UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management and Social Sciences

Evaluation of Reputation in the Context of Online Social Communities

Masterthesis HFE
Jule Landwehr
University of Twente – Enschede
April 2020

Supervisors:
Prof. Dr. Frank van der Velde
Dr. Martin Schmettow
Mieszko Czyzyk

In Cooperation with:

open social

THX

Abstract

Uncertain credibility and reliability are a severe problem in the social online community space. They often cause member inactivity and or endangerment. A system that evaluates individuals' reputation online and displays it on their profile can help solve this problem. In this study, we (1) investigate the concept of reputation in order to find a set of constructs, that organisations can use to develop different reputation systems and (2) introduce two reputation system categories. This study consists of three parts. Firstly, a Word Association was conducted (n= 61) to find words associated with reputation. Secondly, the words are used in a Pilot Card Sorting (n= 30) to elicit users' mental models of reputation. The mental models give an overview of possible constructs and their underlying structure. The results suggest that there are at least two categories of reputation systems automated and peer to peer. Thirdly, a follow-up Card Sorting was conducted for both categories separately (both n = 31). The results are presented in a heatmap and a dendrogram based on a hierarchical cluster analysis. Combining the obtained clusters from both (heatmap and dendrogram) into a tentative cluster structure results in 6 constructs with subconstructs for each reputation system category. *Activeness, Activity, Network, Engagement, Commonness,* and *Content* for automated reputation and *Credibility, Behaviour, Sociability, Irresponsible/Provoking, Reliable, Confidence* and *Positive Influence* for peer to peer reputation. Based on these set of constructs we introduced two reputation systems one focussing on automated reputation and one focussing on peer to peer reputation. The obtained set of constructs can serve as a basis for developing systems that evaluate the reputation of members in online social communities.

Table of Contents

Preface

This master thesis was written in collaboration with Open Social. Open Social is a company that helps organisations build and maintain their online social communities by providing them with an online community platform and community management tools. To make sure the customers can create thriving sustainable communities, Open Social continuously develops new software features and improvements. These software features have three main goals (1) help community managers to maintain their community, (2) enhance the safety of community members (3) and motivate members to participate in the community actively. A big problem for most of their platforms is user-generated spam and fake news, which is hard to detect and can irritate and endanger users. To help fight spam, fake news and enhance member activity and safety, the company is working on a tool that evaluates members online. For the company, it was important to use a scientific approach for the development of this tool. Therefore, this research will suggest a system that can be used to evaluate members of online social communities. First, literature was researched to find out if there already is a system that evaluates individuals online. The results suggest that similar reputation systems do exist but only for e-commerce communities. Therefore, this study aims to develop a reputation system specifically for online communities by identifying the constructs that are involved in measuring reputation and how they can be used as a basis for an evaluation system tool. In short, the following study (1) describes the process of finding structured constructs of reputation for the evaluation tool and (2) introduces two categories of reputation systems for online social communities. The next section draws a broader picture of the problem at hand, consequences and possible solutions.

1. Introduction

In today's society, online communities have become indispensable. Almost everybody is a member of at least one online community, may it be electronic commerce (e-commerce) like eBay and Amazon or social networks like Facebook or LinkedIn (Perrin, 2015; Schrammel et al, 2009). Online interactions and activities often go far beyond staying connected with friends and family. The online space makes it possible to not only connect and exchange information locally (with family and friends) but with people all over the globe (Resnick et al., 2000; Zacharia & Maes, 2000). Connecting globally online can be especially important for individuals that represent a minority in their local communities. The advantage is that the variability of online social communities enables individuals to find at least one community that reflects their interests and values and helps them feel like they belong (Ulusu, 2010).

Besides bringing great possibilities, online social communities can pose a threat for both the individual participating and the organisation behind the community (Fan et al., 2005; Resnick et al., 2000). Most of the time, individuals do not know each other in real life beforehand (Zacharia & Maes, 2000). That can be an advantage if a person does not want his or her real identity to be shared, but it also makes a person vulnerable to fraud, fake news and internet bullying. Hence, knowing who to trust is especially important (Fan et al., 2005).

In order to find out whom to trust, the individual needs to rely on the little personal information available online (Resnick et al., 2000). Unfortunately, it is not certain whether this information is reliable since information over background, character and trustworthiness of members is often missing, which makes it impossible for members to evaluate each other (Yao et al., 2009; Fan et al., 2005). As a consequence, it is tempting to commit fraud, as there are no reputational consequences. In fact, the number of internet fraud is rising (Ba, 2001; Fan et al., 2005). Thus, the danger of being cheated by somebody else is relatively high. For instance, one person can use a fake identity online to abuse or fraud others (catfish). It can be used for romantic scams, trolling or financial gain (Ba, 2001).

Two consequences of uncertain credibility and reliability are (1) hesitance of users to get actively involved in communities and (2) user endangerment (Yao et al., 2009). The reasons why people are hesitant to be active are that users or people around them either do not want to take any risk being an active part of the community to prevent being scammed, or they are overwhelmed by fake news and spam. Young users are especially vulnerable to fraud and scams and active use of chats and forums can put them in physical and or psychological danger. Often, these users still read content and comments but never actively participate (Dellarocas, 2010a;

Zacharia & Maes, 2000). Hence, the problem does not only limit the individual's possibilities but also can be problematic for organisations behind those communities. Organisations need member engagement for a thriving running community (Bishop, 2007; Falor et al., 2014). A lot of online communities fail or face problems due to lack of user engagement or user-generated spam and fake news. For them, it is difficult to detect fake accounts, fake news and getting rid of all the spam because they do not know who is credible and who is not. Thus, the problem affects both organisations and users alike.

In the past, some research was done on solving the issue of uncertain trustworthiness and credibility, with the main focus on finding solutions for e-commerce communities (Yu & Singh, 2002; Xiong & Liu, 2003). E-commerce communities are websites like eBay, where individuals can buy products from different private sellers (Zacharia & Maes, 2000). One solution that has been introduced in different e-commerce communities is the implementation of reputation systems. For example, on eBay or Amazon, buyers and sellers can ask questions, rate each other's products and vote on the quality of a review (Dellarocas, 2010a). In that way, sellers can build up a reputation, either good or bad, on which basis buyers can build an opinion and decide whom they can trust to make transactions with (Jensen et al., 2002; Bishop, 2007). Likewise, a reputation system for online social communities could solve the problem of trust as it did for e-commerce communities. It can help to evaluate members without needing to know their whole personal and private background, protecting the individual's privacy. (Zacharia & Maes, 2000). This way, trust can be built between users without them needing to reveal their real identity.

Unfortunately, little research has been done towards finding a fitting solution for online social communities, even though research is needed to ensure a thriving and safe online space (Ba, 2001). A reputation system designed primarily for online social communities might be the right solution to gain trust and build thriving spaces online. Therefore, the primary goal of this study is to create a reputation system that can be used to evaluate members of online social communities. However, to be able to design that system, we first need to know which constructs are involved in the evaluation of reputation. Therefore, this study will begin by (1) finding and studying structured constructs of reputation that can be used as a basis for an evaluation tool for online social communities and (2) develop a reputation system based on the found constructs. The research question is: *'Can a set of constructs be found that could be used as the basis for developing a reputation system?'*

For that purpose, we are going to do literature research where we first take a look at why a reputation system specially designed for online social communities might be important. We do this to make sure that reputation is the right concept to be used to solve the problem of uncertain credibility and reliability. Secondly, existing reputation systems are going to be examined, to get a bigger picture of how a reputation system possibly could look like. Next, reputation as a concept is going to be investigated, to get to know valuable input on what reputation essentially means, what underlying constructs might be important and how reputation can be measured online. This includes looking at how people use reputation in the offline world. Word Association and Card Sorting are used to find underlying structured constructs of reputation that can be used later on for the evaluation system tool. The constructs found in the Card Sorting studies are interpreted further in order to form general constructs and subconstructs that organisations can use as a starting point for the design of different reputation systems.

2. Literature Research

In the following, the studied topics and results of the literature research are discussed shortly. The main research topics of the literature research were (1) Why reputation systems are important, (2) What existing reputation systems are there, (3) What the concept reputation is and how to measure it.

2.1. Why Reputation Systems for Social Online Communities are Important

Offline we use reputation every day to evaluate others. In fact, it plays a vital role in our society (Borderless Technology Corp [BTC], 2018; Kawamichi et al., 2013). Individuals either work on their reputation to lead a successful life or use reputation to decide whether it is safe to get in contact with other individuals. It helps us to choose our friends and whom to trust (Izuma, 2012). Moreover, reputation makes people countable for their actions. If somebody misbehaves, he will get a bad reputation sooner or later, and others will behave towards that person accordingly (BTC, 2018).

Online it is challenging to evaluate the reputation of another individual. Most online profiles contain little to no information on who the person is in real life, and even if there is information, it cannot be validated (Zacharia & Maes, 2000). Because of that, organisations cannot readily use online reputation to evaluate others. As a consequence, users are tempted to behave in a bad manner. However, introducing a reputation system could help overcome this problem. Organisations can use it to evaluate the reputation of an individual and displaying it on the user's profile without needing to know private information. Thus, a user-based reputation system focusing on generating a reputation for individual users could be one solution for solving the problem of uncertain trustworthiness and reliability (Dellarocas, 2010b).

2.2. Existing Reputation Systems

Since the first introduction of reputation systems as an online evaluation tool, different reputation systems have been developed and integrated to fit business objectives and needs of users of different online communities. Jensen et al. (2002) grouped the existing types into three different reputation system categories: *Ranking Systems, Rating Systems and Collaborative Filtering Systems. Ranking Systems* analyse users' behaviour to achieve a ranking. That could be, for example, how often a user visits a website or how long this person is a member of the community. *Rating Systems* use evaluations like stars given by other users, to compute an average. These ratings are sorted the same for every user, so user preferences are not taken into account. *Collaborative Filtering Systems* work the same as Rating Systems, but additionally, take users' preferences into account. For instance, if somebody is buying a product online and looks at the customer reviews, he will see the ones that are most relevant for him first.

Until now, the introduced categories for reputation systems first and foremost include reputation systems for transaction-based communities, like e-commerce, where there is a seller versus buyer relationship. In this context, reputation stands for the quality of the service given and the product received (Zacharia & Maes, 2000). Likewise, most of the research is done on fitting reputation systems for electronic commerce communities. However, as mentioned before not only electronic commerce communities deal with the problem of uncertain credibility and reliability but also online social communities. Thus, they could also profit from an integrated reputation system (Dellarocas, 2010a; Bishop 2007).

Unfortunately, systems developed for e-commerce cannot be readily integrated into online social communities. The reason is that reputation in the context of e-commerce does represent something else than reputation in online social communities. Social networks want users to connect and become active in a safe space (Bishop, 2007; Falor et al., 2014). Here the focus lies mainly on measuring the reputation of individuals. E-commerce, on the other hand, wants to build trust between buyers and sellers to perform transactions. In order to do that, reputation is measured based on the quality of the given service and product. Thus, systems developed explicitly for e-commerce do not measure any personal information and are based on non-personal characteristics. To summarise this, both types of online communities' deal with the same problem. Nevertheless, they ask for different solutions.

Jensen et al. (2002) introduced a new reputation system category, namely Peer-Based Systems, that focuses mainly on social communities. The main idea behind the peer-based reputation system category is that in real life people often fall back on friends or family recommendations when making decisions, for example, which series to watch or which person to trust. Jensen et al. (2002) argued that recommendations can be used in the online space too. They introduced two systems for the peer-based category — implicit and explicit peer-based reputation systems. The implicit peer-based system uses the behaviour of the 'friend' of a user as data for ratings. By detecting what users' friends do, the system generates recommendations. The explicit peer-based system weights and filters ratings according to whom the user knows and trusts, which makes ratings relevant for the user.

The introduced peer-based reputation systems are meant to be useful for more social-oriented situations like online social communities (Jensen et al., 2002). However, in order for these systems to work, it is assumed that users have friends they actually know and trust. As mentioned earlier, most users do not know each other in real life beforehand and do not know whether they can trust each other. Thus, peer-based reputation systems do not tackle the problem of uncertain credibility and reliability and can only work in communities where users already know each other in real life. Consequently, another system needs to be integrated additionally to the peer to peer one. A system that keeps into account that most people online are anonymous and do not know each other. This new system might either fit in one of the existing categories or a new category.

All in all, a peer-based reputation could be one part of a successfully integrated social online reputation system, but another system should be introduced additionally. A system that focuses more on an individual's reputation to tackle the problem of uncertain credibility and reliability. To do so, first, a clear picture needs to be drawn on what reputation is, and how it can be measured.

2.3. The Concept Reputation

In order to build a reputation system, it is crucial to get a better picture of what reputation means in different contexts and what underlying constructs there might be. This will be explored further in the following.

2.3.1. The Definition of Reputation in Different Contexts

In the *offline* world, reputation is defined as opinions held about somebody based on past behaviour and characteristics (Montes et al., 2017). In other words, reputation is what an individual is known for. That can be, for example, their extraordinary talent or their noble character. The individual can influence this reputation by how he presents himself in front of others (BTC, 2018). Individuals can build up a reputation over time. This reputation can either be good or bad. Others use their knowledge of another person's reputation to assess that person. Therefore, reputation can have a major impact on where a person stands in society. As a result, people generally want to get a good reputation and keep it (BTC, 2018, Izuma et al., 2014). For instance, if somebody is known for their good reputation, they have a better chance of making friends and finding a job position. In contrast, if somebody is known for impulsive and aggressive behaviour, then they might be avoided by others. As a consequence, we will behave in a certain way when interacting with others in order to build or keep a good reputation (Izuma, 2012).

Online, reputation is seen as somebody's significant actions taken in the online community, which are displayed to the user in a way that another user can evaluate the individual (Dellarocas, 2010a). For instance, in e-commerce, the reputation of a seller is displayed by showing other uses stars on products and services at the seller's profile. Before buying a product, most buyers will look at the rating of the seller. If the star rating is low customers will be hesitant to buy from this seller resulting in low selling numbers Therefore, just like in the offline world sellers will seek to build up and uphold a good reputation. In the context of online social communities, this would mean that, if a user behaves appropriately, it will be displayed on their profile and they can build up a good reputation in the community. In contrast, when the person behaves in a wrong way, for example, by trolling others, they can get a bad reputation, and people will keep their distance. Consequently, people will strive to build up a good reputation and work towards keeping it by behaving accordingly.

Comparing both definitions of reputation (online and offline), it stands out that both online and offline people seek to (1) build up a good reputation if there are consequences and (2) evaluate others by accessing their reputation. However, people assess reputation online and offline differently. Offline we combine everything that we know about one person, for example, their beliefs, behaviour and opinions, in order to evaluate their reputation as good or bad. That automatically happens in our brain (Carbo et al., 2003; Izuma, 2012). In contrast, online, we need to depend on a third instance, like a reputation system. This system should give us information readily about that person, for example, their actions and behaviours in the community, so that we can evaluate that person.

2.3.2. Reputation an Abstract Concept

Like trust or creativity, reputation is an abstract concept (Barsalou & Wiemer-Hastings, 2005). It consists of different underlying constructs, that together represent reputation. When we speak about a concept, we differentiate between two types of concepts abstract and concrete. Barsalou and Wiemer-Hastings (2005) define abstract concepts as “entities that are neither purely physical nor spatially constrained” (p.129). This can be, for example, the concept of freedom or truth. Concrete concepts are concepts we have a specific picture of. They often differ depending on a certain context and situation (Barsalou & Wiemer-Hastings, 2005). An example of a concrete concept is a table. If we think about a table, we will think about situations in which we use a table (eating in the living room or working on a project). It is very easy to draw a picture of the concept 'table' in our mind using its attributes. As a contrast, an abstract concept like freedom is harder to access. We might connect it to a feeling, or a picture of what freedom is, but we cannot easily think about attributes as we do for the concept table (Barsalou & Wiemer-Hastings, 2005).

The reason for this is that for understanding concepts, we rely on physical properties of concepts and the settings we find them in (Barsalou & Wiemer-Hastings, 2005). Thus, abstract concepts like freedom or reputation are hard to grasp since these are often not physical and don not appear in a specific setting. As a consequence, measuring an abstract concept is rather difficult (Izuma et al., 2014; Montes et al., 2017; Barsalou & Wiemer-Hastings, 2005). However, that does not mean that abstract concepts are just words with no connections in our brain that cannot be measured. In our brain concepts, whether they are concrete or abstract have an underlying structure with constructs, also referred to as mental models (Barsalou & Wiemer-Hastings, 2005, Clear, n.d.).

Mental models are cognitive representations of how we see the world, for example, how we see freedom (Jones et al., 2011). They help us to understand situations fast and act quickly and thereby enable us to make quick decisions. Every mental model holds variables, possible outcomes and biases that people need in order to make a decision (Jones et al., 2011). To build the mental model, we use our assumptions and experiences, and once it is built, it is difficult to change (Chermack, 2003). For example, when abstract concepts are processed, it triggers associated words (Barsalou & Wiemer-Hastings, 2005). When somebody thinks about the concept of freedom, automatically words associated with freedom come to mind like a forest and fresh air. These words all represent parts of what we think freedom is. However, these words alone do not give any semantic content for the concept and thus cannot readily be used to measure a concept like freedom. They are just words that might give a hint in the direction of underlying structured constructs. To obtain measurable categories, an additional step needs to be taken. The obtained words need to be put into perspective to unravel the underlying constructs and structure of the abstract concept (Barsalou & Wiemer-Hastings, 2005).

Summarizing two steps need to be taken to make an abstract concept like reputation measurable. (1) Words associated with the abstract concept need to be obtained. (2) The obtained words can be used to elicit the underlying structure of the mental model of the abstract concept, revealing measurable constructs.

When it comes to the method used to find words associated with the abstract concept, Word Association is an appropriate choice. Humans automatically use Word Association in their everyday life to simplify abstract concepts. Furthermore, Word association is often being used when one wants to find out more over a particular concept (Istifci, 2010; van der Velde et al., 2015).

For the second step, a method needs to be chosen that enables us to find underlying structures that reveal constructs. At first glance, Card Sorting is not the most obvious choice for designing a reputation system. Previously Card Sorting was used to (1) reveal underlying constructs of concepts like creativity to design a questionnaire (Van der Velde et al., 2015) and (2) to form categories for navigation structures in order to design usability friendly websites (Schmettow & Sommer, 2016). The reputation system design sits somewhere in-between. It has similarities with the design of navigation structures because we want to find underlying structures and use them as a basis to create a logical, usable and friendly reputation system. It also has the character of a questionnaire in the sense that the finished design is meant to assess a person and display the results. This is why the Card Sorting method was chosen in this study. In the end, we want to have a set of constructs and subconstructs that contain a number of words. These constructs should be measurable, and organisations should be able to use them as a basis for reputation systems in online social communities.

Summarizing, the mental models of reputation are investigated with Word Association (WA) and Card Sorting (CS) techniques. These techniques leave us with a semantic map of words related to reputation. The clusters in the semantic map represent groups that can be used to evaluate reputation. For the Word Association, we chose to use a restricted Word Association with three words to discover which words members of online social communities associate with reputation when they think about evaluating other members in their community. For the Card Sorting studies, we chose to use an online two-layer hierarchical open Card Sorting for the Pilot Card Sorting to get a general idea on how a structure for a reputation system might look like. For the main Card Sorting, we used a one-layer open Card Sorting, to get a general picture of the already divided reputation categories. For a more detailed explanation on how Word Association can be conducted, see van der Velde et al. (2015) and for Card Sorting, see Schmettow and Sommer (2016).

3. First Part: Word Association

As introduced above a restricted Word Association (WA) was conducted. We chose a restricted Word Association to avoid the potential bias of a free association. The restricted Word Association aimed to obtain a set of words associated with ‘reputation of members’ in the context of online social communities, which will be used later in the Card Sorting part of the study.

3.1. Method

3.1.1. Participants

91 members from different online communities participated in the first part of the study (Word Association) (47 female, 42 male, age range 18-55, mean age range 18-24). 44 were Dutch, 40 were German, and five were from a different nationality. The Word Association was approved by the University of Twente Faculty of Behavioural Management and Social Sciences Ethics Committee. All participants accepted the informed consent prior to participation. Participants who did not complete the survey were eliminated for incomplete data. 30 participants were excluded in further data analysis. 61 participants remained.

3.1.2. Material

An online questionnaire was designed to measure words associated with the reputation of members in online social communities. The questionnaire consisted of five items. The first four items were questions regarding demographic data like age and gender. The last question asked the participant to give the first three words that come to their mind when they think about reputation in the context of online social communities (see appendix B). The questionnaire was written in English so that people with different languages could participate. Participants who did not complete the questionnaire were excluded from the data.

3.1.3. Procedure

The survey was posted on different social media websites and in different online communities. In the online questionnaire, first, the participant was asked to read and accept the informed consent (see appendix A) and fill in some personal data. After that, the participant was asked to write down the first three words associated with the word reputation. In the end, the participant was thanked for filling in the questionnaire.

3.1.4. Data Analysis

3.1.4.1. Extracting the Data

Firstly, the data set consisting of all the words named in the WA were extracted from the survey into an excel sheet. Afterwards, it was counted how many participants named the same or similar words and scores were given to all words. A score of one means that a word was named by one participant, a score of two means that two different participants named the same word, and so on. Based on these scores, a table was created with a score for every word.

3.1.4.2. Adding Words from a Meeting

In addition to the Word Association, a literature research meeting - with two other persons' who are involved in developing an evaluation tool for online social communities - was held. In the meeting, we discussed the results of the Word Association and words found during the literature research. In the end, we made a list of relevant words obtained during the literature research.

3.1.4.3. Deciding which Words to Use in the Further Research Process

The list of words obtained in the Word Association and during the literature research meeting (see appendix C) could not all be used for the next part of the study (Pilot Card Sorting). A selection of words needed to be made. First of all, all words named at least twice were added to the final word list. Additionally, all three persons' who attended the literature research meeting including myself rated the words named once on the Word Association list and words on the literature research list with '1' (association with reputation) and '0' (no association with reputation). In the following, the three persons will be referred to as raters. Words rated with '1' by all raters were added to the final list (see appendix E).

3.2. Results

One goal of the WA was to obtain items for the Pilot Card Sorting. The online restricted Word Association produced a set of 115 words. The existing reputation systems meeting produced 44 words. This resulted in a list of 159 words. Table 1 illustrated all words that were mentioned more than once (for a complete list see appendix C). The score illustrates how many times, different participants named a word. The final word list contained a 100 words.

Table 1

List of words with a higher score than obtained from the Word Association.

Score	Words
11	Fake
8	Likes
5	Social
4	Advertising medium
4	Fame
4	Followers
4	Influencers
4	Privacy
4	Pictures
3	Addiction
3	Friends/Friend Group
3	Perfection/Perfect
3	Trust/Trustful
2	Achievements
2	Annoying
2	Blog
2	Hater
2	Image
2	Power/full
2	Public
2	Respect
2	Rewards
2	Sharing
2	Status
2	Supportive

3.3. Discussion

The goal of the first part of study was to find words associated with the concept of reputation in the context of online social communities. Looking at the results, it stood out that eleven participants named the word *Fake*. The word *Likes* was written down by eight participants and the word *Social* had a score of five. Six words had a score of four. Four words had a score of three and twelve words had a score of two. Twenty-five words were named more than once. All other 134 words were only named by one participant each. The fact that many words were only mentioned once despite having a high number of participants (61) can mean that there is a broad opinion on what reputation is and means in the context of online communities. For that same reason, it might be interesting to take a broad look at the underlying semantic structure of all words at hand first to get a general idea on what people generally understand under the concept of reputation in online social communities. A Pilot Card Sorting can help to create that general picture.

The next section describes the process of preparing and executing the Card Sorting studies. After that, the results of the Card Sorting studies are presented, discussed, and two reputation systems are introduced.

4. Second Part: Pilot Card Sorting

In the first part of the study, a list of 159 words associated with reputation was obtained from the restricted Word Association and the existing reputation systems meeting (see appendix C). We used this list to select words for the Card Sorting. The selection was based on two conditions. Firstly, all the words that appeared more than once as an answer were selected. Secondly, words that were only named by one participant and all the words obtained in the literature research meeting were rated. This resulted in a list of 100 words, which is a high number of words for the Pilot Card Sorting (see appendix E). The reason that so many words were selected is that the goal of the Pilot Card Sorting was to create a comprehensive picture of how reputation could be evaluated further. The Pilot Card Sorting aimed at getting a first idea on the internal semantic structure of the words associated with 'reputation' and possible constructs to evaluate the reputation of members in online social communities. To do so, we conducted an open hierarchical Card Sorting.

We chose Open Card Sorting to get to know more about their mental model. There are two reasons why the Card Sorting was conducted online. First and foremost, people from online communities are the target audience and can be best reached online. Second of all, more participants could be reached, than it would be possible with a physical Card Sorting. We conducted a two-layer Card Sorting to give the Card Sorting more meaning and help to narrow down possible nested groups to get a general picture of constructs of reputation.

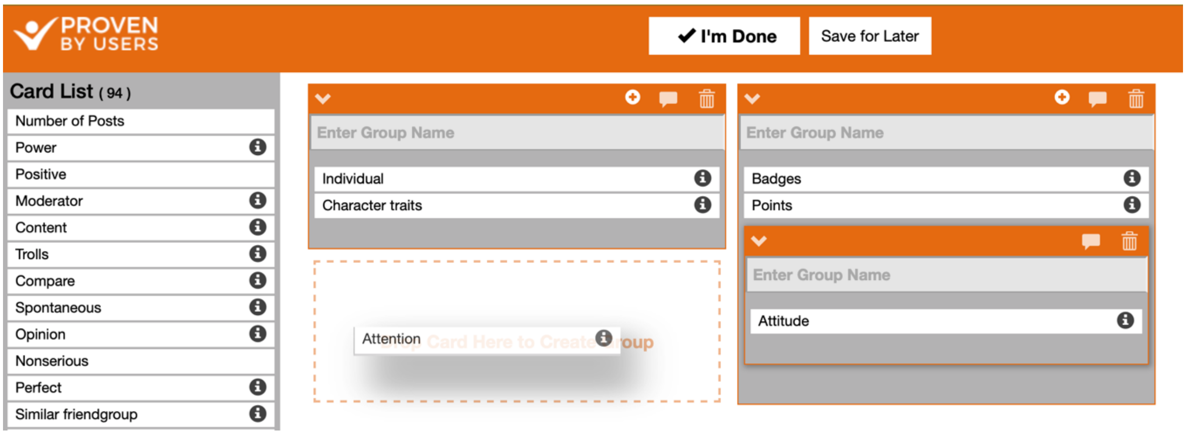
4.1. Method

4.1.1. Participants

30 members from different online social communities, who did not participate in the first part of the study, conducted the second part of the Pilot Card Sorting (16 female, 13 male, 1 other, age range 18-40, mean age 27). 7 were Dutch, 17 were German, and 6 had another nationality. The University of Twente Faculty of Behavioural Management and Social Sciences Ethics Committee approved the Pilot Card Sorting. All participants agreed with the online informed consent prior to participation.

4.1.2. Material

To conduct the Card Sorting, the online Card Sorting tool 'provenbyusers' was used. Figure 1 shows the layout of the tool. The set of cards that need to be sorted are on the left (see figure 1A). The cards can be sorted into groups on the right (see figure 1B and C). To do so, a participant can drop a card somewhere in the right white field (see figure 1B bottom). The sorted groups can also be divided to build subgroups by clicking on the + sign. The participant can put words in the subgroup afterwards (see figure 1C).



A B C

Figure 1. Tool used for Card Sorting Study. (A) Set of words (B) Dropping a card into the right field a group (C) A group with a subgroup.

The set of words we selected prior to the second part of the study were used for the Card Sorting. On every card, a word from the list was written down with a short definition of the word in case a participant did not know the meaning of a word (see figure 2).

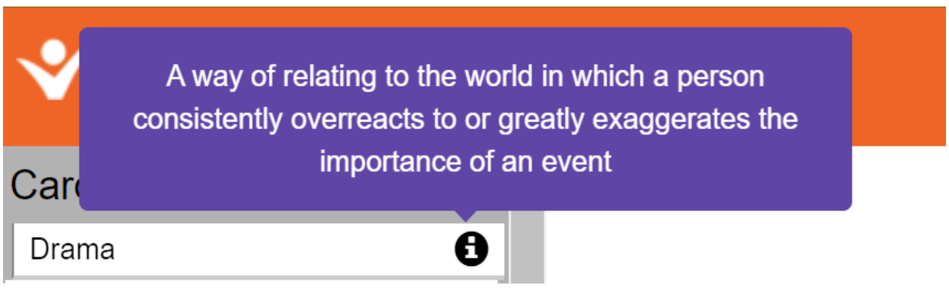


Figure 2. The card *Drama* with definition.

4.1.3. Procedure

First, the participant was asked to read and accept the informed consent (see appendix D). After that, the participant was instructed to share their gender, age and nationality. Next, the participant was asked to read the instructions. Then the participant was asked to sort the cards into groups with a maximum of one subgroup per group.

4.1.4. Data Analysis

4.1.4.1. Jaccard Coefficient Score

The data collected during the Card Sorting was analysed with the Jaccard Coefficient (similarity measure) to obtain similarity scores that can be presented in a similarity matrix. The Jaccard Coefficient Score creates a similarity measure between two items. Two steps are used to obtain the Jaccard Coefficient of the two items *Achievement* and *Trolls*: (1) counting the number of groups *Achievement* and *Trolls* both belong to (2) dividing it by the number of groups to which either *Achievement* or *Trolls* belong to (Schmettow & Sommer, 2016). First, the Jaccard Score for each participant was obtained and written down in a table using Excel. After that, all scores were combined in one table in Excel. Thereby the unorganised heatmap is created.

4.1.4.2. Agglomerative Hierarchical Cluster Analysis

The overall score table was analysed with a vector analysis in the programme 'R' to produce a heat map and dendrogram. The obtained clusters of both the dendrogram and the heatmap were used to build groups. These groups are displayed in a tentative cluster structure.

For the analysis, it was chosen not to use the standard Card Sorting analysis, but a more complex version. In a standard Card Sorting analysis, the two items with the highest score (the two highest associated words) are selected and are replaced by a cluster item (single item). After that other item scores are calculated using the average of the two scores of the two items the cluster was derived from. This procedure is repeated until no items are left (Schmettow & Sommer, 2016; van der Velde, 2018). In this way, different clusters are obtained that can be represented with a heatmap.

However, using this method, the construct relations will be based only on the comparison of one datum. The problem with this is that, for example, the first cluster that is formed (the one with the highest relation) can have a little stronger relation than another cluster. Hence, a small difference in a score can determine the basis of the cluster analysis (van der Velde, 2018). In this part of the study, it was essential to get to know how similar two items are in order to obtain logical categories for evaluation. A vector comparison is more suitable in that case. That is why we chose a more complex comparison.

In the complex comparison, two items are strongly related based on two conditions: (1) they score high in the same cluster; (2) they score similarly in other clusters. That means when an item scores high in a group, the related item should also score high in the same group. In the more complex version, the data is analysed with a vector comparison. To do that the 'Euclidean distance' between the vectors is calculated. The lower the distance, the stronger the relation. The distance between the vectors is then used as the basis for the dendrogram and heatmap (van der Velde, 2018).

4.2. Results

The results of the Pilot Card Sorting are presented in a dendrogram (see figure 3, figure 4) and heat map (see figure 5).

4.2.1. Dendrogram

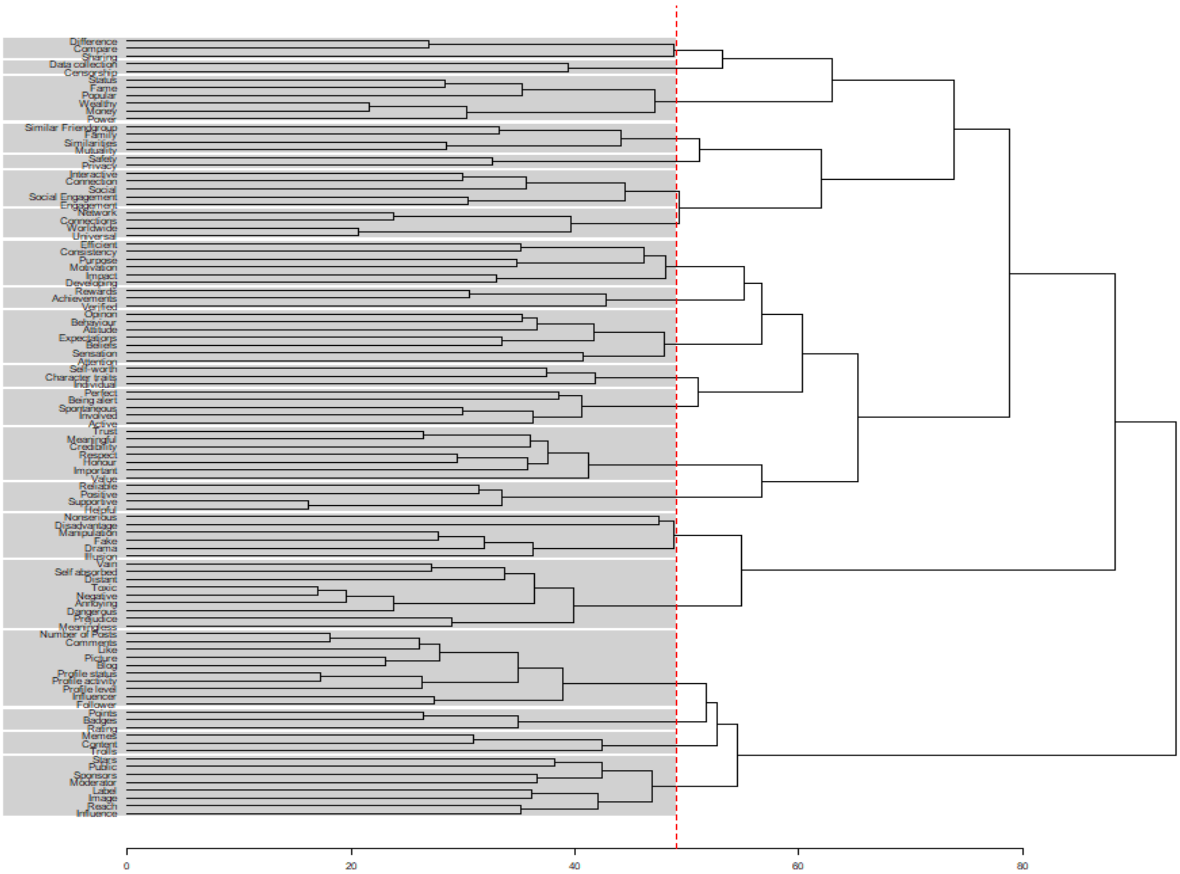


Figure 3. Dendrogram with clusters. The grey rectangles underline the different clusters. The red line indicates where the dendrogram was cut.

The dendrogram presents the distances between the vectors in a tree diagram. The hierarchical cluster structure starts with clusters of one or two words at the left and ends with two overall clusters at the right (see figure 3). The horizontal axis displays the distance between clusters and sub-clusters. The vertical axis represents the set of words and clusters. A vertical line needs to be drawn, to find meaningful clusters. Clusters that are next to each other or in proximity to one another have a higher association than clusters that are far away. The first cluster, for example, has a weak association with the last cluster. Relevant clusters were chosen according to the following criteria (1) number of clusters according to the elbow method (2) number of clusters according to the silhouette method and (3) number of clusters according to the relative distances observed in the graph. All three methods are shortly explained in the following:

The elbow method uses the percentage of variances that can be explained by the number of clusters and displays them in a graph. At first, the variance is high, but at a certain point in the data, the variance drops and gives an elbow like angle in the data. Depending on this point, the number of clusters is chosen. The silhouette method uses consistency within the clusters. Different values are calculated by measuring how similar a word is to its clusters compared to other clusters. High values in the graphic indicate that the words are well matched. One of the highest values is chosen. In the relative distance method, the number of clusters is chosen according to the relative distances in the dendrogram. The researcher looks at possible jumps in the distances that indicate where to cut the dendrogram. For this method, the context of the data is also taken into consideration. Thus for this method, both the context and distances are used to set a line. Additionally, relevant clusters are counted.

The elbow method proposes 6 to 9 clusters. The silhouette method suggests 4 to 5 clusters. The elbow and the silhouette method were only used to get an indication of where the line should be drawn, but this method can be imprecise because the context is not taken into account. That is why all three methods were used in combination to set the red line. Looking at the data at hand it can be seen that from around 45 - see the red line – there are big ‘jumps’ in the data, which indicates that something might be merged that actually should not be merged. That is why the red line was drawn around 45, which results in 20 clusters. Cluster to the left of the red line might give information on relevant constructs which could be used to measure reputation. The grey rectangles at the left of the line underline the relevant clusters.

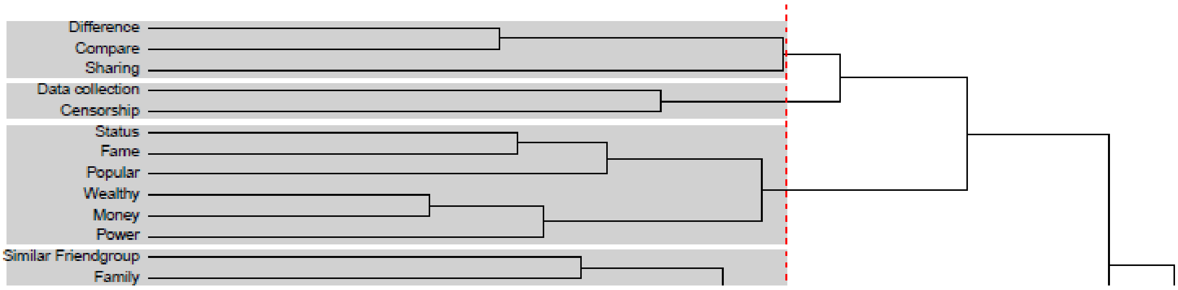


Figure 4. Zoomed in view of dendrogram. The grey rectangles indicate the clusters. The red line shows were the dendrogram was cut.

Figure 4 shows a zoomed-in view of the dendrogram. The zoomed-in view presents three clusters. The red line indicates where the tree is cut off, and the joined leaves on the left side of the red line indicate the clusters. The grey rectangles are used to underline the clusters. The first cluster consists of three words: *Difference, Compare* and *Sharing*. The second cluster consists of two words: *Data Collection* and *Censorship*. The last cluster consists of six words: *Status, Fame, Popular, Wealthy, Money* and *Power*. The first cluster was just cut after the merge of the cluster *Difference*, and the cluster *Compare* and *Sharing*. That means the first cluster could also be split into two clusters.

4.2.2. Heatmap

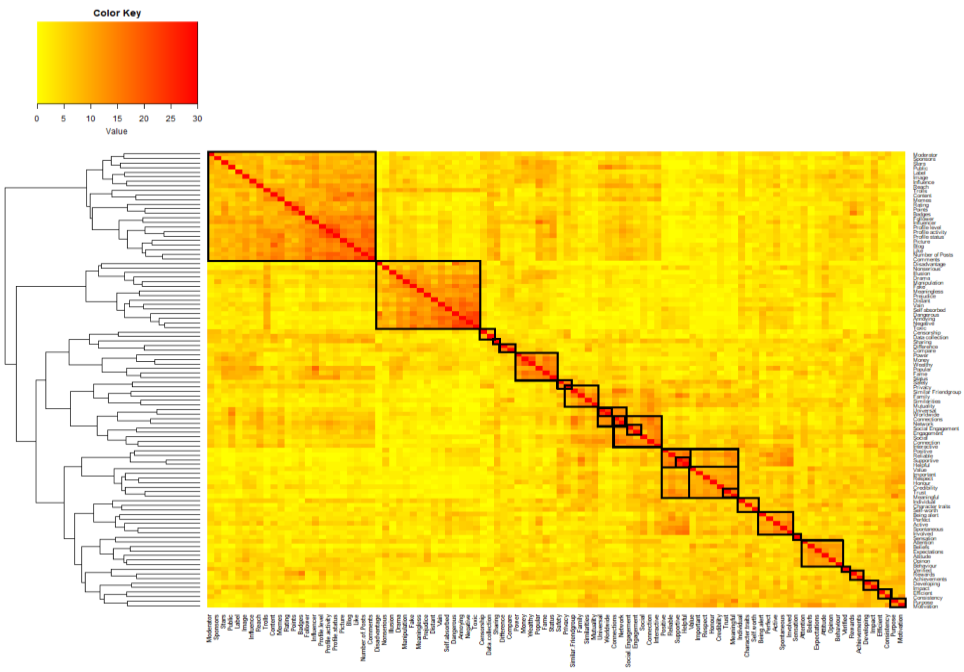


Figure 5. Heatmap with clustered items. The black rectangles underline the cluster.

Figure 5 presents the results of the Pilot Card Sorting in a heatmap. The colour indicates the strength of the association between two words with red = strong and yellow = weak. In other words, the colour in each cell represents how often every word from the row and column belonged to the same group for all participants. There were 30 participants in the Pilot Card Sorting, meaning this number can range between 0 and 30. These numbers are displayed as colour ranging from light yellow (0) to deep red (30). The obtained data showed that the lowest number was 0 and the highest number, 24. That means, for example, that at least two words were sorted in the same group by 24 participants. The squares that form groups of words are related to the clusters of the dendrogram (see figure 3). In the top left corner, there is a 24 x 24 square that is much darker than the yellow around it. The square includes 24 words: *Moderator, Sponsors, Stars, Public, Label, Image, Influence, Reach, Trolls, Content, Memes, Rating, Points*, etcetera. Some words which belong to one group also have a strong association with words outside that group. For example, the word *Profile Level* belongs to the 24x24 group but can also be associated with *Popular*. The distance between groups cannot be seen in the heatmap but only in the dendrogram. Looking at the heatmap, it stands out that the clusters get smaller and the strength of the association gets weaker down the diagonal. In general, there are a lot of dark orange spots all over the heatmap, which indicate that words in the cluster could also be associated with words outside the cluster. The red blocks indicate that there might be 20 groups with seven subgroups that might represent possible constructs to measure reputation.

4.2.3. Tentative Cluster Structure

A set of clusters potentially related to reputation was created based on the dendrogram presented in figure 3 and the heatmap presented in figure 5. In order to create the tentative cluster structure, both the heatmap and the dendrogram were used. Additionally, the distances between clusters from the dendrogram were also taken into consideration. All in all, there are 11 big cluster groups. Counting all the clusters suggested by both the heatmap and dendrogram, there are 20 different clusters with additional sub-clusters. Respectively there is also additional information from the heatmap which suggest combining two or more clusters and create subgroups inside those groups, as the heatmap overall suggests bigger cluster groups than the dendrogram. The main clusters of items associated with the concept of reputation with the groups and subgroups are presented in Figure 6, 7 8 and 9.

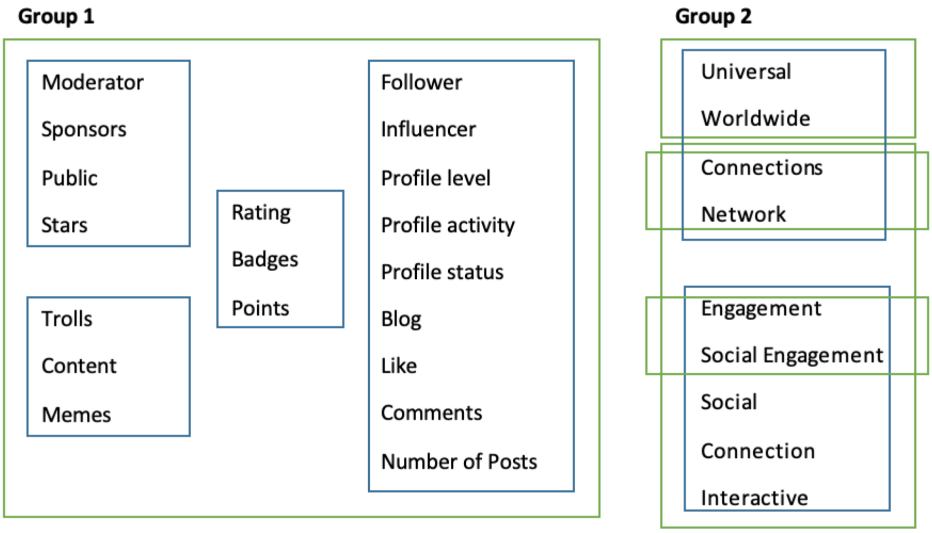


Figure 6. Group 1 and 2. Tentative clusters building a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows cluster obtained from the heatmap.

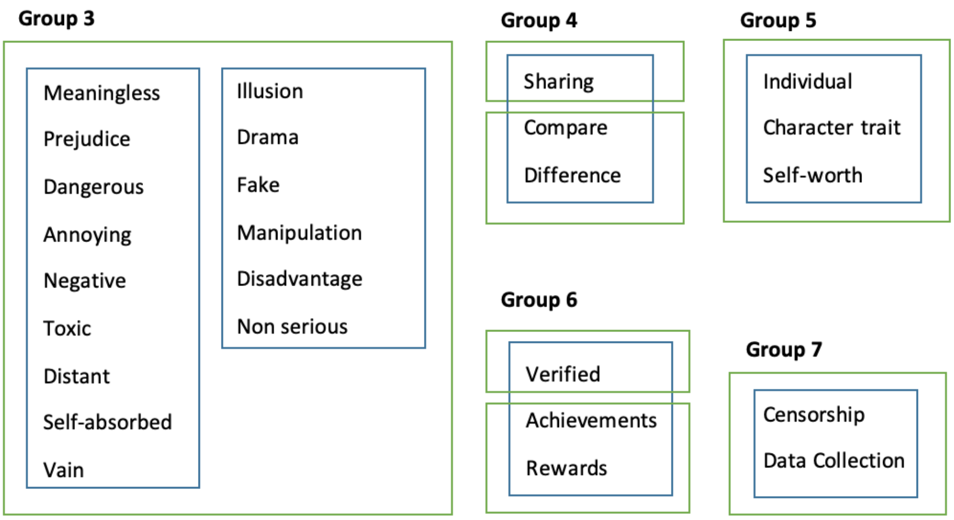


Figure 7. Group 3, 4, 5, 6 and 7. Tentative clusters building a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows cluster obtained from the heatmap.

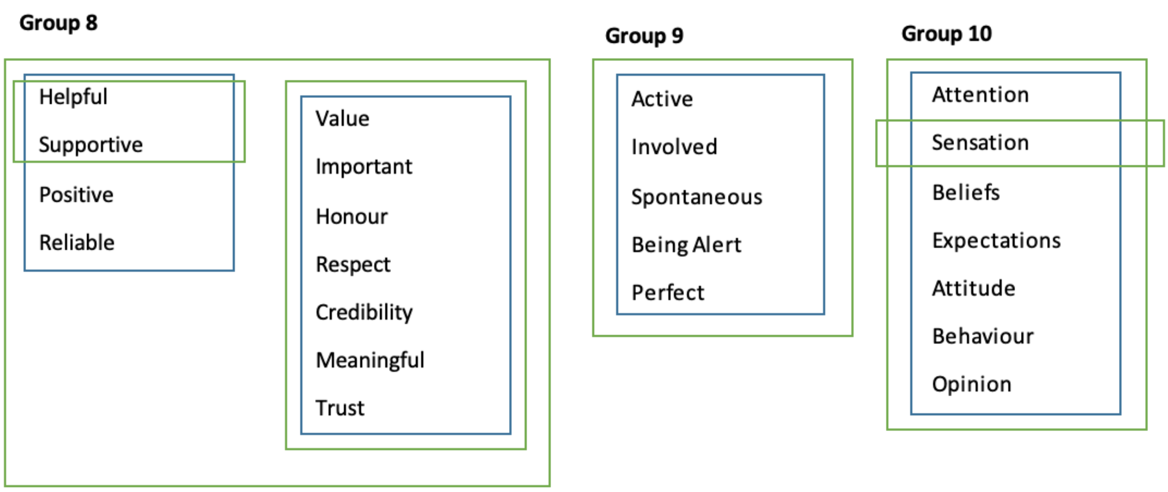


Figure 8. Group 8, 9 and 10. Tentative clusters building a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows cluster obtained from the heatmap.

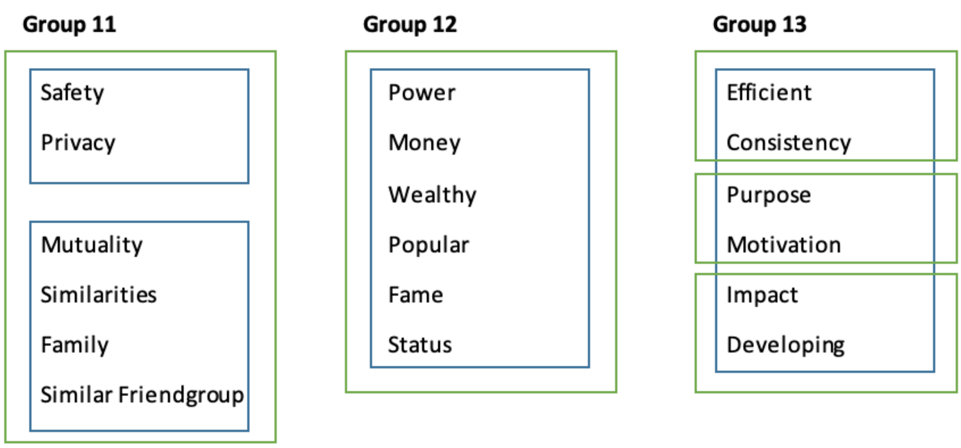


Figure 9. Group 11, 12 and 13. Tentative clusters building a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows cluster obtained from the heatmap.

It is hard to come up with overall terms for clusters and sub-clusters. Most clusters have at least one word that falls out of line not logically fitting in with the rest of the words. For example, Group 9 consists of five words. The words *Active* and *Involved* both describe somebodies *Status* in a community. Users can either be active and involved, which is called an active member or inactive and just looking at the content often referred to as lurkers. *Being alert, Spontaneous* and *Perfect* all seem to fall out of line. However, some sub-clusters seem to have an overall concept like the subcluster in Group 8. The words Helpful, Supportive and Positive, are all positive attributes and could be summarized under the term **Helpful Character** or **Desired Behaviour**. Observing the different clusters focusing on which ones have a meaningful structure it meets the eye that some words make more sense in the context of a rating system and others work in the sense of constructs used for processing data for an automatic system. Constructs from both system categories are mixed in the clusters. Clusters that seem to only contain words from one domain can be summarized by a meaningful category. Likewise clusters that contain words from both domains do not seem to have a summarizing term in common.

4.3 Discussion

The goal of the Pilot Card Sorting was to get a general idea of how reputation in the context of online social communities can be evaluated and if there are possible constructs to be found. In the heatmap, it stands out, that many words which are in one group could also be associated with words in other groups. The dendrogram suggests a few smaller clusters than the heatmap does. However, after combining both results in the tentative cluster structure and analysing each cluster carefully, it stands out that many groups suggested by the dendrogram do not have an overall term (blue rectangles). Often the words do not have something in common. Groups obtained by the heatmap often have an overall category that can represent it (green rectangles).

One reason for this could be that using both a dendrogram and heatmap can create ambiguity. However, using both also gives a more detailed picture of the relationship between the items. The heatmap gives information that the dendrogram does not give and vice versa. In the heatmap, one can see which items belong to a common group while the dendrogram reveals the relationships between groups. The dendrogram displays the distance between groups that help understand why some words in the heatmap that are not in the same group still have a strong association. Furthermore, the dendrogram displays subgroups pretty well, whereas in the heatmap spotting subgroups is rather difficult and less precise. While creating the tentative cluster structure, it was very important to look at the relationship of the items from different angles to get a clear picture on how groups and subgroups might look like and which group makes sense for a reputation system. Combining the results from the heatmap and dendrogram into a tentative cluster structure can help to create a diverse group that represents the concept of reputation as precisely as possible.

Examining the tentative clusters in the tentative cluster structure further (see figure 6, 7, 8 and 9) and looking for other possible reasons to why the majority of the obtained clusters have at least one to two words that fall out of line, it was discovered that at least two different reputation categories might have been mixed. For example, the words *Comments, Number of Posts* and *Likes* from Group 1 could all be measured automatically. In order to do that, a system could use an algorithm to count the *Number of Comments*, *Likes* and *Number of Posts* and give values. In contrast, constructs that belong to Group 2 *Self –absorbed, Distant* and *Non-serious* cannot just be measured automatically but would need an evaluation by a real person, maybe another user.

Thus, in order to measure reputation, two systems might be needed. Two possible systems could be a peer to peer and an automated reputation system. A peer to peer system would take in data of online users based on ratings and feedback of other users in order to give a reputation. An automated system would use automatically generated data like comments, likes and number of posts to rate somebody's reputation. Taking a look at the words from the tentative cluster structure, again it seems some words fit a peer to peer reputations system better like *Trust* and *Respect* and others seem to belong to an automated reputation system *Data Collection* and *Badges*. *Trust* and *Respect* cannot easily be measured by collecting data out of the community and performed actions but often have to be rated by human raters. Conversely, data collection can be used by obtaining data out of the activity stream of a user and badges can be counted easily automatically by the system.

The clustering in each reputation category could be different if these categories are studied separately. Thus, the cards should be sorted again but with the two categories (peer to peer and automated) separated from each other. The next section will describe the process of sorting the words obtained by the Word Association into two different categories and presents the results obtained by the follow-up Card Sorting of both categories. Afterwards, results of both studies are discussed, connections are made, and two possible systems are introduced briefly.

5. Third Part: Automated and Peer to Peer Card Sorting

In the first part of the study, we obtained words associated with reputation. In the second part of the study, participants were asked to sort those words into groups to get a general idea on possible structures of the constructs of reputation. The Pilot Card Sorting gave a nice overview of possible constructs and a broad overview of possibilities. The results gave some interesting insights into the underlying semantic structure of the obtained words. The results hint that there might be at least two different categories of systems in the underlying structure. That means two reputation systems might be needed for online social communities: peer to peer and automated. In order to get more information on these findings, three people were asked to sort all words obtained from the Word Association in one of the three categories (1) automated (2) peer to peer (3) neither automated nor peer to peer. They were also asked to rate the words that belong to (1) or (2) after they sorted them into different categories. They were asked to rate the words according to what they think fits the best into the category 1- … . One was the word that fits the best. A list with 42 words was obtained for the peer to peer category, and a list of 48 words was obtained for the automated category, ten words were identified to not fit in either one of these categories (see appendix G). A heatmap was created for both of the categories from the already existing data. The goal was to see whether clusters might be distributed differently.

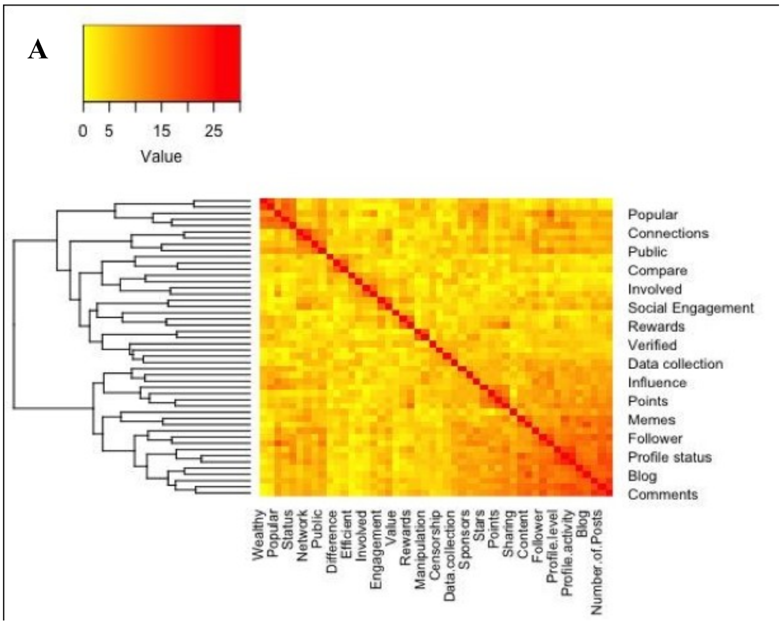


Figure 10. Heatmaps for clusters items with already existing data. (A) Heatmap of the automated domain. Red indicates high association between items. Yellow indicates a low association.

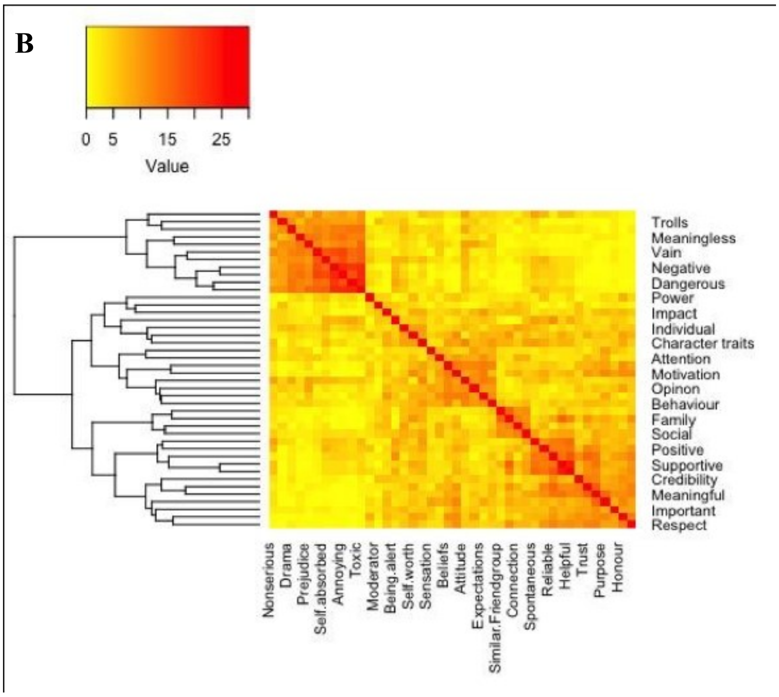


Figure 11. Heatmaps for clusters items with already existing data. (B) Heatmap of the peer to peer system domain. Red indicates high association between items. Yellow indicates a low association.

The heatmaps do not indicate any clear clusters (see figure 10, 11). There are no clear darker rectangles along the diagonal. We decided to do another Card Sorting on each set separately because of two reasons. (1) The existing data set is not a reliable data set to use for analysis because the data was obtained with another purpose in mind. Now we want to look at reputation with the two found domains in mind; thus, conclusions cannot simply be extracted from the old data. (2) The heatmaps with the existing data might not indicate any clear groups because the two domains were mixed together.

5.1. Automated Card Sorting

This part of the study aims to find out whether the found constructs for the automated system category have a meaningful structure. In order to do that, an open Card Sorting was conducted. As in the Pilot Card Sorting, we chose an online open Card Sorting. Different than the Pilot study, we conducted a one-layer Card Sorting for the reason that the main classification was already made by dividing the words into the two reputation categories (automated and peer to peer). The data obtained by the Card Sorting were analysed, clusters were formed, and a tentative cluster structure was created based on a heatmap and a dendrogram.

5.1.1. Method

The same method as the Pilot Card Sorting was used for this study with the following changes: 31 members from different social online communities conducted the automated Card Sorting (16 female, 15 male, age range 18-51, mean age 22). Eight were Dutch, 18 were German, and five were from other nationalities. The same Card Sorting Tool was used but items could only be sorted into groups without subgroups (see figure 12) and the participants were asked to sort all items into groups without creating subgroups.

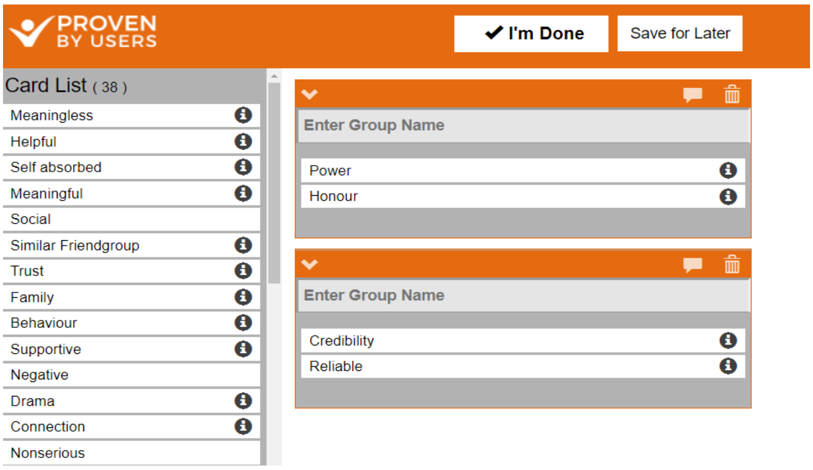


Figure 12. Tool used for Card Sorting Study. Items to be sorted are on the left side. On the right side are groups with items.

5.1.2 Results

The results of the automated Card Sorting are presented in a heatmap (Figure 13) and dendrogram (Figure 14).

5.1.2.1. Heatmap

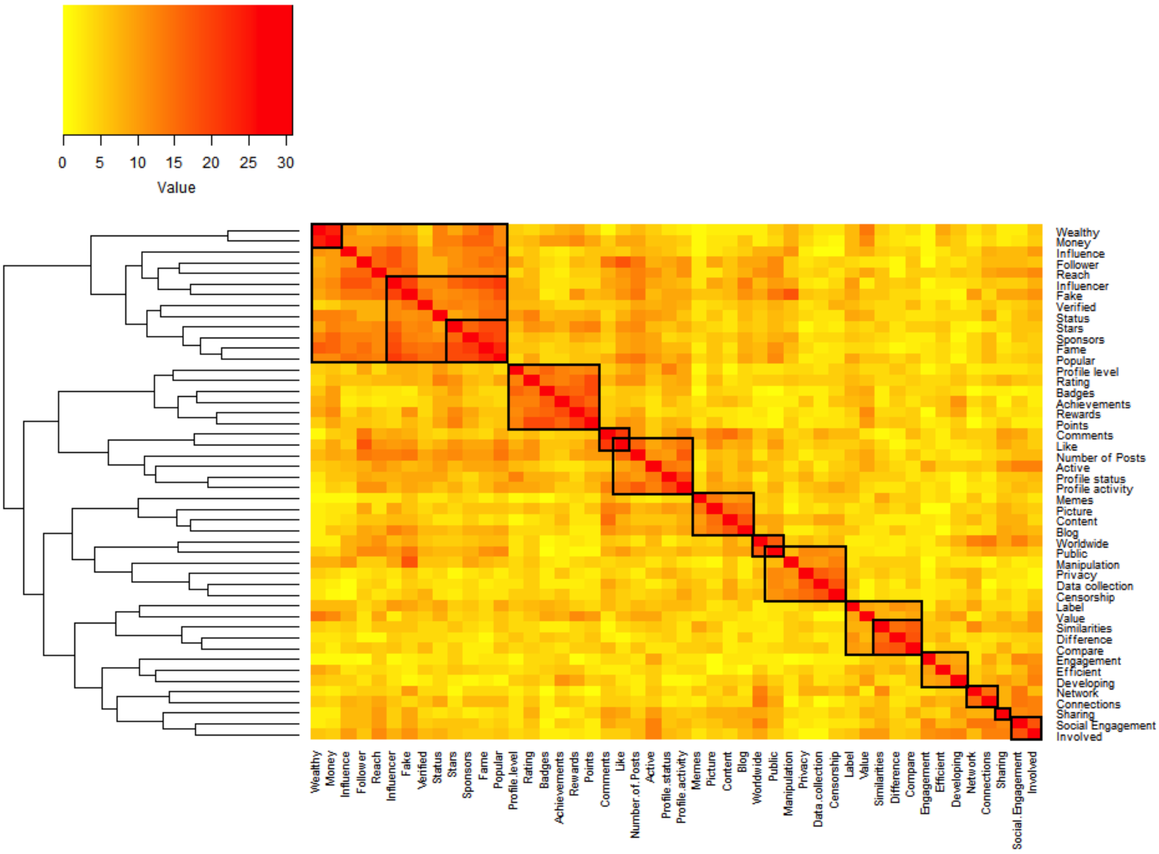


Figure 13. Heatmap with clustered items for the automated reputation category. The black rectangles underline the found clusters. Red indicates strong and yellow a weak association. There are twelve clusters and four subclusters.

The heatmap presents the distances between items obtained from the vector analysis for the automated reputation system. The obtained data shows that the strength of the association of the words ranges between 1 and 24. 1 is the weakest and 24 the strongest association. That means the redder a rectangle, the higher the association. The black rectangles underline the found clusters and subclusters. The heatmap proposes 12 clusters and four sub-clusters. It stands out that there are smaller clusters within bigger clusters, which might indicate that there are sub-clusters. In the top left corner, there is a darker 13x13 square. Within this square, there is a darker 2x2 square consisting of two words *Wealthy* and *Money*. In the same 13x13 square, there are two more groups one 8x8 square, and within this square, there is a 4x4 square. This scheme goes down the whole diagonal. Almost every bigger group has some darker spots which indicate that there are some smaller groups.

Furthermore, there are several bleeding spots. Bleeding spots are off-diagonal darker spots in a heatmap. They indicate that words that belong to one cluster might also be associated with one or more words from other clusters. The most bleeding spots are in the upper left half of the heatmap. Words in the upper clusters might not only belong to one cluster but can also be associated with other words from another cluster. For example, the word *Follower* might also be associated with the words *Comment*, *Likes* and *Number of Posts*.

5.1.2.2. Dendrogram

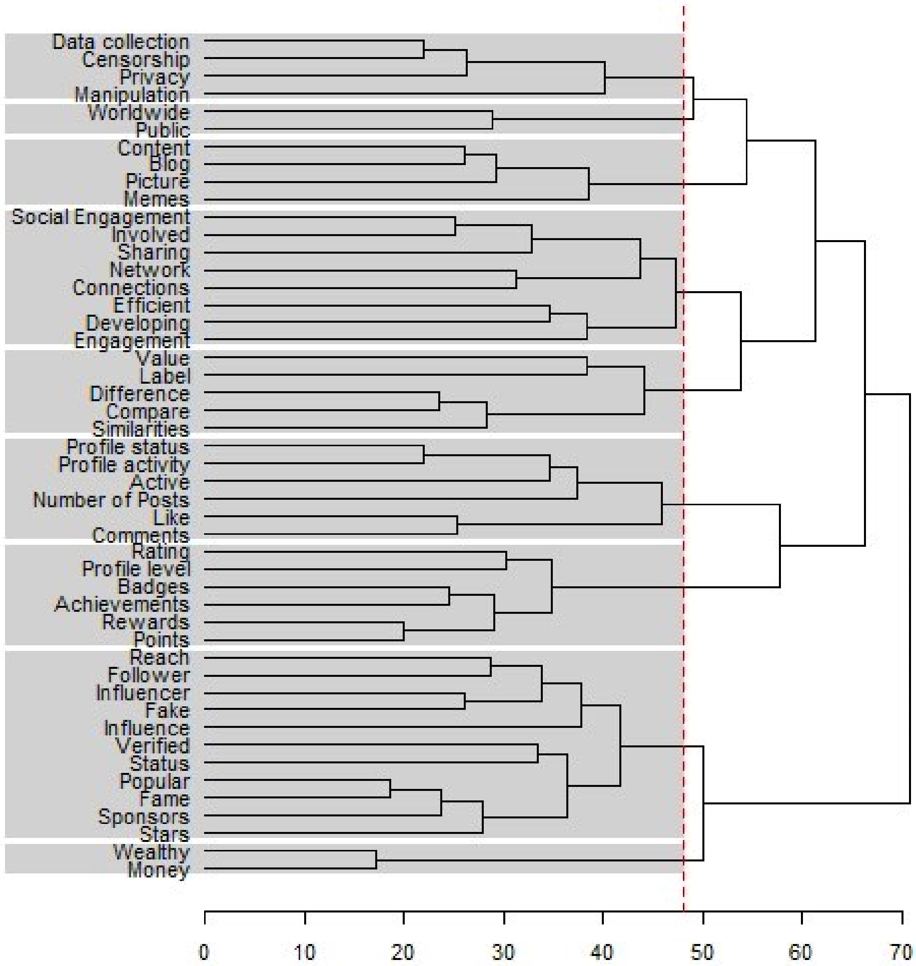


Figure 14. Dendrogram with 9 clusters (see grey rectangles) for the automated reputation system category. The red line indicates where the dendrogram was cut.

The hierarchical cluster structure in figure 14 represents the distances between the different items. It starts with clusters of one or two words at the left and ends with two overall clusters at the right. According to the elbow method, there should be around 7-9 clusters, and according to the silhouette method, there are around six relevant clusters. Again both of the methods were only used as an orientation to see where the line should be roughly drawn. It can be seen that from 58 onwards there are a few big jumps and distances are getting bigger, which means the association of the words gets weaker.

Additionally, it can be seen that the distances within the sub-clusters on the left side of the red line are comparably short in comparison with the distances between the sub-clusters on the right side of the line. Thus, the line for the relevant clusters was drawn around 58, which results in 9 clusters. The grey rectangles on the left underline the 9 clusters. They consist of a varying amount of words. Two of the clusters contain two words. The rest of the clusters consist of 4 to 11 words. The first and the second cluster could be possibly merged if it would make more sense in the context, as the dendrogram is just cut before the two clusters merge.

5.1.2.3. Tentative Cluster Structure

Based on the dendrogram and heatmap (figure 13, figure 14), groups with subgroups were created. The blue rectangles mark groups created with the obtained data from the dendrogram, and the green ones are groups obtained from the heatmap. In general, the heatmap often more subgroups than the dendrogram. The mixture of displaying both clusters from the heatmap and dendrogram creates seven big groups with additionally 14 subgroups. The first group has four subgroups. The second, third and fourth group have three subgroups. The last group has two subgroups. All groups will be discussed in more detail in the following:

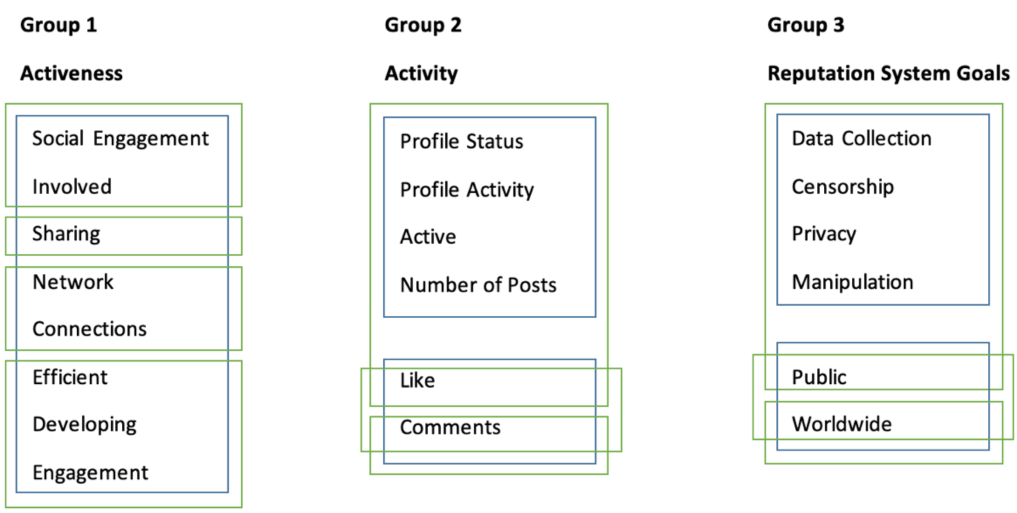


Figure 15. Group 1, 2 and 3. Tentative clusters building a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows cluster obtained from the heatmap.

The first group consists of eight words, with additionally four subgroups. The words *Social Engagement, Involved, Sharing, Network, Connections and Engagement*, could be summarised by the overall term of **Activeness**. All words have something to do with how active a user is. If he is active, he is involved in the community. This often means that the person has a lot of connections and a big network. The subgroup consisting of the words *Social Engagement* and *Involved* could be summarised under the term **Participation**. *Sharing* stands alone in the second subgroup. Somebody that shares content of knowledge is active in the community so *Sharing* can also be summarised by the terms **Engagement** and **Involvement**. A fitting term for the subgroup consisting of the words *Network* and *Connections* could be **Socializing.** Networking with people and establishing meaningful connections with others is seen as a socialising process. The subgroup which contains the words *Efficient, Developing* and *Engagement* all fit the term of **Productivity**.

Group 2 consists of six words with three subgroups: *Profile Status, Profile Activity, Active, Number of Posts, Like* and *Comments*. The whole group could be summarised with the term **Activity**. All words can be used to measure how active a user is. If a user changes his profile status often of posts much content, he is very active in the community.

Furthermore, all words have something to do with creating content. The subgroup containing the words *Like* and *Comment* can be best summarised under **Content Creation**. The other subgroup consisting of the rest of the words can be described by the word **Profile Activity**.

The third group consists of six words, with three subgroups. In general, it is difficult to come up with one term for the overall group. However, looking at the words carefully, the category that describes the whole group might be **Reputation System Goals**. Reputation System should work *Worldwid*e. They are meant to *Collect Data*, protect *Privacy* and prevent *Manipulation*. Separately in the different subgroups, constructs can be found. *Data Collection, Censorship, Privacy* and *Manipulation* might be seen as **Endangerment.** *Data Collection* on its own can be seen as a **Measurement Tool** to measure other constructs. For example, on one hand, *Data Collection* could endanger user’s privacy, and they might be manipulated more easily. On the other hand, *Data Collection* could be used to measure other constructs as long as keeping the user’s privacy is kept in mind and the user’s wellbeing is the priority. The second subgroup consisting of the words *Public* and *Worldwide* can be summarised as **Inclusiveness** or **International**. *Public* and *Worldwide* can be seen as a description of social communities in general. Almost everybody has a mobile phone and is a member of an online social community, which makes it easy to get in touch with people all over the world.

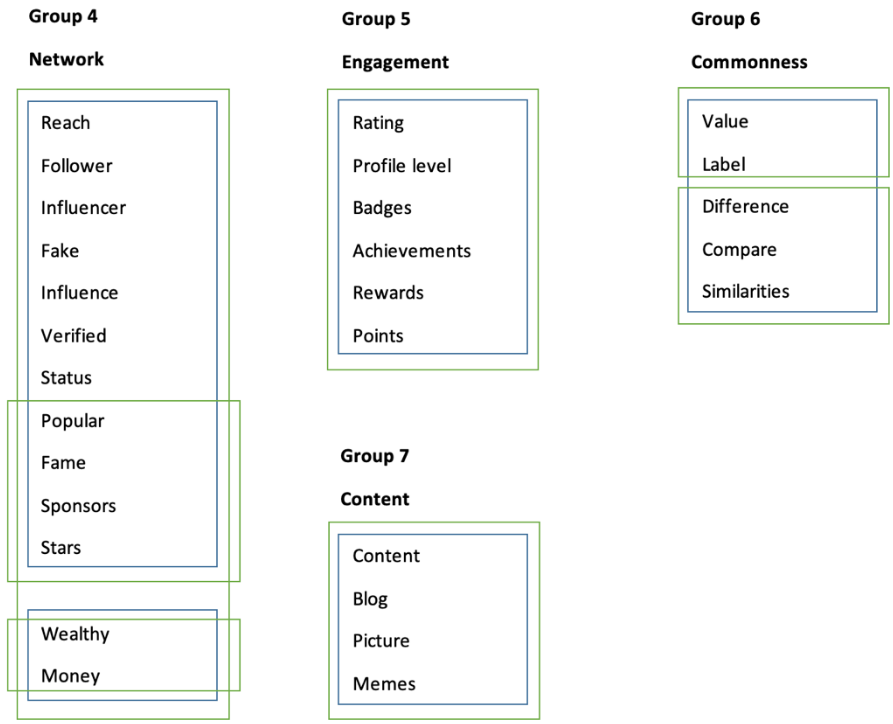


Figure 16. Group 4, 5, 6 and 7. Tentative clusters building a tentative cluster structure for an automated reputation system. Blue are clusters indicated by the dendrogram and green shows clusters obtained from the heatmap.

Group 4 consists of 13 words, with additionally three subgroups. The word that best describes the whole group is **Network**. The subgroup consisting of the words *Reach, Follower, Influencer, Fake, Influence, Verified, Status,* can be best summarised by the word **Reach**. The subgroup with the words *Popular, Fame, Sponsors* and *Stars*, describes different **Types of Users**. That can be, for instance, influencers often are very popular, raise to fame, and they often have sponsors like cloth or makeup brands.

Group 5 consists of six words. All the words *Rating, Profile level, Badges, Achievements, Rewards* and *Points,* fit with the term **Gamification** or more specific user **Engagement**. Gamification is used in online social communities to motivate members to become active and compete against each other. All of these words can also be used to assess somebody's reputation. *Achievements* and *Rewards* can be used to display somebody's trustworthiness and reliability can be used for reputation. *Wealthy* and *Money* might show the circles somebody chooses to stay in.

Group 6 consists of five words with two subgroups: *Value, Label, Difference, Compare* and *Similarities*. The whole group can be summarised with the term of **Commonness**. The first subgroup (*Values*, *Labels*) can be best summarised under the term **Labels** are often used by people to oversee their environment quickly. A person, for example, labels another person as dangerous based on their looks. *Values* can be either used for measuring or to compare values of different persons. The second subgroup (*Compare, Differences, Similarities*) fits the term of **Comparison**. Online two or more persons can be compared, and differences and similarities can be determined. All these words can also be used to label another person.

Group 7 consists of four words: *Content, Blog, Picture* and *Memes*. The word **Content** can be used as the term of the overall group. *Blog*, *Picture* and *Memes* are all content posted online in social communities. It can be further divided into **Created** and **Viewed Content**.

5.2 Peer to Peer Card Sorting

This part of the study aims to find out whether the constructs from the peer to peer system category form a meaningful structure. In order to do that, an on-layer open Card Sorting was conducted like in the automated system part. The data obtained by the Card Sorting were analysed, clusters were formed, and a tentative cluster structure was created based on a heatmap and a dendrogram.

5.2.1 Method

The same method as the Automated Card Sorting was used in this study with an exception of the participants. In this study, there was a total of 31 participants (32 female, 19 male, 2 neither, age range 19-35, mean age 22), nine were Dutch, 20 were German, and three were from other nationalities.

5.2.2. Results

The results of the peer to peer Card Sorting are presented in a heatmap (Figure 17) and dendrogram (Figure 18).

5.2.2.1. Heatmap

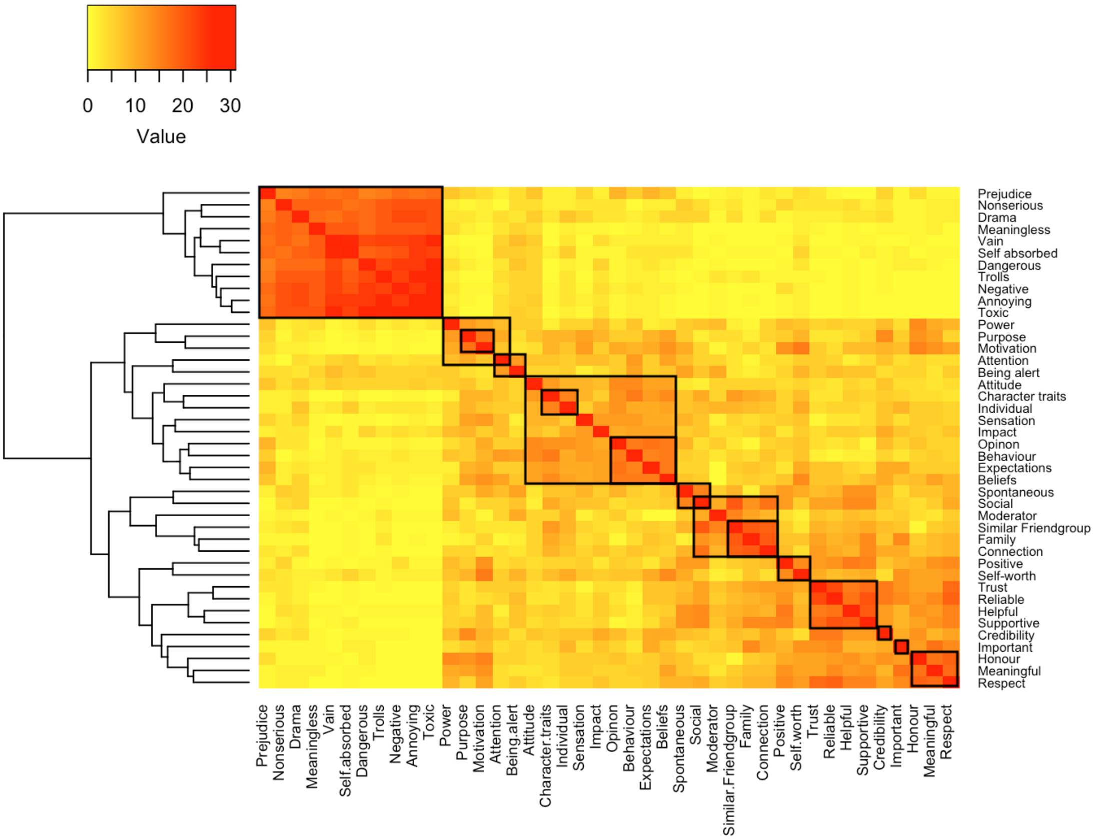


Figure 17. Heatmap with clustered items for the peer to peer reputation category. The black rectangles underline the different clusters and subclusters. There are eleven clusters and four subclusters.

The heatmap presents the distances between the items for the peer to peer reputation system. The obtained data shows that the strength of the association of the words ranges between 1 and 28, with one being the weakest association and 28 the strongest. In the heatmap eleven clusters with additionally, four sub clusters are presented. It can be seen that three clusters have smaller groups within a big cluster. Furthermore, there are two 2x2 squares where one of the words in the square might also belong to another bigger cluster. The first 2x2 square consists of the words *Being Alert* and *Attention*. The word *Attention* might also belong to a 4x4 square containing the words *Power, Purpose* and *Motivation*. The second 2x2 square consists of the words *Spontaneous* and *Social*. The word *Social* might also belong to the 5x5 square underneath the 2x2 square. In the middle of the heatmap, there is a darker 9x9 square. Within this square, there is a darker 2x2 square which consists of two words I*ndividual* and *Character Traits*. In the same 9x9 square there is one 4x4 square. This scheme appears a few times in the heatmap. Almost every bigger group has some darker spots within, which indicates that there are some smaller groups inside of those bigger groups. The first square in the topleft corner

11x11 has a very dark colour and the squares around it have a very light colour. That means the words in this square have a strong association. The last four clusters on the right side of the diagonal have a lot of darker bleeding spots around them, which makes it difficult to pin down specific groups. The groups are somewhat unclear, and clustered words also have a high association with other words around them. For example, the word *Respect* which belongs to the last cluster together with the words *Honour* and *Meaningful*. However, to the left and upwards, there are a few darker orange spots that shows that *Respect* does also have a relatively strong association with the words *Trust*, *Reliable* and *Helpful*.

5.2.2.2. Dendrogram

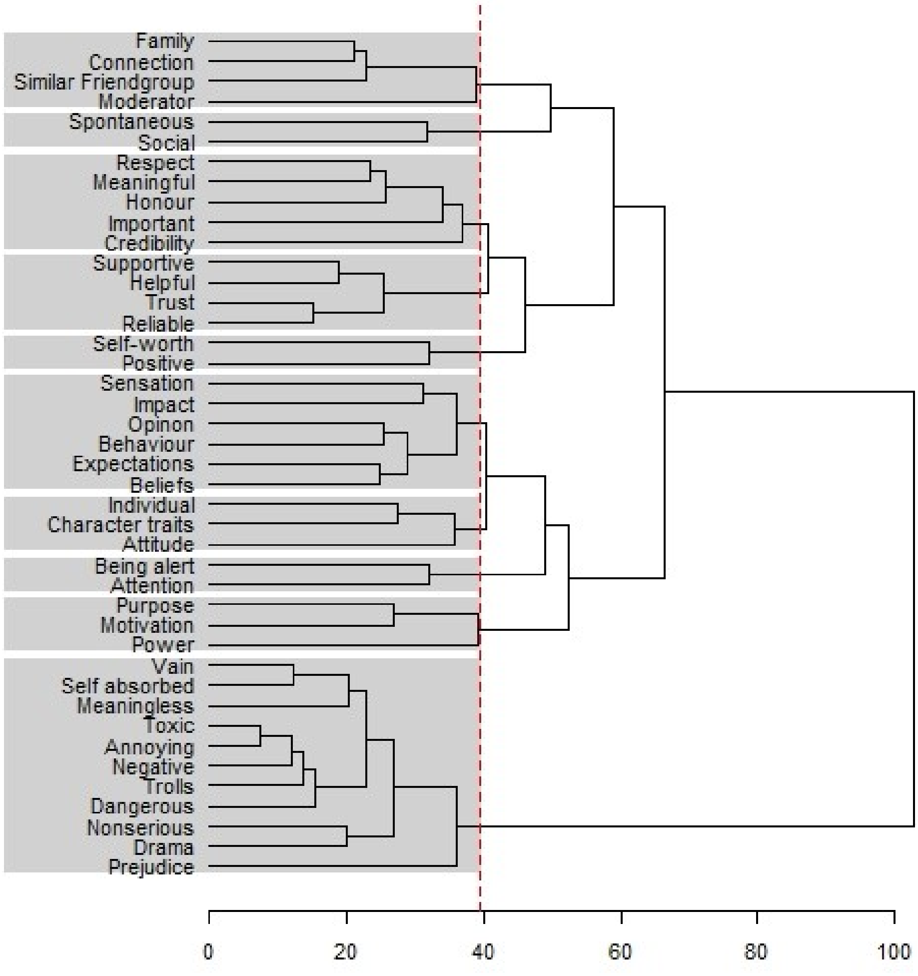


Figure 18. Dendrogram with 10 clusters (see grey rectangles) for the peer to peer reputation system category. The red line indicates where the dendrogram was cut.

The hierarchical cluster structure in figure 18 starts with clusters of one or two words at the left and ends with two overall clusters at the right. The elbow method proposes 8-12 clusters and the silhouette method suggests around 6 clusters. The graph presented by the elbow method is somewhat unclear. It does not show a clear angle. Carefully examining the dendrogram and taking both methods into consideration, the line was set at 39, which leaves us with 10 overall clusters. From 39 on bigger jumps can be seen in the distances of the clusters. This means that the association of words get approximately weaker at the right side of the line. The grey rectangles in the figure underline the 10 clusters. Three of the contained clusters consist of two words. The number of words in the other clusters range between 3 and 11.

5.2.2.3. Tentative Cluster Structure

The tentative cluster structure displays possible groups that can be used to develop a peer to peer reputation system. The groups with subgroups were created based on the dendrogram and heatmap (figure 17, figure 18). The blue rectangles mark groups created with the obtained data from the dendrogram, and the green rectangles are obtained from the heatmap. The mixture of displaying both clusters from the heatmap and dendrogram creates seven big groups with additionally 15 subgroups. Group 1 has three subgroups. Groups 2, 3 and 5 have four subgroups. Groups 4, 6 and 7 do not have any subgroups. In the following, the groups are discussed in more detail.

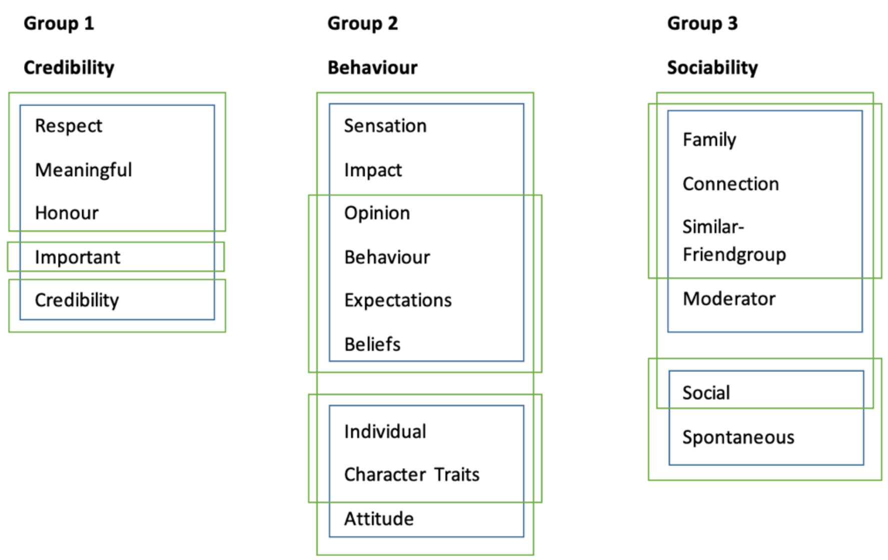


Figure 19. Group 1, 2 and 3. Tentative clusters for a peer to peer reputation system presented in a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows clusters obtained from the heatmap.

Group 1 consists of five words: *Respect, Meaningful, Honour, Important* and *Credibility*. The word which describes this group the best is **Credibility**. A person that others trust and believe in often is highly respected. Being respected and trusted can be an honour. In the context of online social communities, *Credibility* might be important to have in order to build a safe network. To measure credibility, only meaningful peer reviews should be taken into consideration. The subgroup consisting of the words *Respect, Meaningful* and *Honour* could describe **Respect.** For example, in a social online community, people should respect each other, have meaningful interaction and demonstrate honourable behaviour. In the context of the reputation system, this could mean that it should be looked at whether somebody follows those guidelines.

The second group consists of nine words: *Sensation, Impact, Opinion, Behaviour, Expectations, Beliefs, Individual, Character Traits* and *Attitude*. It is difficult to find something that all words have in common. Nevertheless, a very general term that could describe all the words is **Behaviour**. People hold certain opinions and have certain expectations towards one another. These opinions and expectations often come from what they believe is true. That can influence their attitude towards one another. One person might have a bad attitude towards the other person, but whether he acts upon it can be influenced by their character. If they are very peaceful, they might not act on it and behave friendly towards the other person, nonetheless. Two people with the same personal preference might be more inclined to meet one another. The subgroup consisting of the words *Opinion, Behaviour, Expectations* and *Belief,* can be best summarised by the term **Beliefs**. The Beliefs one person holds influence their opinion, expectations and behaviour. The other subgroup (*Individual*, *Character Traits*, *Attitude*) fits best with the word **Attitude**.

Group 3 consists of six words: *Family, Connection, Similar Friend group, Moderator, Social* and *Spontaneous.* In general, a fitting term could be **Sociability**. A social person has many relations to others and possesses skills for social interactions. *Moderators* can also make recommendations on users' behaviours and their connection with others. One subgroup consists of the words *Social* and *Spontaneous.* This best describes the **Social Skills**. A social person often is spontaneous and reacts to harmful content fast and efficiently in a positive way. The word *Relations* can best describe the other subgroup (*Family*, *Connection*, *Similar Friendgroup*, *Moderator*).

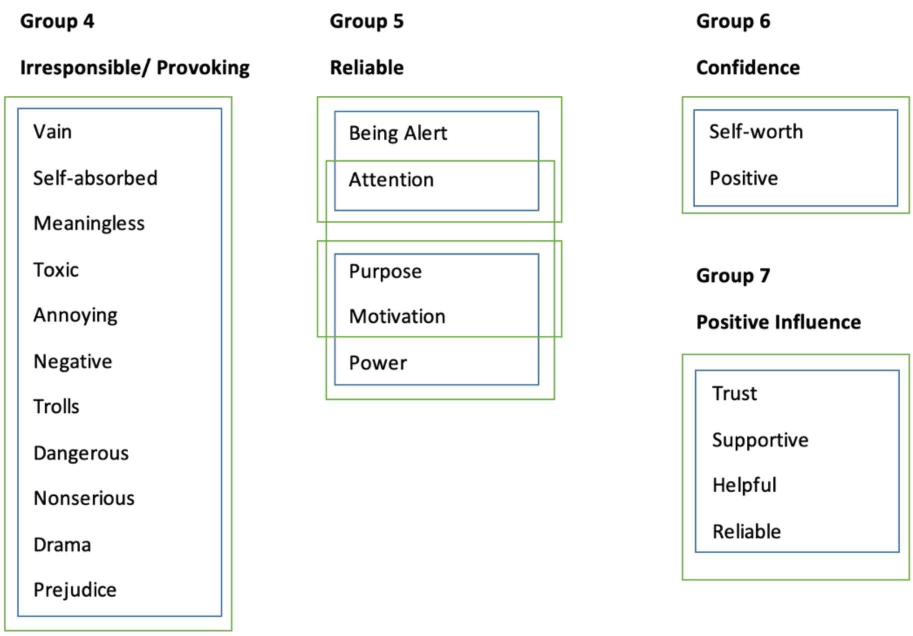


Figure 20. Group 4, 5, 6 and 7. Tentative clusters for a peer to peer reputation system presented in a tentative cluster structure. Blue are clusters indicated by the dendrogram and green shows clusters obtained from the heatmap.

The fourth group consists of 11 words: *Vain, Self-absorbed, Meaningless, Toxic, Annoying, Negative, Trolls, Dangerous, Non-serious, Drama* and *Prejudice*. All words in this group can be associated with negative behaviour. The overall term could be **Irresponsible** or **Provoking**. Words in the group could be split further. *Vain* and *Toxic* behaviour can be often found by *Trolls* and often leads to *Drama*. It can also be an *Endangerment* of another person's reputation. Trolls are often seen as annoying and are known for their meaningless negativity. Some users also might be *Self-absorbed* and only think about their gain.

Group 5 consists of five words: *Being Alert, Attention, Purpose, Motivation* and *Power.* The term representing all the words is **Reliable**. A reliable person pays attention to others, is there to motivate others if needed and is always alert. The subgroup containing the words *Being Alert* and *Attention*, fit with the term of **Responsibility**. The other subgroup can be summarised by the word **Trustworthy**.

The next group (Group 6) includes two words: *Self-worth* and *Positive.* These words can be summarised by the term of **Confidence**. Somebody can be aware of himself positively or negatively.

Group 7 consists of four words: *Trust, Supportive, Helpful* and *Reliable*. All words fall under the **Positive Influence** of a social community member. A member should be reliable and trustworthy and supportive and helpful towards others. That often depends on somebodies **Nature**.

5.3 Discussion

The goal of the Automated and Peer to Peer Card Sorting was to find out whether the found constructs have a meaningful structure by looking at the reputation system categories peer to peer and automated separately. For both categories, several groups were found; 18 possible groups for the automated domain and 13 groups for the peer to peer domain. The questions that remain for both domains are: (1) Can these groups be used for a reputation system and if yes (2) how would such a system look like?

To answer both questions, first a decision needs to be made on whether one or two reputation systems are needed. As mentioned earlier, automated systems collect data from different individuals, analyse it and display it in a way that is understandable for the users. Contrary, peer to peer reputation systems rely on data given by a peer in order to display an individual's reputation. Thus, both systems use different data inputs to generate value. An automated system puts out statistical data and is objective. In contrast, a peer to peer one is based on peer opinions and is subjective. Additionally, the peer to peer system uses a three-step process.

First peers need to evaluate another peer. Secondly, the obtained data from the evaluations need to be analysed, weighed, and thirdly a value needs to be calculated. Still, both systems should be presented together at the user's profile, as they might complete each other. Therefore, two systems, one for the automated and one for the peer to peer category, are needed. However, in terms of data input and scope, there is a difference. That means there are at least two reputation system categories for online social communities. The automated system has similarities with the rating system category, and the peer to peer system has similarities with the ranking system category from Jensen et al. (2002). Whether and how organisations can use the found set of constructs, will be explored separately for both systems. In the following, the automated reputation system is explored further.

5.3.1 An Automated Reputation System

The constructs found for the automatic reputation system categpry are *Activeness, Activity, Network, Engagement, Commonness* and *Content*. The table displays the constructs and subconstructs abstracted from the tentative cluster structure for automated reputation systems. The third cluster group was not used for this table as it displays Reputation System Goals and not measurable constructs for reputation. However, the constructs from this groups were used for the building and guiding process of the reputation system design.

Table 2

Constructs and Subconstructs as the basis for the automated reputation system.

Constructs	Subconstructs	Words
Activeness	Participation	Involvement. Social Engagement
	Socialising	Network, Connections
	Productivity	Developing, Efficient, Engagement
Activity	Content Generation	Like, Comment
	Profile Activity	Profile Status and Activity, Activity, Number of Posts
Network	Reach	Reach, Follower, Influencer, Fake, Influence, Verified, Status
	Type of User	Popular, Fame, Sponsors, Stars
	Circles	Wealthy, Money
Engagement	Profile Level	Ratings, Points, Profile Level
	Achievements	Badges, Achievements, Rewards
Commonness	Labels	Values, Labels
	Comparison	Compare, Differences, Similarities
Content	Created Content	Blog, Content, Memes, Pictures
	Viewed Content	Blog, Content, Memes, Pictures

The constructs and subconstructs in table 2 are parts of reputation. The words describing the subconstructs symbolise the data input that is being used to obtain values for the scores. To find out whether these groups (constructs with subconstructs) can be used for a reputation system, it is first discussed how an automatic system works: Firstly, the system collects data. Secondly, the collected data is evaluated and lastly displayed on a user's profile. For this process, an algorithm is used. An algorithm takes in data and uses logic to generate an output. Here, the algorithm would collect data suggested by the constructs - like *number of posts* - calculates a value and logically displays the value on the user profile.

Most of the constructs from the table represent numerical data. In other words, the constructs can be measured by using numbers like *Numbers of Posts, Number of Likes* or

Comments. However, there is also another type of data. Subconstructs like *Type of User* and *Values* cannot be measured by a number. They can only be measured as labels. Categorical data consist of labels and classifications. That means there are two different kinds of data inputs and therefore, the system needs to use two different kinds of algorithms.

One algorithm uses numbers and calculates values from those numbers which are then outputted as scores (see figure 21A). The other algorithm takes in different words related to the construct, calculates the most frequent or fitting words and displays them on the user's profile (see figure 21B). The construct **Commonness** and **Content** and their subgroups consist of categorical data, and the subconstruct *User Type* is also categorical. The rest of the constructs use numerical data. Therefore, the constructs can be used for a reputation system.

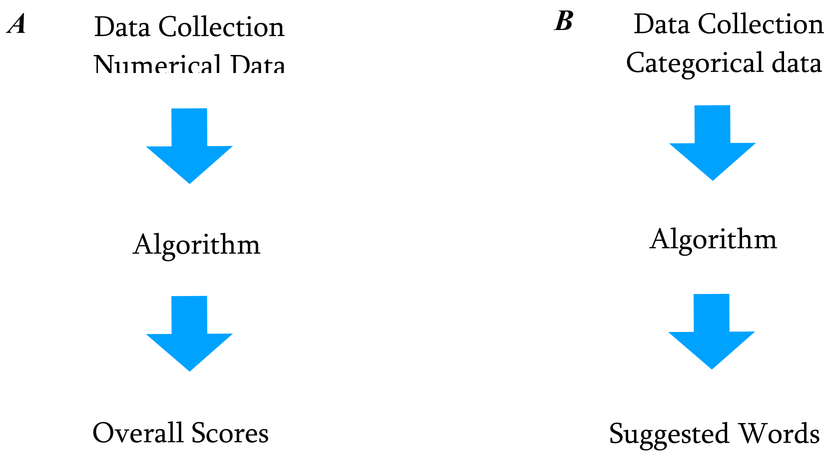


Figure 21. (A) Algorithm Numerical Data, (B) Algorithm Categorical Data.

To illustrate how the discovered constructs can be used in order to design a reputation system, an example is provided. The constructs **Activity** and **Commonness** are used for the example. For the numerical data, for this example, there is a system in place where a score from 1-10 is little a score from 11-20 is middle and 21-30 is high activity. Later every community needs to choose this depending on how active their users are in general. These scores are only used for this example. **Activeness** has three subconstructs: *Participation, Socialising* and *Productivity*. First, the system collects data on all three subconstructs. **Socialising** is used as an example to explain the data collection in more detail. The system collects data on how active a person was in his network and how many active connections the person had. The data is not only measured once but frequently. Importantly, only active networks and connections are taken into consideration. For measuring how active a user was in their networks, data on interactions, likes and posts is collected. The algorithm collects the number of active connections and the level of activeness of the user. After that, it calculates an overall score for the suconstruct. In this example, the score is 15 for the current week and was 22 the week before. In figure 22A the score for the previous week and 22B the score for the current week are displayed.

Figure 22. Example of scores for the subconstruct Socializing. The blue indicates the amount of social activity.

However, the data output only shows how active a user was but not if the content that had been put out (posts, comments) was appropriate and helpful for the community. But the point of having a reputation system is to see in what way somebody behaves. That is why, additionally to this algorithm, the categorical algorithm can be used to find either positive or negative connotations in the data. Based on that and the numerical data, the activity can be either negative or positive. In the **Socialising** example, this means if a person was active in his network but only acted in a bad manner, he or she will have a negative score of 15 (see figure 23A). If the person acted appropriately at times but not properly at others, positive and negative connotations are detected, and the bar goes both ways (see figure 23B). The bar is like a scale: zero is in the middle, plus (positive connotations) goes to the right side and minus (negative connotations) to the left side. In the example, it can be seen that the negative score is dark blue, and the positive score is light blue.

Figure 23. Example of scores for the subconstruct Socializing with negative ratings. Dark blue shows the amount of negative behavior (negative socializing). Light blue shows the amount of positive behaviour.

Commonness consists of only categorical data (labels and classifications). Therefore, the categorical algorithm is needed. The subconstruct **Comparison** is used to explain how data is measured. Firstly, all the data on different *Labels* of users is collected. After that, the system compares *Similarities* and *Differences*, for instance, inactiveness, networks, connections and overall online behaviour with the same data on other users. The system then suggests the users that are the most similar to that person. In the following, an example is given: Jane Smith is a member of the Breed Dog Group and a Chess Club Group. She is very active and posts a lot over her dog Berry, who is an Australian Shepherd and she also likes to write blogs over new chess strategies. The algorithm will search for words like chess, dog, breeding, Australian Shepherd and chess strategies and compare it with users with similar words. The most fitting users will be displayed on the profile (see figure 24). Figure 25 shows how the reputation could be displayed on the user's profile. It puts all parts explained above together.

Jane Smith

Similar Users: John Roberson, Steve Bose

Figure 24. Suggestion for similar users.

Jane Smith

Type of User: Influencer, Blogger
Labels: Helpful, Informative
Similar Users: John Roberson, Steve Bose
Created Content: 10 Blogs, 5 Pictures, 2 Memes
Viewed Content: 100 Memes, 2 Blogs, 300 Pictures

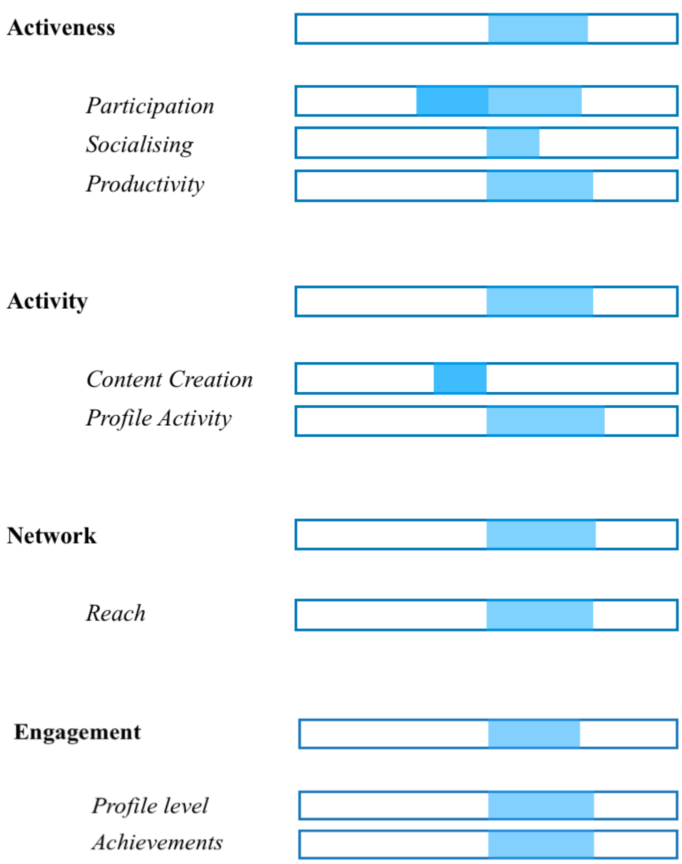


Figure 25. Example of a user’s reputation profile. Dark blue indicates bad activity. Light blue presents positive activity. The bluer the more activity.

5.3.2. A Peer to Peer Reputation System

The constructs found for the peer to peer reputation system are *Credibility, Behaviour, Sociability, Irresponsible/ Provoking, Reliable, Confidence* and *Positive Influence*. The table displays the constructs with subconstructs abstracted from the tentative cluster structure for peer to peer reputation system.

Table 3

Constructs and Subconstructs as the basis for the automated reputation system.

Constructs	Subconstructs	Words
Credibility	Respect	Respect, Meaningful, Honour, Important, Credibility
Behaviour	Beliefs	Beliefs, Expectations, Behaviour, Opinion, Impact, Sensation
	Attitude	Individual, Character Traits, Attitude
Sociability	Relations	Family, Connection, Similar Friendgroup, Moderator
	Social Skills	Social, Spontaneous
Irresponsible/ Provoking	-	Vain, Self-absorbed, Meaningless,Toxic, Annoying, Negative, Trolls, Dangerous, Nonserious, Drama, Prejudice
Reliable	Responsible	Being Alert, Attention
	Trustworthy	Purpose, Motivation, Power
Confidence	-	Self-worth, Positive
Positive Influence	Nature	Trust, Supportive, Helpful, Reliable

The words at the right in table 3 describe the subconstructs. The subconstructs are part of the constructs, and the constructs are parts of reputation. First, it needs to be discussed how a peer to peer system works in general, in order to answer the question of whether the found constructs can be used for a reputation system. First of all, peers evaluate another member. After that, peer feedback is collected. This feedback is then weighed and analysed by, for example, an algorithm. Lastly, the results are displayed on the members' profile. This means there are three processes in place. (1) There is a scale where peers can evaluate other peers (2) There is an algorithm that calculates the weight of every evaluation and (3) an algorithm that calculates an overall value/score that can be displayed on the profile. In the following section, an example is given.

First, a decision needs to be made on what evaluation tool is used. For this example, a bar is used the same as in the automation system example. The category **Behaviour** is used to explain the process in more detail. First, a peer gives a rating to another peer (see figure 26)

Figure 26. Example of how a user rates another user. Light blue indicates a positive rating. Dark blue presents a negative rating.

The system weighs the given rating. First, it searches for clues on what connection both peers have, what their *Beliefs* are, and what kind of relationship they have in general. If the two persons are the opposite from each other and they are known to get into fights, the evaluation will be changed accordingly. Contrary if the peer has a good reputation himself and is known for his positive influence on the community, the evaluation is weighed more. After the weight is determined, the algorithm calculates an overall score for the evaluated member. For a detailed view on how this could look like on the user's profile see figure 27.

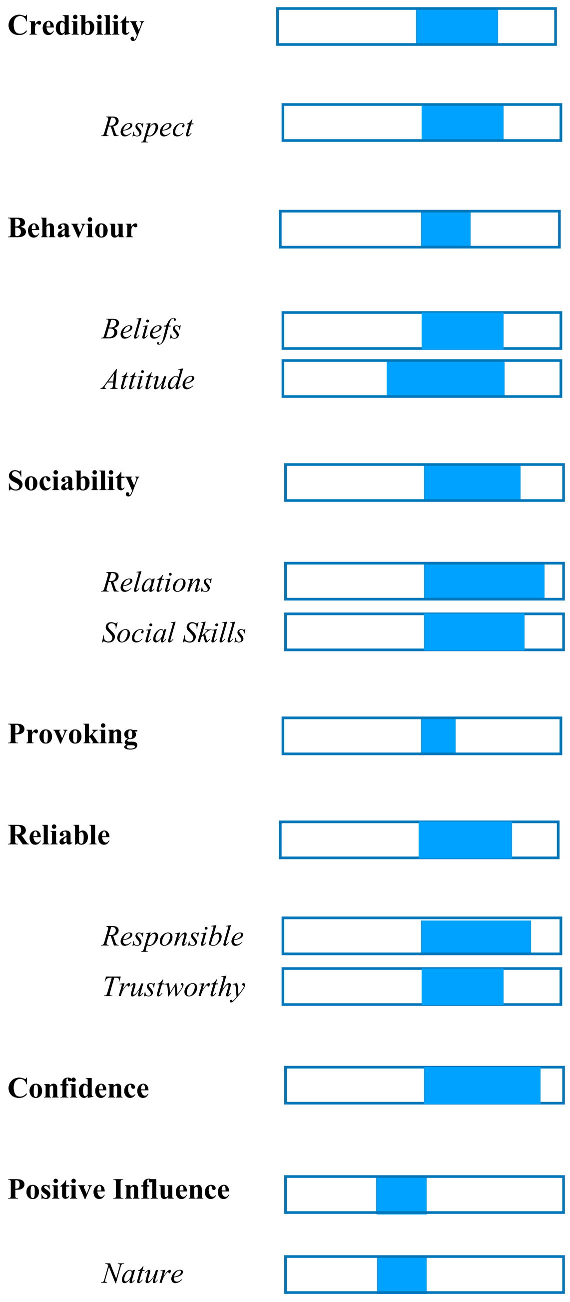


Figure 27. Example of how peer to peer-based reputation can look on a user's profile. Dark blue indicates negative ratings and light blue represents positive ratings.

It can be concluded that for both systems, the discovered set of constructs can be used in order to build and design a reputation system.

6. General Discussion

Uncertain credibility and reliability between members of online social communities make individuals hesitant to get actively involved in online communities or can put them in danger if they do get involved. Sooner or later, this can lead to inactive communities. As a consequence, possibilities given by the online space cannot be exploited, and organisations are left with failed communities. In this study, two reputation systems based on a set of constructs are introduced to help solve this problem.

6.1. Using the Concept Reputation to Evaluate Individuals Online

Past research suggested that reputation systems can act as an evaluation tool. However, research mainly focussed on integrated reputation systems for e-commerce, where the focus lies on evaluating a product or service and not individuals. That is why a reputation system specially developed for online social communities was needed. In order to develop such a system, we needed to find out first what reputation meant in the context of online social communities, how individuals decided whether someone has a good or bad reputation and how this can be transformed into a reputation system. To do so, we decided to explore the concept of reputation further. The problem was that we did not entirely know how information is processed to assess someone's reputation, as it is a very abstract concept. Consequently, reputation could not readily be transformed into a measurement scale. In order to solve this problem, we looked into what an abstract concept is and how it can be measured. We found out that by unfolding the mental model of reputation we could find underlying structured constructs of reputation that can serve as a basis for reputation systems. To unfold constructs of reputation we conducted a Word Association. After that, we used Card Sorting to find a structure. In the following, the findings of all three studies are presented and discussed briefly.

6.2. A Set of Constructs as a Basis for Online Social Reputation Systems

The Word Association (first part of the study) resulted in a list of 159 words associated with reputation. One hundred of these words were used for the Pilot Card Sorting (second part of the study), to get a first broad picture of the mental models of reputation. Results from the Pilot Card Sorting suggested that there might be at least two different reputation system categories. One for automated systems, where data is collected automatically and a peer to peer system, where individuals evaluate each other. Ergo, constructs of two mental models might have been mixed in the Pilot Card Sorting. That could have led to the confusion of participants and to a semantic map that does not represent the mental model correctly. Therefore, a second Card Sorting (third part of the study) was conducted, where the found constructs were sorted for both categories separately. The main Card Sorting resulted in two tentative cluster structures, one for each category with several meaningful groups and subgroups, that could potentially be used to measure reputation.

The groups and subgroups for both categories represent constructs of reputation. The words in those groups describe what needs to be explicitly measured for every part. Names chosen for the groups and subgroups are meant to represent all the words inside the specific groups. The following set of constructs were found for automated reputation systems; *Activeness, Activity, Network, Engagement, Commonness* and *Content.* The constructs *Credibility, Behaviour, Sociability, Irresponsible/ Provoking, Reliable, Confidence* and *Positive Influence* were found for the peer to peer reputation systems.

All findings in the study are significantly important because the three parts of the study are built on one another. The findings from the first part of the study are used for the second and third part of the study. The findings from the second part of the study are used for the third part of the study. The major findings of all three studies are the structured constructs and subconstructs with the related words resulting from the two tentative cluster structures found in the third part of the study (see figure 15, 16 and figure 19, 20). The discovered constructs and subconstructs break down the abstract concept reputation, into smaller, measurable pieces, making it possible to use reputation as a tool that can evaluate members in online social communities.

In other words, what makes these findings (constructs and subconstructs) especially important is that these set of constructs can be used as a basis to build different reputation systems that members can use to evaluate different individuals online. However, it is not proven yet that in practice, these systems will work. Thus, the research question *'Can a set of constructs be found that could be used as the basis for developing a reputation system?'* can be positively answered because not only one but two set of constructs could be found that can potentially be used as a basis for creating reputation systems for two reputation categories (peer to peer and automated). However, the limitation is that it still needs to be tested out in a real-time online environment.

6.3. Using Card Sorting to Design a Reputation System

Card Sorting has been used in the past to unfold underlying structures in order to design user-centred questionnaires and navigation structures for websites. This study is a hybrid in a sense that first a concept (reputation) is investigated in order to create a design later on. In the following, it is discussed shortly why Card Sorting was used, if it was successful and whether this method can be recommended for further use.

The primary goal was to design a system. However, in order to design that system, we needed to find a set of flexible constructs. These constructs should ideally reflect the users' idea of what reputation is in the context of online social communities. This is why Card Sorting was chosen as a method although it is typically not used for this kind of design. By using Card Sorting as a method we were able to unfold the mental model of reputation of users', and thereby obtain two sets of constructs that we used to design two reputation systems. Thus, Card Sorting was integrated successfully as a method in our study and helped us to build a solid basis for future reputation systems. In general, it can be said that Card Sorting is a great method to use when (1) a hands-on user-friendly system design needs to be developed and (2) the given concept is abstract and measurable constructs and their structure cannot be abstracted easily from it or out of existing literature. It is also important that a list of words associated with the concept is obtained beforehand either from literature or by using a method like word association. Furthermore, Card Sorting can only be fully recommended for this specific kind of design and it still needs to be tested out for other designs to test if its generalizable.

6.4. Using E-commerce Reputation Systems for Online Social Community Reputation Systems

In the discussion section of the automated and peer to peer Card Sorting, two reputation systems one for the automated and one for the peer to peer category were presented. Both systems give a first idea on the usage of the set of constructs and subconstructs for the design process. Furthermore, they show how underlying mechanisms of such systems could look like. The automated reputation system uses an algorithm in order to calculate a value that organisations can use to rank the different parts of an individual's reputation. The peer to peer reputation system makes use of both a rating and a ranking system. Members can use the rating system to evaluate other members of the community. The system then ranks the different evaluations and calculates an overall value.

Jensen et al. (2002), present different reputation system categories in their research. Comparing the characteristics of the found system categories (peer to peer and automated) with the characteristics of the suggested categories (Ranking System, Rating System, Collaborative Filtering System and Peer based system), it meets the eye that the designed automated system might fall partially under the category ranking system and the peer to peer system might fall partially under the rating system category regarding the rating process. Furthermore, it appears that based on every category suggested by Jensen et al. (2002), an automated or peer to peer reputation system can be built using the discovered constructs, depending on what purpose the system should serve. Therefore, it can be concluded that systems for both community types (e-commerce and online social communities) might be relatively similar in appearance and strategies used with the difference that the content and the constructs used to measure reputation are completely different for both of them. Consequently, parts of reputation systems introduced by e-commerce could be used in order to build a reputation system for online social communities, given that it fits with the measurable set of constructs found in the study. The advantage of this is that systems do not need to be built from the start. Algorithms and rating mechanisms developed for e-commerce could be used as a starting point for online social communities and might even be transferred directly with a few small changes.

6.5. Limitations

One limitation of the study is that the words found to be associated with reputation were not scientifically double-checked. The danger of not double-checking is that all the results from the second and third part of the study also become invalid. A reversed association could have been conducted to double-check the obtained words. Still, the words obtained by the study were double-checked by three rater in the literature research meeting, in order to be sure that they indeed represent the concept of reputation.

Another limitation is that mental models, in general, are incomplete. Mental models are simplified ideas of how we see the world. We break down complicated concepts in a way we understand. As a consequence, there might be even more constructs representing reputation than the ones that we found. Nevertheless, the study gives an idea of several possible constructs and thus draws a broad picture of the mental model of reputation.

The last limitation is that the study only draws a theoretical idea on reputation. Real-life issues are not taken into account yet. One potential issue could be, for example, data privacy. The problem is that the algorithm uses personal data of user's and displays it on the profile.

Therefore, the data needs to be displayed in a way that data privacy is still in place. Furthermore, it needs to be ensured that the system does not misuse the user's data. Another issue is that it can be difficult to programme these smart algorithms. A solution could be to use existing algorithms, for example from e-commerce systems, so the developer does not need to start from scratch. Montes et al. (2017), for example, proposed HFLTS to evaluate the reputation of individuals in online social communities. They used an algorithm that analyses content in the community, including connotations, in order to evaluate users. This algorithm could potentially be used for an automatic reputation system.

6.6. Future Research

The implementation of evaluation systems into online social communities brings along several risks. One risk is that it could pose a threat to data privacy. In order for a reputation system to work, the system takes in data. Data with information over the actions of individuals that are going to be pinned to their profile. In the future, it needs to be investigated, in how far these systems interfere with privacy rules and how this can be avoided. Another risk is that first and foremost peer to peer reputations systems are going to be misused. People can, for instance, group together in order to ruin or boost someone's reputation. Safety systems need to be integrated to prevent this from happening. In general, what needs to be investigated further at this point is how these theoretical results will work in practice. That means research needs to be done on what parts of existing systems can be used for online social communities and on fool proof algorithms for these systems.

7. Conclusion

It can be concluded that there are at least two reputation system categories peer to peer and automated. For both system categories the discovered sets of constructs can theoretically be used to build at least two reputation systems to evaluate members in online social communities. The designed automated and peer to peer reputation systems are only one example of how such systems could look like. The discovered constructs that were presented by table 2 and table 3 can be used to build more reputation systems. This study acts as a starting point by presenting two sets of constructs that can be used as a basis for a social online reputation system. Additionally, research on data privacy and fool-proof systems needs to be conducted.

References

Ba, S. (2001). Establishing online trust through a community responsibility system. *Decision Support Systems*, *31*(3), 323–336. doi:10.1016/s0167-9236(00)00144-5

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating Abstract Concepts. *Grounding Cognition*, 129–163. doi: 10.1017/cbo9780511499968.007

Bishop, J. (2007). Increasing participation in online communities: A framework for human–computer interaction. *Computers in Human Behavior, 23*(4), 1881-1893. doi:10.1016/j.chb.2005.11.004

Borderless Technology Corp. (2018, February 9) [BTC]. The Importance and Psychology of Reputation In Human Lives [Blog Post]. Retrieved from https://medium.com/@Borderless/the-importance-and-psychology-of-reputation-in-human-lives-c72362393c91.

Carbo, J., Molina, J. M., & Davila, J. (2003). Trust Management Through Fuzzy Reputation. International Journal of Cooperative Information Systems, 12(01), 135–155. doi:10.1142/s0218843003000681

Chermack, T. J. (2003). Mental Models in Decision Making and Implications for Human Resource Development. *Advances in Developing Human Resources*, *5*(4), 408–422. doi:10.1177/1523422303257373

Clear, J. (n.d.). Mental Models: Learn How to Think Better and Gain a Mental Edge. Retrieved from https://jamesclear.com/mental-models.

Dellarocas, C. (2010a). Designing Reputation Systems for the Social Web. *The Reputation Society*. doi:10.7551/mitpress/8777.003.0005

Dellarocas, C. (2010b). Online Reputation Systems: How to Design One That Does What You Need. *MIT Sloan Management Review, 51*, 33-38.

Falor, R. A., Feit, R., Boyington, M., Morgan, M., Ziman, B. (2014). Reputation and Engagement System for Online Community Management. *United States Patent No. US,10,049,138 B1.* Retrieved from https://patents.google.com/patent/US10049138B1en

Fan, M., Tan, Y., & Whinston, A. B. (2005). Evaluation and Design of Online Cooperative Feedback Mechanisms for Reputation Management. *IEEE Transaction on Knowledge and Data Engineering, 17*(3), 244-254. doi:10.1109/TKDE.2005.26

Istifci, I. (2010). Playing with Words: A Study on Word Association Responses. *Journal of International Social Research, 3*(10).

Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience*, 5(2), 133–147. doi:10.1080/17470910903202559

Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience Research*, *72*(4), 283–288. doi:10.1016/j.neures.2012.01.003

Jensen, C., Davis, J., & Farnham, S. (2002). Finding others online. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Changing Our World, Changing Ourselves - CHI 02*. doi:10.1145/503453.503456

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*, *16*(1). doi:10.5751/es-03802-160146

Kawamichi, H., Sasaki, A. T., Matsunaga, M., Yoshihara, K., Takahashi, H. K., Tanabe, H. C., & Sadato, N. (2013). Medial Prefrontal Cortex Activation Is Commonly Invoked by Reputation of Self and Romantic Partners. *PLoS ONE*, *8*(9). doi:10.1371/journal.pone.0074958

Montes, R., Sanchez, A. M., Villar, P., & Herrera, F. (2017). A decision making model to evaluate the reputation in social networks using HFLTS. *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. doi:10.1109/fuzz-ieee.2017.8015519

Perrin, A. (2015). *Social Media Usage: 2005-2015.* Retrieved from the Pew Researcher Center website: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/

Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. (2000). Reputation Systems. *Communications of the ACM, 43*(12), 45-48. doi:10.1145/355112.355122

Schmettow, M., & Sommer, J. (2016). Linking Card Sorting to browsing performance – are congruent municipal websites more efficient to use?. *Behaviour & Information Technology, 35*(6), 452-470. doi:10.1080/0144929x.2016.1157207

Schrammel, J., Köffel, C., & Tscheligi, M. (2009). Personality Traits, Usage Patterns and Information Disclosure in Online Communities. *Proceedings of the 23rd British HCI, Group Annual Conference on People and Computers: Celebrating People and Technology,* 69-174. doi:10.14236/ewic/hci2009.19

Ulusu, Y. (2010). Determinant Factors of Time spent on Facebook: Brand Community Engagement and Usage Types. *Journal of Yaşar University, 5*(18), 2949 - 2957

Van der Velde, F. (2018). Vector analysis in R of concept clusters based on Card Sorting. CPE-BMS, University of Twente

Van der Velde, F., Wolf, R. A., Schmettow, M., & Nazareth, D. (2015). A Semantic Map for Evaluating Creativity. In H. Toivonen, S. Colton, M. Cook, & D. Ventura (Eds.), *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)* (pp. 94-101). Provo, UT, USA: Brigham Young University.

Xiong, L., & Liu, L. (2003). A reputation-based trust model for peer-to-peer ecommerce communities [Extended Abstract]. *Proceedings of the 4th ACM Conference on Electronic Commerce - EC 03*. doi:10.1145/779928.779972

Yao, G., TU, A., & Zaner-Godsey, M. (2009). *United States Patent No. US 2009/0327054A1*. Retrieved from https://patents.google.com/patent/US20090327054

Yu, B., & Singh, M. P. (2002). Distributed Reputation Management for Electronic Commerce. *Computational Intelligence, 18*(4), 535-549. doi:10.1111/1467-8640.00202

Zacharia, G., & Maes, P. (2000). Trust management through reputation mechanisms. *Applied Artificial Intelligence, 14*(9), 881-907. doi:10.1080/08839510050144

Appendix A

Informed Consent for the WA

You are being invited to participate in a research study titled ***Word Association study towards a Reputation System in Online Communities***. This study is being done by ***Jule Landwehr*** from the Faculty of Behavioural, Management and Social Sciences at the University of Twente.

The purpose of this research study is to ***find constructs related with reputation to build a reputation system that meets users' needs*** and will take approximately ***5 minutes*** to complete. The data will be used for the testing of my master thesis.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question. You need to be over 18 in order to participate in this study.

We believe there are no known risks associated with this research study; however, as with any online related activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks ***by treating the information confidentially. Datasets obtained by the study do not contain personal information and might be shared in their original form to inform future research or validate research results but will not be reported to serve any other goal than that of the research. Results will be reported in my Master thesis and will be made accessible via the University of Twente library service (https://essay.utwente.nl/).***

Study contact details for further information: Jule Landwehr, j.landwehr@student.utwente.nl

If you click on 'Next' you consent, and the study will start.

Appendix B

Word Association Test

UNIVERSITY OF TWENTE.

Please write down the first three words that immidiately come to your mind when you think about reputation of members in the context of social online communities. After writing down the words click on 'Next' to submit the questionnaire.

First word

Second word

Third word

Back

Next

Appendix C

List of words obtained from the Word Association

Score	Words
11	fake
8	likes
5	social
4	advertising medium
4	fame
4	followers
4	influencers
4	privacy
4	pictures
3	addiction
3	Friends/ Friend Group
3	Perfection/ Perfect
3	Trust/ Trustful
2	Achievements
2	Annoying
2	Blog
2	Hater
2	Image
2	Power/full
2	Public
2	Respect
2	Rewards
2	Sharing
2	Status
2	Supportive
1	Act
1	Active
1	Attention
1	Attitude
1	Badges

1	Beauty
1	Being alert
1	Beliefs
1	Blogger
1	Chance
1	Censorship
1	Comments
1	Compare
1	Connection
1	Content
1	Convincible
1	Credibility
1	Dangerous
1	Data collector
1	Developing
1	Difference
1	Disadvantage
1	Distant
1	Drama
1	Efficient
1	Engagement
1	Expectations
1	Facebook
1	Family
1	Female
1	FOMO
1	Food
1	Gamers
1	Helpful
1	Honour
1	Illusion
1	Impact
1	Important
1	Individual
1	Instagram

1	Interactive
1	International
1	Involved
1	Label
1	Long hair
1	Mainstream
1	Manipulation
1	Marketplace
1	Meaningful
1	Meaningless
1	Memes
1	Moderator
1	Money
1	Motivation
1	Mutuality
1	Negative
1	Negativity
1	Network
1	Non Serious
1	Number of Posts
1	Opinion
1	Points
1	Popular
1	Positive
1	Purpose
1	Prejudice
1	Profile activity
1	Profile level
1	Profile status
1	Promotions
1	Rating
1	Reach
1	Safety
1	Self-absorbed
1	Selfie

1	Self-worth
1	Sensation
1	Social Engagement
1	Social Media
1	Sponsors
1	Spontaneous
1	Stars
1	Toxic
1	Trolls
1	Universal
1	Unnecessary
1	Vain
1	Value
1	Wealthy
1	Worldwide
1	Character traits
1	Hearsay
1	Similar Friends
1	Gamification
1	Activity
1	Verification
1	Similarities
1	Advantages
1	Rating on services
1	Influence
1	Connections
1	Truth
1	Behaviour
1	Consistency
1	Age
1	Creditworthy
1	Reliable
1	Loved
1	Liked
1	Reputable

1	Right
1	Wrong
1	Factual
1	Positive
1	Negative
1	Emphatic
1	Interesting
1	Source
1	Creative
1	Eccentric
1	Praised
1	Expensive
1	Worthwhile
1	Rated
1	Gained
1	Lost
1	Spam
1	Audited
1	Checked
1	Confirmed

Appendix D

Informed Consent of the Card Sorting

Opening Statement for an Online Survey

You are being invited to participate in a research study titled ***Card Sorting Study towards a Reputation System in Online Communities***. This study is being done by ***Jule Landwehr*** from the Faculty of Behavioural, Management and Social Sciences at the University of Twente.

The purpose of this research study is to ***find meaningful clusters regarding reputation*** and will take you approximately ***20*** minutes to complete. The data will be used for the testing of my master thesis.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question.

We believe there are no known risks associated with this research study; however, as with any online related activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks ***by treating information confidentially. Datasets obtained by the study do not contain personal information and might be shared in their original form to inform future research or validate research results but will not be reported to serve any other goal than that of the research. Results will be reported in my Master thesis and will be made accessible via the University of Twente library service (https://essay.utwente.nl/).***

Study contact details for further information: Jule Landwehr, j.landwehr@student.utwente.nl

Appendix E

List of words used for the Card Sorting

Words
Achievements
Active
Annoying
Attention
Attitude
Badges
Behaviour
Being alert
Beliefs
Blog
Censorship
Character traits
Comments
Compare
Connection
Connections
Consistency
Content
Credibility
Dangerous
Data collection
Developing
Difference
Disadvantage
Distant
Drama
Efficient
Engagement
Expectations
Fake
Fame

Family
Follower
Helpful
Honour
Illusion
Image
Impact
Important
Individual
Influence
Influencer
Interactive
Involved
Label
Like
Manipulation
Meaningful
Meaningless
Memes
Moderator
Money
Motivation
Mutuality
Negative
Network
Nonserious
Number of Posts
Opinion
Perfect
Picture
Points
Popular
Positive
Power
Prejudice

Privacy
Profile activity
Profile level
Profile status
Public
Purpose
Rating
Reach
Reliable
Respect
Rewards
Safety
Self-absorbed
Self-worth
Sensation
Sharing
Similar Friendgroup
Similarities
Social
Social Engagement
Sponsors
Spontaneous
Stars
Status
Supportive
Toxic
Trolls
Trust
Universal
Vain
Value
Verified
Wealthy
Worldwide

Appendix F

Informed Consent of the Peer to Peer and Automated Card Sorting

You are being invited to participate in a research study titled ***Card Sorting Study towards a Reputation System in Online Communities***. This study is being done by ***Jule Landwehr*** from the Faculty of Behavioural, Management and Social Sciences at the University of Twente.

The purpose of this research study is to ***find meaningful clusters regarding reputation*** and will take you approximately ***10*** minutes to complete. The data will be used for the testing of my master thesis.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any question.

We believe there are no known risks associated with this research study; however, as with any online related activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks ***by treating information confidentially. Datasets obtained by the study do not contain personal information and might be shared in their original form to inform future research or validate research results but will not be reported to serve any other goal than that of the research. Results will be reported in my Master thesis and will be made accessible via the University of Twente library service (https://essay.utwente.nl/).***

Study contact details for further information: Jule Landwehr, j.landwehr@student.utwente.nl

If you **click** on begin, you **accept** the informed consent

Appendix G

Table of words for Peer to Peer, Automated and Neither

Rating	Automated System	Rating	Peer to Peer System	Neither	Named one time in all categories
1	Data collection	1	Credibility	Illusion	Consistency
2	Points	2	Reliable	Perfect	Disadvantage
3	Rating	3	Helpful	Universal	Distant
4	Badges	4	Supportive		Image
5	Achievements	5	Behaviour		Interactive
6	Rewards	6	Important		Mutuality
7	Like	7	Character Traits		Safety
8	Label	8	Social		
9	Stars	9	Meaningful		
10	Network	10	Purpose		
11	Profile level	11	Respect		
12	Compare	12	Positive		
13	Sharing	13	Family		
14	Value	14	Self-worth		
15	Verified	15	Beliefs		
16	Efficient	16	Individual		
17	Status	17	Spontaneous		
18	Profile activity	18	Trust		
19	Profile status	19	Moderator		
20	Engagement	20	Dangerous		

21	Censorship	21	Self absorbed
22	Number of Posts	22	Attitude
23	Follower	23	Negative
24	Social Engagement	24	Attention
25	Connections	25	Power
26	Fake	26	Impact
27	Money	27	Motivation
28	Reach	28	Being alert
29	Comments	29	Opinion
30	Involved	30	Expectations
31	Privacy	31	Annoying
32	Difference	32	Honour
33	Manipulation	33	Nonserious
34	Similarities	34	Drama
35	Content	35	Meaningless
36	Developing	36	Vain
37	Influence	37	Connection
38	Popular	38	Prejudice
39	Active	39	Sensation
40	Influencer	40	Toxic
41	Fame	41	Similar Friendgroup
42	Sponsors	42	Trolls
43	Wealthy		

44	Public
45	Picture
46	Worldwide
47	Memes
48	Blog

Appendix H

Syntax for the data analysis

Data analysis Pilot Card Sorting

```
install.packages('gplots')
install.packages('factoextra')
install.packages('dendextend')
install.packages('pheatmap')
library(gplots)
library(RColorBrewer)
library(tidyverse)  # data manipulation
library(cluster)    # clustering algorithms
library(factoextra) # clustering visualization
library(dendextend) # for comparing two dendrograms
library(pheatmap)

# Read the data file (.csv format)
data2 <- read.csv("Masterthesis/Analysis/General Card Sorting 1/Datasheet_analysis_2.csv", comment.char="#")
rnames <- data2[,1]


# Transform data in numerical format and give names
mat_data2 <- data.matrix(data2[,2:ncol(data2)])
rownames(mat_data2) <- rnames


# Define colors of heatmap: red for high numbers
my_palette <- colorRampPalette(c("yellow","red"))(n = 299)


#Heatmap & Dendrogram
dev.off()
heatmap.2(dendrogram = "row", mat_data2, key = TRUE, keysize = 1.0, col = my_palette, density.info="none", trace="none",
        revC = TRUE, cexCol = 0.4, cexRow = 0.4, margins = c(5, 5), offsetRow = 0.1,
        offsetCol = 0.1)
```

```
#Dendrogram vertikal
#par(mfrow=c(1,1))
hc.rows<- hclust(dist(mat_data2))
dend <- as.dendrogram(hc.rows)
par(cex=0.6, mar=c(10, 4, 10, 4))
plot(dend)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 20, lower_rect = -22, lty = 5, lwd = -1, col = rgb (0.1,0.2,0.4,0.1))
abline(h = 45, col = 2, lty = 2)

#Dendrogram horizontal
dend2 <- as.dendrogram(hc.rows)
par(cex=0.5, mar=c(4, 4, 4, 4))
hang.dendrogram
plot(dend2, horiz = TRUE)
plot_horiz.dendrogram(dend2, side = TRUE)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 9, lower_rect = -11, horiz = TRUE, lty = 5, lwd = -1, col = rgb
(0.1,0.1,0.1,0.2))
abline(v=49, col = 2, lty = 2)

wss <- function(k) {
  kmeans(mat_data2, k, nstart = 10 )$tot.withinss
}

# Compute and plot wss for k = 1 to k = 15
k.values <- 1:20

# extract wss for 2-15 clusters
```

```
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")

# function to compute average silhouette for k clusters
avg_sil <- function(k) {
  km.res <- kmeans(mat_data2, centers = k, nstart = 25)
  ss <- silhouette(km.res$cluster, dist(mat_data2))
  mean(ss[, 3])
}

# Compute and plot wss for k = 2 to k = 15
k.values <- 2:60

# extract avg silhouette for 2-15 clusters
avg_sil_values <- map_dbl(k.values, avg_sil)

plot(k.values, avg_sil_values,
     type = "b", pch = 19, frame = FALSE,
     xlab = "Number of clusters K",
     ylab = "Average Silhouettes")
```

Data analysis Automated Card Sorting

```
library(gplots)
library(RColorBrewer)
library(tidyverse)  # data manipulation
library(cluster)    # clustering algorithms
library(factoextra) # clustering visualization
library(dendextend) # for comparing two dendrograms
library(pheatmap)


# Read the data file (.csv format)
data_automated <- read.csv("Masterthesis/Analysis/Automated/Datasheet_analysis_automated.csv",
comment.char="#")
rnames <- data_automated[,1]


# Transform data in numerical format and give names
mat_data_automated <- data.matrix(data_automated[,2:ncol(data_automated)])
rownames(mat_data_automated) <- rnames


# Define colors of heatmap: red for high numbers
my_palette <- colorRampPalette(c("yellow","red"))(n = 299)


#Heatmap & Dendrogram
dev.off()
heatmap.2(dendrogram = "row", mat_data_automated, key = TRUE
        , keysize = 1.5, key.title = NA, col = my_palette, density.info="none", trace="none",
        revC = TRUE, cexCol = 0.9, cexRow = 0.8, margins = c(8, 8), offsetRow = 0.2,
        offsetCol = 0.1)


#Dendrogram vertikal
#par(mfrow=c(1,1))
```

```
hc.rows<- hclust(dist(mat_data_automated))
dend <- as.dendrogram(hc.rows)
par(cex=0.6, mar=c(10, 4, 10, 4))
plot(dend)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 20, lower_rect = -22, lty = 5, lwd = -1, col = rgb (0.1,0.2,0.4,0.1))
abline(h = 45, col = 2, lty = 2)

#Dendrogram horizontal
dend2 <- as.dendrogram(hc.rows)
par(cex=0.7, mar=c(4, 8, 0, 9))
hang.dendrogram
plot(dend2, horiz = TRUE)
plot_horiz.dendrogram(dend2, side = TRUE)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 7, lower_rect = -20, horiz = TRUE, lty = 5, lwd = -1, col = rgb
(0.1,0.1,0.1,0.2))
abline(v=48, col = 2, lty = 2)



wss <- function(k) {
  kmeans(mat_data_automated, k, nstart = 10 )$tot.withinss
}

# Compute and plot wss for k = 1 to k = 15
k.values <- 2:30

# extract wss for 2-15 clusters
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
```

```
    type="b", pch = 19, frame = FALSE,
    xlab="Number of clusters K",
    ylab="Total within-clusters sum of squares")

# function to compute average silhouette for k clusters
avg_sil <- function(k) {
  km.res <- kmeans(mat_data_automated, centers = k, nstart = 25)
  ss <- silhouette(km.res$cluster, dist(mat_data_automated))
  mean(ss[, 3])
}

# Compute and plot wss for k = 2 to k = 15
k.values <- 2:30

# extract avg silhouette for 2-15 clusters
avg_sil_values <- map_dbl(k.values, avg_sil)

plot(k.values, avg_sil_values,
    type = "b", pch = 19, frame = FALSE,
    xlab = "Number of clusters K",
    ylab = "Average Silhouettes")
```

Data analysis Peer to Peer Card Sorting

```
library(gplots)
library(RColorBrewer)
library(tidyverse)  # data manipulation
library(cluster)    # clustering algorithms
library(factoextra) # clustering visualization
library(dendextend) # for comparing two dendrograms
library(pheatmap)


# Read the data file (.csv format)
data_peer <- read.csv("Masterthesis/Analysis/Peer to Peer/Datasheet_analysis_peer.csv",
comment.char="#")
rnames <- data_peer[,1]


# Transform data in numerical format and give names
mat_data_peer <- data.matrix(data_peer[,2:ncol(data_peer)])
rownames(mat_data_peer) <- rnames


# Define colors of heatmap: red for high numbers
my_palette <- colorRampPalette(c("yellow","red"))(n = 299)


#Heatmap & Dendrogram
dev.off()
heatmap.2(dendrogram = "row", mat_data_peer, key = TRUE, keysize = 1.2, key.title = NA, col =
my_palette, density.info="none", trace="none",
        revC = TRUE, cexCol = 0.7, cexRow = 0.6, margins = c(7, 7), offsetRow = 0.2,
        offsetCol = 0.1)


#Dendrogram vertikal
#par(mfrow=c(1,1))
hc.rows<- hclust(dist(mat_data_peer))
```

```
dend <- as.dendrogram(hc.rows)
par(cex=0.6, mar=c(10, 4, 10, 4))
plot(dend)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 20, lower_rect = -22, lty = 5, lwd = -1, col = rgb (0.1,0.2,0.4,0.1))
abline(h = 45, col = 2, lty = 2)

#Dendrogram horizontal
dend2 <- as.dendrogram(hc.rows)
par(cex=0.7, mar=c(4, 8, 0, 9))
hang.dendrogram
plot(dend2, horiz = TRUE)
plot_horiz.dendrogram(dend2, side = TRUE)
#k = 4
#n = nrow(mat_data2)
#MidPoint = (hc.rows$height[n-k] + hc.rows$height[n-k+1]) / 2
#abline(h = MidPoint, lty=2)
dend %>% rect.dendrogram(k = 10, lower_rect = -30, horiz = TRUE, lty = 5, lwd = -1, col = rgb
(0.1,0.1,0.1,0.2))
abline(v=39.5, col = 2, lty = 2)

wss <- function(k) {
 kmeans(mat_data_peer, k, nstart = 10 )$tot.withinss
}

# Compute and plot wss for k = 1 to k = 15
k.values <- 2:27

# extract wss for 2-15 clusters
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
    type="b", pch = 19, frame = FALSE,
    xlab="Number of clusters K",
```

```
    ylab="Total within-clusters sum of squares")

# function to compute average silhouette for k clusters
avg_sil <- function(k) {
  km.res <- kmeans(mat_data_peer, centers = k, nstart = 25)
  ss <- silhouette(km.res$cluster, dist(mat_data_peer))
  mean(ss[, 3])
}

# Compute and plot wss for k = 2 to k = 15
k.values <- 3:27

# extract avg silhouette for 2-15 clusters
avg_sil_values <- map_dbl(k.values, avg_sil)

plot(k.values, avg_sil_values,
     type = "b", pch = 19, frame = FALSE,
     xlab = "Number of clusters K",
     ylab = "Average Silhouettes")

head(data_peer)
```