their social circles and improving their gameplay experience.

Online video gaming has been booming over the last two decades and some games nowadays attract several million players each month from all over the world. The players can get set up randomly together which may lead to disconnect with the offline world, and especially the social circles of the players.

Firstly, we went to look for academic proof that online video gaming actually was correlated with smaller and lower quality offline social circles. Then, to put our application in perspective, we sought out to look whether online video gaming indeed improved offline social contacts when gamed together. Finally, the need to enhance the overall gameplay experience is highlighted.

13.2 Background Research

This section first highlights the importance of improving social circle and academic support to back this statement. Then, we have a look at how gameplay experience can be improved by linking up players from the same social circles.

13.2.1 Improving social circles

It is commonly known that offline social circles are important for human, and specifically adolescent's physical and mental health. This was confirmed by a study done by Voorhees et al. [28] in 2005. Moreover, social circles outside of the online sphere also have shown to reduce deviant behaviour.[6]. This demonstrates the importance of maintaining offline contacts and improving social circles.

It was found by Cole and Griffiths [5], as well as Shen and Williams[23], that online video gaming led to a decline in the size and quality of the player's offline social circles. The same studies showed that players engaging in online video gaming tend to replace their offline social contacts with the online contacts.

Griffiths, Davies and Chapell [10] asked players "what they sacrificed to play the game". It was reported that both adolescent and adult players state that they commonly sacrifice socializing with their offline social contacts in order to play. Especially for the adolescent gamers, the study shows that they agree more often to sacrifice sleep and work/education to play games.

Kowert, Domahidia, Festl, and Quandta [15] confirm these findings in their study. In fact, they found that social online game play corresponds with smaller social circles of a lower quality. This indicates that online video game players have less close friends, and that the friends that they have are of less emotional support than individuals engaging in less or no online video gaming.

Social proximity according to Trepte, Reinecke, Juechems found that social proximity is the most reliable predictor of social support. Thus, by diminishing the distance between the players, we can improve their social support. Duchenaut et al. [8] noted that if games allow for the opportunity to relax and socialize, the players may tend to increase their social support.

Kaye and Bryce [13] found that gamers preferred other gamers to non-gamers. This means that adults and adolescents who engage in video gaming make friends with people who do the same. For adolescents, it was found that this effect was positive across the second term (ending of the high-school year), but not during the first term. This is because, it takes time for adolescents to get to know each other and discover that they both enjoy gaming

These studies unanimously confirm that social circles decline with online gaming, thus presenting risks to lead to mental and physical health issues. Adolescents especially are at risk to sacrifice their physical health (e.g. sleep deprivation) or social contacts to play online. Also, it was found that social proximity has a direct effect on social contact with other players offline.

13.2.2 Improving gameplay experience

Piirainen-Marsh [21] found that communication problems can arise from online video gaming with not sim-

ilar native speakers. Additionally, the study examines Finnish online players with English native speakers, but one can imagine that for other countries bilingual literacy is less common - Finnish people are commonly known to have a good command of the English language. This study also demonstrated that bilingual skills actually improved from playing with speakers from other countries when the initial communication barrier was overcome. In our application, there is a trade-off between forcefully (by random assignment) becoming familiar with different languages and thus become acquainted with foreign languages, and sharing the same language, thus having no linguistic communication problems to start with.

Kaye and Brye [13] looked at how social processes in gameplay influence the individual and group experiences of players. They found that social gaming experiences were considered as fun, regardless of collaboration in the game. Feeling of social belonging and interactions with friends are also highlighted in this context. Thus, adding social aspects enhances the emotional experiences felt during the activity.

Moreover, knowledge of the skills of other players were also predictors for positive group experience. de Kort, Ijsselstein and Gajadhar [7] found similar results to Kaye and Brye [13] in that increased social presence in cooperative, particularly in the online sphere, leads to better gameplay experience. This suggest that the social component leads to better gameplay experience than without it.

13.3 Conclusion

In light of aforementioned studies, it was thus found that online videogaming may lead to decreased physical and mental health as a results of smaller social circles which are of lesser quality. However, by linking up players who know each other in the offline sphere, online video gaming can have a positive effect on offline social contacts.

It was also found that the gameplay experience enhances whilst playing with people which stem from the offline social circles. In fact, the players experienced better communication and thrill from gaming as well as improved their relationships by playing together online.

The players thus have an incentive from a health perspective, but also to improve their social contacts and have an overall better gaming experience. These findings confirm our hunches and entail us to believe our application would have a positive effect on the online video gaming community, and online video game players in general.

14 Steven's contribution: building the application

14.1 Introduction

Within the application project, my contribution consists of the idea of the application and the execution of it.

I came up with this idea since I had never played online videogames much but was surprised once I started playing Fortnite, by the amount of new people I met through this game. The new people I met are mostly real-life friends of my own real-life friends who also play Fortnite. In this way I got introduced to a lot of friends of friends, some of them I have even met a few times in real life. This felt like a win-win situation because I met new people whom I interact with both online and offline. Therefore, when we got the ability to make a application in a group concerning a social web problem, I proposed to my group members a way to find even more friends through video games, focussing on Fortnite for the sake of time.

14.2 Limitations

The application created has been build solely by me. Building the application I experienced some mayor flaws, mostly concerning the Facebook API. Because of the restrictions of the API the application could not function perfectly within the scope of the project. This is also due to me having no webprogramming skills when starting with this project.

14.3 Conclusion

Concluding, I can say that in advance the aim of our application was to high. Maybe it would have been better to stick to a Twitter analysis, since we know this was possible within the scope of our project. Also, due to my lack of webprogramming skills I experienced a lot of problems creating this application. However, personally I am very interested in this application myself and will probably finish and polish the application in my spare time to optimze the user experience.

15 Josh's contribution: Quality Assurance

My contribution has been of that of a group member who was willing to put in the work where necessary. Besides writing my share of this portfolio and doing my part in the presentations. During the creation of

the project I was one to often ask critical questions which we as a group considered before moving one with a decision within the project.

Finally, when trying to get a working version of our online application I tried hosting it on a VPS and link a SSL certificate to the web-address in order for the application to be accessible online after this project, since the Facebook API which the app uses . This however did not work within the scope of the project sadly.

References

[1] URL: https://stats.wikimedia.org/v2/#/all-projects.

[2] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[3] Shvetank ShahAndrew HorneJaime Capellá. *Good Data Won't Guarantee Good Decisions*. 2014. URL: https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions.

[4] Chris Charron, Jaap Favier, Charlene Li, et al. "Social computing: How networks erode institutional power, and what to do about it". In: *Forrester Customer Report* (2006).

[5] Helena Cole and Mark D Griffiths. "Social interactions in massively multiplayer online role-playing gamers". In: *Cyberpsychology & behavior* 10.4 (2007), pp. 575–583.

[6] Robert Crosnoe, Kristan Glasgow Erickson, and Sanford M Dornbusch. "Protective functions of family relationships and school factors on the deviant behavior of adolescent boys and girls: Reducing the impact of risky friendships". In: *Youth & Society* 33.4 (2002), pp. 515–544.

[7] Yvonne AW De Kort, Wijnand A IJsselsteijn, and Brian J Gajadhar. "People, Places, and Play: A research framework for digital game experience in a socio-spatial context." In: *DiGRA Conference*. 2007.

[8] Nicolas Ducheneaut et al. "Alone together?: exploring the social dynamics of massively multiplayer online games". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM. 2006, pp. 407–416.

[9] Angele A Gilroy. "Access to broadband networks: The net neutrality debate". In: Congressional Research Service, Library of Congress. 2009.

[10] Mark D Griffiths, Mark NO Davies, and Darren Chappell. "Online computer gaming: a comparison of adolescent and adult gamers". In: *Journal of adolescence* 27.1 (2004), pp. 87–96.

[11] Harry Halpin and Mischa Tuffield. "A standards-based, open and privacy-aware social web". In: *W3C. W3C Incubator Group Report* (2010).

[12] Xin Jin et al. "LikeMiner: a system for mining the power of'like'in social media networks". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2011, pp. 753–756.

[13] Linda K Kaye and Jo Bryce. "Putting the fun factor into gaming: The influence of social contexts on the experiences of playing videogames". In: *International Journal of Internet Science* 7.1 (2012), pp. 24–38.

[14] Heather McCreery Kellert. "Mining and Crafting Mathematics: Designing a Model for Embedding Educational Tasks in Video Games". PhD thesis. The Ohio State University, 2018.

[15] Rachel Kowert et al. "Social gaming, lonely life? The impact of digital game play on adolescents' social circles". In: *Computers in human behavior* 36 (2014), pp. 385–390.

[16] Steve Lohr. *The Origins of 'Big Data': An Etymological Detective Story*. 2013. URL: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/.

[17] Alessandro Mantelero. "The EU Proposal for a General Data Protection Regulation and the roots of the 'right to be forgotten'". In: *Computer Law & Security Review* 29.3 (2013), pp. 229–235.

[18] Brent Daniel Mittelstadt and Luciano Floridi. "The ethics of big data: current and foreseeable issues in biomedical contexts". In: *Science and engineering ethics* 22.2 (2016), pp. 303–341.

[19] *Number of social media users worldwide 2010-2021*. URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

[20] Jeremiah Owyang. "Collaborative economy honeycomb". In: *Creative commons* (2014).

[21] Arja Piirainen-Marsh. "Bilingual practices and the social organisation of video gaming activities". In: *Journal of Pragmatics* 42.11 (2010), pp. 3012–3030.

[22] Juan Ramos et al. "Using tf-idf to determine word relevance in document queries". In: *Proceedings of the first instructional conference on machine learning*. Vol. 242. 2003, pp. 133–142.

[23] Rabindra A Ratan et al. "Schmoozing and smiting: Trust, social institutions, and communication patterns in an MMOG". In: *Journal of Computer-Mediated Communication* 16.1 (2010), pp. 93–114.

[24] David Reinsel, John Gantz, and John Rydning. "Data age 2025: The evolution of data to life-critical". In: *Don't Focus on Big Data* (2017).

[25] Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr. *How Trump Consultants Exploited the Facebook Data of Millions*. 2018. URL: https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

[26] Limor Shifman. *MIT press essential knowledge: Memes in digital culture*. 2013.

[27] *Social media in numbers*. URL: https://www.cendrinemedia.com/Blog/Resources/Social-media-numbers?fbclid=IwAR21wgttaeNptngqSW9Mq7aOah3AcAS_mOUPx2wLdWvD6tiWBmXGtiCV8A4.

[28] Carolyn C Voorhees et al. "The role of peer social network factors and physical activity in adolescent girls". In: *American journal of health behavior* 29.2 (2005), pp. 183–190.

[29] Fei-Yue Wang et al. "Social computing: From social informatics to social intelligence". In: *IEEE Intelligent systems* 22.2 (2007), pp. 79–83.