applied sciences

MDPI

Article

Predicting Reputation in the Sharing Economy with Twitter Social Data

Antonio Prada † **and Carlos A. Iglesias** *,†

Intelligent Systems Group, Universidad Politécnica de Madrid, 28040 Madrid, Spain; toniprada@gmail.com

* Correspondence: carlosangel.iglesias@upm.es

† These authors contributed equally to this work.

Received: 5 March 2020; Accepted: 14 April 2020; Published: 21 April 2020

Abstract: In recent years, the sharing economy has become popular, with outstanding examples such as Airbnb, Uber, or BlaBlaCar, to name a few. In the sharing economy, users provide goods and services in a peer-to-peer scheme and expose themselves to material and personal risks. Thus, an essential component of its success is its capability to build trust among strangers. This goal is achieved usually by creating reputation systems where users rate each other after each transaction. Nevertheless, these systems present challenges such as the lack of information about new users or the reliability of peer ratings. However, users leave their digital footprints on many social networks. These social footprints are used for inferring personal information (e.g., personality and consumer habits) and social behaviors (e.g., flu propagation). This article proposes to advance the state of the art on reputation systems by researching how digital footprints coming from social networks can be used to predict future behaviors on sharing economy platforms. In particular, we have focused on predicting the reputation of users in the second-hand market Wallapop based solely on their users' Twitter profiles. The main contributions of this research are twofold: (a) a reputation prediction model based on social data; and (b) an anonymized dataset of paired users in the sharing economy site Wallapop and Twitter, which has been collected using the user self-mentioning strategy.

Keywords: sharing economy; reputation; natural language processing; Wallapop; Twitter

1. Introduction

The increasing adoption of social media is transforming the way we live and do business [1]. As of October 2018, there are more than 2.2 billion monthly active users and 1.5 billion daily active users on Facebook [2]. Other social networks such as Youtube and Twitter have 1900 and 300 million monthly active users, respectively [2].

Social Media interactions, from commenting on a post to liking a photo, leave behind digital traces—voluntary or not—that can be used for mining patterns of individual, group, and societal behaviors [3]. These social data created by users when they interact with social media channels are usually known as *digital footprints* [4]. The full potential of digital footprints for generating insights about users comes from their combination with other sources, such as bank transactions [5], sensor information [6], or other social networks [3]. Several works have used many techniques for user profile matching across social networks, such as user profile similarity [7,8] as well as based on spatial, temporal, and content similarity [9].

In this work, we address how social media footprints can improve the social markets of the so-called sharing economy. The sharing economy can be defined as "*Collaborative consumption that takes place in organized systems or networks, in which participants conduct sharing activities in the form of renting, lending, trading, bartering, and swapping of goods, services, transportation solutions, space, or money*" [10]. The success of the sharing economy lies in connecting individuals who have underutilized assets

with other individuals interested in their use in the short term [11]. The shared assets range from vehicles and spare time to rooms, second-hand objects or spare time for performing everyday tasks. Some of examples of these services are BlaBlaCar (http://wwww.blablacar.com) for long-distance car- sharing, Turo (http://turo.com) for short-term car renting, Uber (http://www.uber.com) and Cabify (http://www.cabify.com) that offer an alternative to taxis, Airbnb (http://www.airbnb.com)) that enable sharing rooms and apartmetns, Wallapop (http://www.wallapop.com) for second-hand objects, and Fiverr (http://www.fiverr.com) for performing everyday tasks [11].

One of the potential barriers for the development of the sharing economy is the high level of trust needed to do potentially dangerous activities with strangers as driving for them or staying at their houses [12]. Several solutions attempt to overcome this problem. Some examples are online reputation systems where users can publicly rate each other after a transaction outcome or insurance-like protections that cover material and monetary risks [13].

Online reputation systems enable buyers and sellers to rate each other after each transaction (e.g., ride or apartment sharing), in the same way as eBay transactions. In this way, they help to build a reputation profile that makes transactions safer and less uncertain [12]. Nonetheless, some authors [14,15] have pointed out some drawbacks of these systems, such as their manipulation with opinion spam [16] or extremely high averaged ratings [15]. Another drawback is the need for historical data, and therefore the total uncertainty of new users' future behavior, which translates to mistrust from other users of the platform.

Such drawbacks could be mitigated by predicting the probability of a user being negatively rated by others. That information could enable preemptive measures to be taken. For this purpose, we propose to use social network data to build models that predict user reputation on sharing economy platforms, the same way that insurance companies use demographic data to predict customer risk. We choose to use social network data to train the models because of three reasons: (1) wide availability [2], (2) easiness to obtain data thanks to the availability of interfaces and (3) easiness of integration, since multiple sharing economy platforms, such as Airbnb or Blablacar, already allow users to link their social network profiles with their existing reputation profiles.

To further investigate this solution, we want to answer the following research questions:

- RQ1. Can social digital footprints be used for predicting bad user behavior on sharing economy sites?
- RQ2. What are the most relevant social features for predicting reputation in the Sharing Economy?

The remainder of the paper is organized as follows. Section 2 describes the state of the art. Then, Section 3 introduces the case study analyzed in this article as well as the proposed general approach to predicting user reputation based on Online Social Network (OSN) activity. It consists of three phases: the creation of a dataset of paired identities (Section 3.1), expanding this dataset with information from both sites (Section 3.2), and developing a reputation model (Section 4), which also includes the results and analysis from our experimental evaluation. Finally, the article concludes with Section 5 that discusses the conclusions and main findings of our work.

2. Background

2.1. Trust and Reputation Notions and Computational Models

Trust and reputation play an important role in human society. Thus, they have become subject of study of a wide array of disciplines, including sociology [17,18], social psychology [17,19], anthropology [20], political science [21], philosophy [22,23], law [24], economics [25], and computer science [26–36].

Several factors in today's society make the concept of trust more relevant [37]. Contemporary society has become more interdependent because of the globalization process. In addition, people have more available options in all domains of life, which makes both our decisions and those of

others less predictable. Moreover, organizations have become more opaque and complex. Another source of unpredictability is the increasing dependence on technology, which brings new threats and vulnerabilities.

Trust and trustworthiness is our strategy to deal with uncertain and uncontrollable future situations. As defined by Sztompka, "Trust is a bet about the future contingent actions of others" [37]. In this regard, Yamagishi [38] treats trust as a form of social intelligence that allows people to assess the degree of risks they face in social situations. Granting trust to a target is the result of the estimation of its trustworthiness, which can be determine using three main mechanisms [37]: reputation, performance, and appearance. Reputation is "a record of past deeds" [37]. Let's say we are evaluating how trustworthy an Uber driver is. The reputation would be based on the driver's rating. Performance refers to the present behavior. In the previous example, we would focus on how the driver is currently driving and maybe her conversation. Finally, appearance consists of superficial external signs relevant for trust, such as, for example, the driver's photo [39] or if the car is in good condition.

Online transactions challenge our traditional mechanisms for evaluating trustworthiness since we cannot see and try the products or the providers. In addition, reputation rating could be dishonest or biased towards positive feedback [40]. Therefore, trust and reputation systems have tried to provide adequate substitutes for the traditional signs we use in the physical world and take advantage of Information and Communications technology (ICT) technologies for collecting this information and providing useful trust and reputation metrics [32].

Trust and reputation have a rich research tradition in Computer Science. Several surveys have provided an excellent overview of its representation, computational models, and applications in Web services [28], the Semantic Web [29], web sciences [41], Multi-agent systems [27,31,42], online service provision [32,34,36], Peer to Peer (P2P) applications [26], ad hoc networks [30,33], and social networks [35].

The main mechanisms for managing trust are [26,43]: (i) *policy-based trust* that rely on security mechanisms such as credentials and access policies to grant access rights; and (ii) *reputation-based trust*, where past interactions are used for predicting future behavior; and (iii) *social network-based trust* that also use social relationships between peers to compute trust.

Policy-based trust relies on a trusted third party that serves as an authority for providing and verifying credentials. Policies can be expressed in policy specification languages [44]. Trust negotiation mechanisms enable building trust through a bilateral exchange of credentials preserving privacy [45].

Reputation-based trust systems should provide mechanisms for storing past experiences and calculating trust score. There are two main architectural types, centralized and distributed systems [32]. In centralized reputation systems, a central authority, the so-called reputation center, collects all the ratings and computes a public reputation score for every user. This is the most popular architecture [32] and is used in applications such as auction services such as eBay (http://ebay.com, product review sites such as BizRate (http://bizrate.com) and Amazon (http://www.amazon.com), or discussion fora such as Slashdot (http://slashdot.org). Distributed reputation systems lack a central component, and each participant can obtain ratings from other members. Many research solutions have been proposed [46,47], a recent research topic being its combination with blockchain technology as described in the survey [48].

As introduced previously, social network-based trust systems take advantage of the social network structure. Following the classification by Sherchan et al [35], three steps are followed for deriving social trust in social networks: (i) trust collection, (ii) trust evaluation, and (iii) trust dissemination. The first step, *trust collection*, is the process of collecting trust information from three main sources [35], attitudes [49], behaviors [50], and experiences [51]. Some of the signs that can be used for inferring trustworthiness from a social network are the opinions towards a user, topic or another entity (attitude), and the interaction patterns (behavior, for example, the level of participation), and previous experiences, usually reported through feedback mechanisms. The second step, *trust evaluation*, consists of deriving trust from the information collected. There are two main approaches for

social trust evaluation: network-based and interaction-based models, which can also be combined in hybrid models. *Networks-based social trust models* compute social trust using the social graph structure. Some works use networks based metrics (e.g., in-degree, out-degree), while others are based on the notion of "Web of Trust" and extend Friend-Of-A-Friend (FOAF) [52] and exploit the propagative nature of trust [53,54]. Regarding *interaction-based social trust evaluation models*, they rely the interaction patterns and the behavior of users [55]. Finally, the third step, *trust dissemination*, consists of exposing trust in the social network. The main dissemination models are *trust recommendation models* and *visualization models*. Trust recommendation models use a trust network where nodes are users and the edges are the trust place on them. This trust network is exploited by trust recommendation algorithms for generating personalized recommendation based on the aggregation of opinions of users in the trust network [56]. Accordingly, visualization models show the trust network that helps to understand trust relationships between the members of the social network. The reader can refer to the detailed survey by Liu et al. [57] for a comprehensive overview of the Machine Learning (ML) based methods for pairwise trust prediction.

2.2. Trust and Reputation in the Sharing Economy

Sharing economy transactions have more significant risks involved than traditional online transactions as the relationships are more intimate and sophisticated (e.g., allowing a stranger to sleep in your house or entering a stranger's car). For these platforms to succeed, it is necessary that their users feel safe and trust each other when using them because without that safety feeling it is unlikely that people would want to expose themselves to such a danger: trust is a mandatory requirement for sharing economy platforms development. As stated by Tadelis [58], *"the success of these marketplaces is attributed not only to the ease in which buyers can find sellers but also to the fact that they provide reputation and feedback systems that help facilitate trust."*

As mentioned previously, the notion of trust may vary in different domains, but, in the sharing economy, it can be defined as "the psychological state reflecting the willingness of an actor to place themselves in a vulnerable situation for the actions/intentions of another actor" [59].

To facilitate trust, sharing economy markets usually incorporate reputation and feedback systems [58]. After each transaction, users can provide feedback (e.g., five-star feedback score) and the market can derive the reputation of sellers and buyers from both public information (e.g., feedback) and not public information (e.g., messages). Reputation is only one of the available mechanisms to foster trust. Other mechanisms for fostering trust are long reciprocity in long-term relations, regulations, or professional qualifications [60]. In addition, personal reputation plays an important role in the shared economy [59,61]. The presence of storytelling narratives, connected accounts, and photos can be used for verifying the digital identity and increase trust as well as the sense of personal contact. Moreover, some studies [39] have reported that consumers infer sellers' trustworthiness from their photos, and this visual-based trust and attractiveness could affect more than their reputation, although this conclusion has not been confirmed in other sharing economy transactions [62].

An interesting approach to improve trust in the sharing economy is the possibility of trust transfer across sharing economy applications. The term "trust transfer" has been used in different research contexts, but in the sharing economy [63], the trust transfer situation can be described as a trustor (e.g., a user) that has some initial trust on a trustee (e.g., a service provider in AirBnB) and transfers his trust on a different context (e.g., the same service provider in a different platform such as BlaBlacar). There are two main approaches to trust transfer in the sharing economy [63]: a direct trust transfer possibility from platform to platform, and providing a reputation board from a third party that acts as trust authority. A number of startups have proposed solutions for reputation boards (e.g., erated.co, miicard.com, traity.com, truste.com), but they have not become popular yet. Reputation dashboards aim to aggregate users activity and reputation and to provide a trust score (or reputation passport) to become a trusted member in new social sites.

In this work, the social site Twitter is used as the source for reputation data, and we aim to transfer this reputation to the sharing economy platform *Wallapop* (http://wallapop.com) that is an online marketplace for second-hand articles launched in 2013. According to internal sources, the platform has more than 25 million downloads [64] and more than 5 million monthly active users in Spain, France, and US [65]. It follows the two-sided reputation schema, where both sellers and buyers receive a rating (from 1 to 5 stars) as well as an optional comment. Users' reputation can be measured by observing the ratings they receive after each transaction.

As shown in Figure 1, *Wallapop* user profiles consist of a nickname, a history of products sold with their associated feedback and list of products currently on sale. In addition, user profiles and products can be shared in many social networks (e.g., Facebook and Twitter) as well as by email. Wallapop can be accessed through iOS and Android applications and the official website.

Figure 1. Wallapop user profile.

2.3. Characterizing and Interlinking User Profiles in Social Networks

Technology enables new ways of communication, but it also enables ways to extract information about human interaction at a scale not available for classic ethnographic research. For example, mobile phone data have been used for inferring friendship relationships [66]. Moreover, behavioral patterns from mobile phones correlate with personal information such as gender [67] or job satisfaction [66].

In the context of OSNs, the volume of data generated related to personality traits, behavioral patterns, and social connections is even bigger: from self-reported networks of friendship to user-generated content. These data are usually publicly available and voluntarily generated by users. A large body of research has exploited this available information for inferring both social

aspects (e.g., influenza trends [68], sentiment in the stock market [69] or unemployment rates [70]) and individual characteristics (e.g., personality [71], location [72], ethnicity [73], or political affiliation [74]).

Since people have different accounts in many social networks, a body of research has addressed matching user profiles across different social networks to obtain a global profile of individuals. Several approaches have addressed this problem [75]: profile matching, where two user identities are compared to the similarity of their profile attributes [7,8,76]; network matching that compares network attributes [77] in two or multiple social networks [78], content matching that looks at the similarity of the content generated by users in different networks [75]; a combination of spatial, temporal and content matching [9] that compare spatial, temporal and content similarities; and self-reporting that searches for self-mentions to different networks [79].

3. Case Study and Methodology

We propose a model to *predict user reputation* in the sharing economy platform Wallapop, *using the social footprints* users leave when interacting on the Twitter platform. This way we are predicting the behavior of users in real-world peer-to-peer transactions (measured as the ratings received afterwards) by just using social data, frequently available in these platforms.

The methodology of this work is depicted in Figure 2. Three steps have been followed: (a) identification of paired identities between Wallapop and Twitter using a *self-mentions* strategy (Section 3.1); (b) enrichment of those identities by downloading user profiles from both platforms and generating numeric features (Section 3.2); and (c) development of a predictor of the probability of getting bad reviews in Wallapop based on the paired Twitter profile (Section 4).

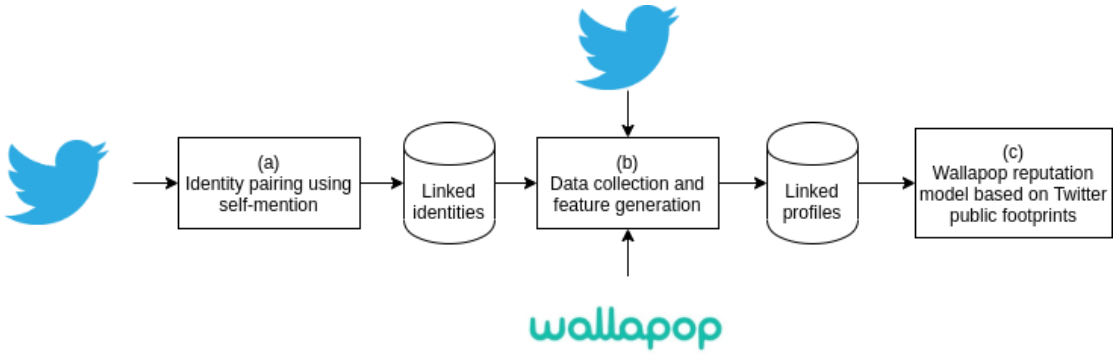


Figure 2. Proposed methodology to build a model for reputation prediction.

3.1. Identity Pairing

Past research has proved that publicly available information such as names, self-reported descriptions and social connections can be used to disambiguate users from different platforms [7]. While these methods are useful to match profiles between different platforms, they suffer from a significant error rate at the same time that are difficult to scale. Instead, we have designed an alternative process which tracks social media patterns related to sharing economy platforms to match identities from both platforms (Figure 2a). This way we are creating a dataset of matched identities that we can enrich later with both platforms data (Figure 2b).

This process leverages certain actions that can be performed in sharing economy platforms and have an impact on their connected social networks, such as posting referral invites or sharing some content. In the case of Wallapop, after users upload an item to the platform, they are offered the possibility to share such listing on multiple social networks to increase its potential audience. If users choose to share it on Twitter, a default message is offered, such message format has been constant through time and contains a link to the Wallapop listing. An example of this default message is: *"I am selling PlayStation 4 on #wallapop <item URL at Wallapop>"*.

Users can modify such messages, but most of the time, they do not change it or perform only minor changes. Based on this observation, we have searched tweets matching those patterns to collect Twitter users that sell a product in Wallapop so that we can match Twitter and Wallapop identities. In particular, we have searched for tweets that match the following two conditions:

- *Contain a link to a Wallapop product* that we can use to reach the Wallapop profile of the seller.
- *Tweets similar to the default tweet message*, including keywords as *selling* or hashtags as *#wallapop*.

Accessing Twitter data is a straightforward task thanks to the Application Programming Interface (API) offered by the platform. Twitter offers different *endpoints* to access different sets of data in the same way that the official applications do. For the profile matching, the *search tweets* endpoint was used, which returns a collection of relevant Tweets matching a specified pattern. After frequently tracking Twitter for an extensive period (from December 2015 to December 2016), we ended with a dataset formed by 34,981 paired Wallapop and Twitter identities. We expected a very low error rate with this matching process as it searches for the default tweet created when users share their profiles. In addition, we have created a random sample of 100 matches and manually verified them by comparing the names and pictures of both profiles, resulting in 98 confirmed matches and two uncertain matches.

3.2. Data Collection and Feature Generation

Once a dataset of paired user identities in Twitter and Wallapop has been collected, our goal is to extend that dataset with information from both platforms (Figure 2b): a metric for user reputation coming from Wallapop and a set of social features coming from Twitter.

3.2.1. Data Collection

Downloading Twitter public data is a straightforward process thanks to the existing API that allows us to download structured data for each one of the paired identities. The different characteristics available through different endpoints should be included to build a full Twitter profile:

- *users/show*: profile characteristics as name, picture, or description.
- *statuses/user_timeline*: list of tweets posted by the user.
- *followers/list*: list of followers.
- *friends/list*: list of friends.

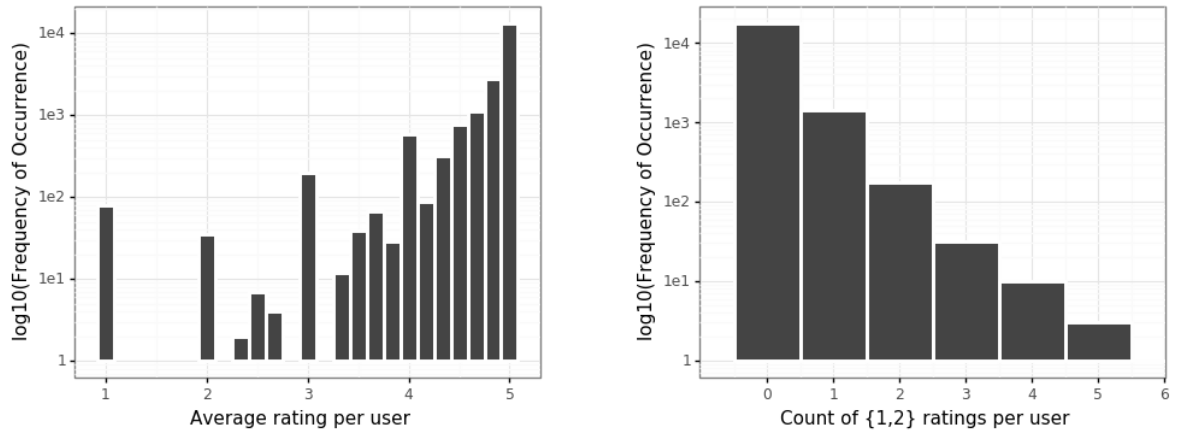
In contrast with Twitter, Wallapop does not offer a documented API to access its service. One alternative could be to use scraping techniques to extract data from the website, but this technique has the downside of not being resilient to changes in the web interface. Given that we wanted to obtain data for an extended period, we opted for a different option focused on accessing the data through the internal API that the official mobile applications use to communicate with the servers.

Wallapop mobile applications follow a typical pattern of server-client communication: a set of servers that perform most of the business logic feed applications with structured data to be rendered to users. Both parties communicate with each other through the use of an API that is usually stable for a long time, which makes it ideal for gathering data for an extensive period. To download data from the internal API, the first step was to identify which endpoints the mobile apps were connecting to and how they exchanged the information. The process started by configuring and deploying a transparent web proxy and an HTTP interceptor that logged all the data sent through it. Then, a smartphone was configured to use the proxy; the official Wallapop application was installed and used for some time. The output of the process was the full set of requests exchanged between the server and the application. By studying these logs, it was possible to get all the endpoints we needed:

- *Search*: a list of items available near a given location.

- *Item*: description of an item (including price and seller profile).
- *User*: user public profile, including data such as the number of reviews and verifications.
- *User reviews*: full set of reviews given or received by a user.

The output is a dataset of size 35,706 samples. As user reputation is based on ratings left by others, the only samples that are useful for the experiments are the ones who have received at least one review (n = 19,325). Users are rated from 1 to 5 stars. The average reviews rating distribution for each user is highly skewed to the right, due to the majority of positive reviews, with a big mass around integer values, as shown in Figure 3a. If we consider a review as negative to the ones with ratings below three stars out of five, the distribution of the count of negative reviews for each user has a mean value of 0.099 and a variance of 0.131, as shown in Figure 3b.



(**a**) Histogram of average rating per user (**b**) Histogram of negative ratings per user

Figure 3. Histograms of average rating per user, and negative rating per user in logarithmic scale.

3.2.2. Feature Generation

According to the general framework defined by Pennacchiotti and Popescu [80], Twitter user features can be classified into four categories: profile features ("who you are"), tweeting behavior ("how you tweet"), linguistic features ("what you tweet") and social network ("who you tweet"). The authorship framework [81] can complement linguistic features, which includes stylistic features.

Based on these frameworks, for the sake of this work, we have processed the following features, as shown in Table 1:

- **Profile**: we selected the account creation date (in seconds since epoch) based on the hypothesis that users with active accounts for a long time are likely to be more trustable [82].
- **Behavior**: based on previous research [83,84], we have selected activity metrics: *most frequent tweeting hours*, *tweets count*, *number of tweets marked as favorites by the user*, and the *tweet average length*. Other authors [85,86] use other network metrics such as centrality metrics since they analyze trust networks. We have not used them since our dataset is very sparse, based on the followed data collection strategy.
- **Linguistic**: the *average count of bad words by tweet* has been calculated using a publicly available list (https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words) of known bad words in Spanish and English (the two majority languages in our dataset). Then, we checked the presence of each tweet word in this list.
- **Social**: the Twitter network is unweighted and directed, with edges formed through the *following action*. Thus, we can analyze these edges separating them between the inner (followers) and outer edges (friends). The extracted features include features such as the count of friends and followers.

Table 1. Twitter features.

Type	Features
Profile	Account creation date
Behavior	Tweets count, Favourites count, Most frequent tweeting hours, Tweets average length
Linguistic	Bad words ratio
Social	Times added to list, Avg. retweeted per tweet, Avg. favourited per tweet, Followers count, Friends count, Followers' followers avg. count, Followers' friends avg. count, Followers' tweets avg. count, Friends' friends avg. count, Friends' tweets avg. count

These features were generated by processing the nested data structures returned from the Twitter API to generate numerical variables that can be used to train statistical and Machine Learning models. This process was an ad hoc one-time process that resulted in a set of numerical features for each one of the matched Twitter profiles. The statistics of these numerical features are shown in Table 2.

Table 2. Twitter features statistics.

	Mean	std	Min	Max
Account creation date	-	-	2006-12-26	2017-03-14
Tweets count	7086.63	16,277.20	0	556,756
Bad words ratio	0.02	0.02	0	0.46
Times added to list	9.76	42.45	0	2735
Favourites count	3066.26	11,826.47	0	399,054
Avg. retweeted per tweet	1397.67	5039.60	0	365,745.75
Avg. favourited per tweet	0.49	9.77	0	906.60
Most frequent tweeting hours	15.24	6.83	0	23
Followers count	556.66	5519.25	0	458,789
Friends count	476.43	2035.52	0	258,934
Tweets average length	88.96	26.75	0	179.53
Followers' followers avg. count	12,194.23	30,909.87	0	1,990,020
Followers' friends avg.count	6945.60	12,969.62	0	441345.57
Followers' tweets avg. count	7120.12	8148.27	0	327347.40
Friends' followers avg. count	941,730.90	1,921,375.52	0	105,297,483
Friends' friends avg.count	4323.63	8205.70	0	477,768.33
Friends' tweets avg. count	17,643.89	12,186.34	0	391,207

4. Prediction Model

Our primary research question (RQ1) is if social network footprints can be used for predicting bad user behavior on sharing economy sites. We base our assumptions in the existing research about the capabilities of Social Media for inferring personal traits. To explore such a hypothesis, we have designed the following experiment: predicting the proportion of bad ratings for each user by only using social features extracted from their Twitter profiles as predictors (Figure 2c). Given the polarization of reputation systems [15], we will define a bad rating as one with a value of less than half of the maximum, in this case, one or two stars out of five. In our dataset, there are 1922 bad ratings out of a total of 149,195.

We can make the following assumptions: (i) the majority of the Wallapop users are not fraudulent since reviews with low ratings are infrequent; and (ii) there is a certain probability of a transaction to go wrong. If this happens, the fault for such an unpleasant transaction can lie in one or both participants in the transaction. Based on this, we can then model the rate of reviews with low ratings in a certain user profile with a Poisson distribution where infrequent events, bad reviews, can occur with a certain probability. Such probability will be different for each user, and we are particularly interested in what are the user characteristics that relate to such probability. In fact, we want to predict such probability

by using only a set of user characteristics that are present on Twitter, as a way of predicting user behaviors at Wallapop by only using social data.

We could think about using a Poisson regression to model the count of bad ratings for each user. However, while Poisson distributions have a variance equal to the mean, in our data, the count of bad ratings suffers from overspreading: the 0.099 and the variance is 0.131. For that reason, we will model the data using the Negative Binomial distribution, which, as a gamma mixture of Poissons, can be used to model count data with overdispersion [87].

This experiment shares two characteristics with the challenges that actuaries face when doing risk modeling in the insurance industry: first, the events are infrequent counts, and second, there is high uncertainty about the predictions of a single entity [88]. Because of these similarities, we have designed our experiment with the tools of actuarial science, such as Generalized Linear Models for modeling of count data and gain curves for model performance evaluation [89].

Generalized Linear Models (GLMs) [90] are a means of modeling the relationship between a variable whose outcome we wish to predict and one or more explanatory variables. GLMs are parametric models, which are simpler and easier to understand than non-parametric models as the ones usually used in ML. At the same time, as a generalization of ordinary linear regression, they can be used to model the probability of occurrence or non-occurrence of counts, which is not possible with a simple Linear Regression. These characteristics make GLMs a good fit for the task of explaining the effects of the features on the predictions of the target variable.

Two of the limitations of parametric models are: i) the necessity of choosing the right distribution for the model and ii) the lack of robustness to correlated features. Nevertheless, determining accurate estimates of relativities in the presence of weakly correlated rating variables is a primary strength of GLMs versus other univariate analyses, ensuring that no information is double-counted [89]. No strong correlation (understood as a Pearson coefficient bigger than 0.4 [91]) is found between any of the features.

In a generalized linear model, each outcome of the dependent variables is assumed to be generated from a particular distribution in the exponential family. GLMs model the relationship between μ (the model prediction) and the predictors as follows:

$$g(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \tag{1}$$

Equation 1 states that some transformation g (link function) of μ (target variable) is equal to a linear combination of an intercept coefficient β_0 and a set of weighted predictors, where $x_1 ... x_n$ are the predictors, $\beta_1 ... \beta_n$ are the weight coefficients and n the number of predictors. By training the model, the coefficients ($\beta_0 ... \beta_n$) are estimated and used to predict the target variable. GLMs, and more specifically, multiplicative models, are widely used in the risk assessment insurance industry for their capacity to produce a multiplicative rating structure that easily explains the effect of each predictor in the final score [89]. In case of using the log link (natural logarithm ln):

$$ln(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \tag{2}$$

$$\mu = e^{\beta_0 + \beta_1 x_1 + ... + \beta_n x_n} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot ... \cdot e^{\beta_n x_n} \tag{3}$$

For our experiments, we have built a GLM from the Negative Binomial distribution family, using the *set of Twitter features* as predictors (x_1, ... , x_n). While the target variable y, defined as the ratio of bad ratings for each user $count_{bad}$ to the total count of ratings for each user $count_{total}$, is not a discrete variable but $y \in [0, 1]$, we can still perform a Negative Binomial regression by using the logarithm of the total count of ratings as an offset to the regression. By multiplying both sides of the equation by the count $count_{total}$, we move it to the right side of the equation. When both sides of the equation are then logged, the final model contains $ln(count_{total})$ as an offset term that is added to the regression:

$$y = count_{bad} / count_{total} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot ... \cdot e^{\beta_n x_n} \tag{4}$$

$$count_{bad} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot ... \cdot e^{\beta_n x_n} \cdot count_{total} \quad (5)$$

$$count_{bad} = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot ... \cdot e^{\beta_n x_n} \cdot e^{ln(count_{total})} \quad (6)$$

$$ln(count_{bad}) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + ln(count_{total}) \quad (7)$$

The results of training this model are shown in Table 3, where we can observe that multiple predictors have a significant p-value (0.05, marked in bold). This result confirms the predictive capabilities of Twitter social footprints for Wallapop reputation prediction and positively answers the research question RQ1.

Table 3. GLM regression summary. Predictors statistically significant at the 5% level are shown in bold.

	Coef	**Std Err**	**P > |z|**
Intercept	-7.8458	0.657	**<0.001**
Account creation date	2.957×10^{-9}	4.77×10^{-10}	**<0.001**
Tweets count	1.867×10^{-7}	2.2×10^{-6}	0.932
Bad words ratio	2.8823	1.246	**0.021**
Times added to list	-0.0013	0.001	0.350
Favourites count	8.7×10^{-7}	2.59×10^{-6}	0.737
Avg. retweeted per tweet	-4.387×10^{-6}	7.57×10^{-6}	0.562
Avg. favourited per tweet	0.0037	0.002	0.056
Most frequent tweeting hours	0.0052	0.005	0.259
Followers count	-4.28×10^{-5}	3.63e-05	0.238
Friends count	6.774×10^{-5}	3.84×10^{-5}	0.078
Tweets average length	-0.0078	0.001	**<0.001**
Followers' followers avg. count	-1.384×10^{-6}	2.21×10^{-6}	0.531
Followers' friends avg. count	8.631×10^{-6}	3.86×10^{-6}	**0.025**
Followers' tweets avg. count	-4.663×10^{-6}	4.21×10^{-6}	0.268
Friends' followers avg. count	4.773×10^{-8}	1.16×10^{-8}	**<0.001**
Friends' friends avg.count	1.236×10^{-6}	3.46×10^{-6}	0.721
Friends' tweets avg. count	3.512×10^{-6}	2.23×10^{-6}	0.115

Regarding the research question RQ2, we can observe in Table 3 that the features with a significant effect on the probability of a review being poorly rated are: (1) *account creation date* (as seconds since epoch) where the regression coefficient indicates that older accounts are related to a lower probability of a bad rating; (2) *bad words ratio*, having a higher ratio of bad words per tweet is related to being a worse rated user; (3) *average tweet length* indicating that worse users create shorter tweets; (4) the *average count of friends for the user followers* and (5) the *average count of followers for the user friends*, relating a lower risk of receiving a bad rating to having a smaller and closer Twitter network.

The *account creation date* was expected to be a relevant predictor, as past research relates this characteristic to trustworthiness [82]. We can also interpret that Twitter early adopters are more experienced in similar peer-to-peer platforms. The *bad words ratio* was also expected to have a significant effect in this particular direction: users that tweet more bad words receive more bad ratings on average. The other three significant predictors are more challenging to interpret: first, the *average tweet length* indicates that bad users create shorter tweets, maybe as a signal of certain personality traits, such as conscientiousness, which has been related to less aggressive behavior [92]. The other two significant predictors, the *average count of friends for the user followers*, and the *average count of followers for the user friends*, suggest that closer networks are related to fewer risky users.

We also trained the model against the good rating count but found that, while the coefficients had the opposite sign (as expected), there were no parameters with p-values below 0.05.

To answer RQ1, we have trained the same Generalized Linear Model with a 10 fold cross-validation strategy and assessed the model performance at differentiating between low and high-risk users using the Area Under the Curve (AUC) of the Gains Curve. Gain Curves are frequently

used in risk assessment to evaluate the performance of the models as a simple and intuitive method for evaluating count data. They are built similarly to Lorenz curves: first, the samples are sorted by the predicted target variable on the x-axis, with higher values on the left and lower values on the right. In this case, users with a higher predicted probability of receiving a bad rating are on the left of the graph. The y-axis represents the amount of the *accumulative percentage of the target variable*, in this case, the percentage of all bad ratings. For each user percentile, the cumulative percentage of bad reviews for those users is plotted, resulting in a curve that will be better the bigger the area under it.

The 45-degree line is the line of the *random model*, which does not have any predictive power at the task of identifying users with a higher probability of getting a bad rating. For example, when the random model is used to sort the users, 50% of users will have 50% of all the bad ratings. The top line is the *perfect model*, one that gives the higher probability of a bad rating to the ones that end up receiving the bad reviews. Therefore, the line will increase very steeply at first until 100% of the target variable is quickly reached [93]. A metric that summarizes the performance of the model is the AUC, also known as the lift index, understood as the area between the model gain curve and the random model curve.

After training and obtaining the output of the model with a 10 fold cross-validation strategy, we generated the Gains Curve displayed in Figure 4. The AUC is 9.83%, which proves that the model has a certain level of predictive power at the task of separating users by their probability of receiving a bad rating: a random model would have an AUC very close to 0%, and a model worse than a random model would have a negative AUC. While the insurance industry is very opaque about their risk assessment models' performance, there are some examples in the literature of models from which we can compare with, for example at the insurance [94] (19,9% AUC) or marketing [95] industries (from 15.6% to 28.9% AUC). As shown in Figure 4, a Random Forest Regressor (RFR) model is included as a baseline with an AUC of 1.52%. We decided to use the GLM model not only because of the better performance but also because, as as a parametric model, it allows us to understand the effect of the predictors in the model output and also answer RQ2.

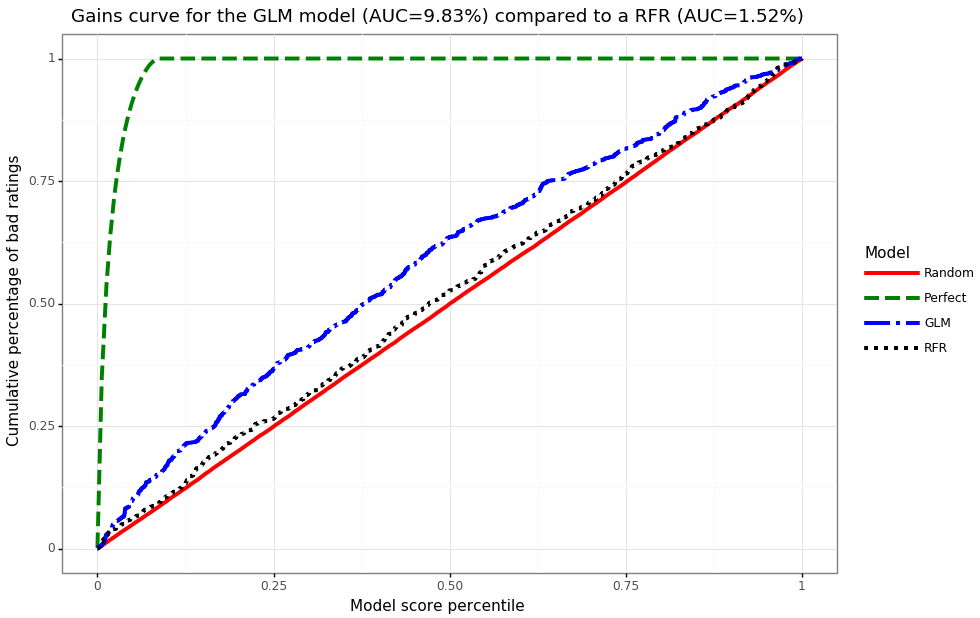


Figure 4. Gains curve for GLM predictions through a 10 fold cross-validation process, compared to a set of baselines: a random model, a perfect model with maximum gain and a RFR.

To summarize, by using the Gains Curve and the AUC, we can answer research question RQ1 and confirm the predictive power of Twitter social features for modeling the users' probability of receiving a bad rating when transacting at Wallapop. At the same time, we can use the same parametric model

to understand which Twitter social features are significant predictors of the outcome, by looking at the significance of the regression parameters, and answer RQ2.

5. Conclusions and Future Works

In this paper, we have explored the potential of social network footprints for predicting user reputation on sharing economy platforms. The initial findings show that our social network footprints can reveal our behavior, and this knowledge can be transferred across different social sites.

We have developed a process to match users from sharing economy platforms and OSNs. This process that leverages a self-mention strategy as the link between both platforms allowed us to create a dataset of profiles formed by Twitter and Wallapop profiles. Later on, we created a set of experiments that explained the relation between social traits and reputation and confirmed such effects. These results could be used to improve the cold-start problem for new users in sharing economy platforms by providing a prediction of how a new user can behave in the future.

One limitation of this study is that our approach has been tested only in one dataset, given that there are not available datasets of linked user-profiles across OSNs that enable trust computation as proposed in this article.

Several future works arise from this work. First, we aim at linking data from other OSNs to have a better understanding of the user, at the same time that targeting different platforms with Reputation Systems to compare different models and test their portability.

The results of this research have significant potential as they connect real-world behaviors with widely available social footprints. It can enable new ways to tackle information asymmetry problems, such as the ones that appear in financial systems in developing countries [96]—for example India having the largest Facebook user base in the world [97] but at the same time having a weak *credit scoring* record infrastructure [96], or in the insurance industry where part of the risk modeling is based on basic demographic traits [98].

Author Contributions: Conceptualization, Methodology, Writing, Review, Edition and Investigation A.P. and C.A.I.; Software, Validation and Data curation: A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors want to express their gratitude to the Traity team for fruitful technical discussions related to some contents of this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
AUC	Area Under the Curve
C2C	Consumer to Consumer
FOAF	Friend-Of-A-Friend
ICT	Information and Communications technology
GLM	Generalized Linear Model
OSN	Online Social Network
P2P	Peer to Peer
RFR	Random Forest Regressor
SVM	Support Vector Machine

References

1. Qualman, E. *Socialnomics: How Social Media Transforms the Way We Live and Do Business*; John Wiley & Sons: Hoboken, NJ, USA, 2010.

2. Statista. Social Media User Generated Content. 2018. Available online: https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/ (accessed on 16 April 2020).
3. Zhang, D.; Guo, B.; Li, B.; Yu, Z. Extracting social and community intelligence from digital footprints: An emerging research area. In *Proceedings of the International Conference on Ubiquitous Intelligence and Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 4–18.
4. Muhammad, S.S.; Dey, B.L.; Weerakkody, V. Analysis of factors that influence customers' willingness to leave big data digital footprints on social media: A systematic review of literature. *Inf. Syst. Front.* **2018**, *20*, 559–576. [CrossRef]
5. Sobolevsky, S.; Sitko, I.; Grauwin, S.; Combes, R.T.D.; Hawelka, B.; Arias, J.M.; Ratti, C. Mining urban performance: Scale-independent classification of cities based on individual economic transactions. *arXiv* **2014**, arXiv:1405.4301.
6. Psomakelis, E.; Aisopos, F.; Litke, A.; Tserpes, K.; Kardara, M.; Campo, P.M. Big IoT and social networking data for smart cities: Algorithmic improvements on Big Data Analysis in the context of RADICAL city applications. *arXiv* **2016**, arXiv:1607.00509.
7. Malhotra, A.; Totti, L.; Meira, W., Jr.; Kumaraguru, P.; Almeida, V. Studying user footprints in different online social networks. In Proceedings of the 2012 IEEE Computer Society International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), Istanbul, Turkey, 26–29 August 2012; pp. 1065–1070.
8. Vosecky, J.; Hong, D.; Shen, V.Y. User identification across multiple social networks. In Proceedings of the IEEE 2009 1st International Conference on Networked Digital Technologies, NDT 2009, Ostrava, Czech Republic, 28–31 July 2009; pp. 360–365.
9. Li, Y.; Zhang, Z.; Peng, Y.; Yin, H.; Xu, Q. Matching user accounts based on user generated content across social networks. *Future Gener. Comput. Syst.* **2018**, *83*, 104–115. [CrossRef]
10. Möhlmann, M. Collaborative consumption: Determinants of satisfaction and the likelihood of using a sharing economy option again. *J. Consum. Behav.* **2015**, *14*, 193–207. [CrossRef]
11. Cusumano, M.A. How Traditional Firms Must Compete in the Sharing Economy. *Commun. ACM* **2014**, *58*, 32–34. [CrossRef]
12. Belk, R. You are what you can access: Sharing and collaborative consumption online. *J. Bus. Res.* **2014**, *67*, 1595–1600. [CrossRef]
13. Walsh, G.; Bartikowski, B.; Beatty, S.E. Impact of customer-based corporate reputation on non-monetary and monetary outcomes: The roles of commitment and service context risk. *Br. J. Manag.* **2014**, *25*, 166–185. [CrossRef]
14. Jøsang, A. Robustness of trust and reputation systems: Does it matter? In *IFIP Advances in Information and Communication Technology*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 374, pp. 253–262.
15. Zervas, G.; Proserpio, D.; Byers, J. A First, Look at Online Reputation on Airbnb, Where Every Stay Is Above Average. 2015. Available online: http://dx.doi.org/10.2139/ssrn.2554500 (accessed on 3 April 2020).
16. Fei, G.; Li, H.; Liu, B. Opinion Spam Detection in Social Networks. In *Sentiment Analysis in Social Networks*; Pozzi, F., Fersini, E., Messina, E., Liu, B., Eds.; Morgan Kauffman: Burlington, MA, USA, 2017; Chapter 9, pp. 141–156.
17. Rousseau, D.M.; Sitkin, S.B.; Burt, R.S.; Camerer, C. Not so different after all: A cross-discipline view of trust. *Acad. Manag. Rev.* **1998**, *23*, 393–404. [CrossRef]
18. Zucker, L.G. Production of trust: Institutional sources of economic structure, 1840–1920. *Res. Organ. Behav.* **1986**, *8*, 53–111.
19. Rotter, J.B. A new scale for the measurement of interpersonal trust 1. *J. Personal.* **1967**, *35*, 651–665. [CrossRef] [PubMed]
20. Corsín Jiménez, A. Trust in anthropology. *Anthropol. Theory* **2011**, *11*, 177–196. [CrossRef]
21. Levi, M.; Stoker, L. Political trust and trustworthiness. *Annu. Rev. Political Sci.* **2000**, *3*, 475–507. [CrossRef]
22. Faulkner, P.; Simpson, T. *The Philosophy of Trust*; Oxford University Press: Oxford, UK, 2017.
23. Dormandy, K. *Trust in Epistemology*; Routledge: Abingdon, UK, 2019.
24. Mindus, P.; Gkouvas, T. Trust in Law. In *Routledge Handbook of Trust and Philosophy*; Simon, J., Ed.; Routledge: Abingdon, UK, 2020.
25. Williamson, O.E. Calculativeness, trust, and economic organization. *J. Law Econ.* **1993**, *36*, 453–486. [CrossRef]

26. Suryanarayana, G.; Taylor, R.N. *A Survey of Trust Management and Resource Discovery Technologies in Peer-to-Peer Applications*; Technical Report UCI-ISR-04-06; Institute for Software Research, University of California: Irvine, CA, USA, 2004.
27. Sabater, J.; Sierra, C. Review on computational trust and reputation models. *Artif. Intell. Rev.* **2005**, *24*, 33–60. [CrossRef]
28. Ruohomaa, S.; Kutvonen, L. Trust management survey. In *International Conference on Trust Management*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 77–92.
29. Artz, D.; Gil, Y. A survey of trust in computer science and the semantic web. *J. Web Semant.* **2007**, *5*, 58–71. [CrossRef]
30. Ahmad, M.; Salam, A.; Wahid, I. A survey on Trust and Reputation-Based Clustering Algorithms in Mobile Ad-hoc Networks. *J. Inf. Commun. Technol. Robot. Appl.* **2018**, *9*, 59–72.
31. Pinyol, I.; Sabater-Mir, J. Computational trust and reputation models for open multi-agent systems: A review. *Artif. Intell. Rev.* **2013**, *40*, 1–25. [CrossRef]
32. Jøsang, A.; Ismail, R.; Boyd, C. A survey of trust and reputation systems for online service provision. *Decis. Support Syst.* **2007**, *43*, 618–644. [CrossRef]
33. Momani, M.; Challa, S. Survey of trust models in different network domains. *arXiv* **2010**, arXiv:1010.0168.
34. Beatty, P.; Reay, I.; Dick, S.; Miller, J. Consumer trust in e-commerce web sites: A meta-study. *ACM Comput. Surv. CSUR* **2011**, *43*, 1–46. [CrossRef]
35. Sherchan, W.; Nepal, S.; Paris, C. A survey of trust in social networks. *ACM Comput. Surv. CSUR* **2013**, *45*, 1–33. [CrossRef]
36. Rahimi, H.; Bekkali, H.E. State of the art of Trust and Reputation Systems in E-Commerce Context. *arXiv* **2017**, arXiv:1710.10061.
37. Sztompka, P. *Trust: A Sociological Theory*; Cambridge University Press: Cambridge, UK, 1999.
38. Yamagishi, T. Trust as a form of social intelligence. In *Trust in Society*; Cook, K., Ed.; Russell Sage Foundation: New York, NY, USA, 2001; Chapter 4, pp. 121–147.
39. Ert, E.; Fleischer, A.; Magen, N. Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tour. Manag.* **2016**, *55*, 62–73. [CrossRef]
40. Tavakolifard, M.; Almeroth, K.C. A taxonomy to express open challenges in trust and reputation systems. *J. Commun.* **2012**, *7*, 538–551. [CrossRef]
41. Golbeck, J. Trust on the world wide web: A survey. *Found. Trends Web Sci.* **2008**, *1*, 131–197. [CrossRef]
42. Ramchurn, S.D.; Huynh, D.; Jennings, N.R. Trust in multi-agent systems. *Knowl. Eng. Rev.* **2004**, *19*, 1–25. [CrossRef]
43. Bonatti, P.; Duma, C.; Olmedilla, D.; Shahmehri, N. An integration of reputation-based and policy-based trust management. *Networks* **2007**, *2*, 10.
44. Kolar, M.; Fernandez-Gago, C.; Lopez, J. Policy Languages and Their Suitability for Trust Negotiation. In *IFIP Annual Conference on Data and Applications Security and Privacy*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 69–84.
45. Paci, F.; Bauer, D.; Bertino, E.; Blough, D.M.; Squicciarini, A.; Gupta, A. Minimal credential disclosure in trust negotiations. *Identity Inf. Soc.* **2009**, *2*, 221–239. [CrossRef]
46. Sharples, M.; Domingue, J. The blockchain and kudos: A distributed system for educational record, reputation and reward. In *Proceedings of the European Conference on Technology Enhanced Learning*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 490–496.
47. Veloso, B.; Leal, F.; Malheiro, B.; Moreira, F. Distributed Trust & Reputation Models using Blockchain Technologies for Tourism Crowdsourcing Platforms. *Procedia Comput. Sci.* **2019**, *160*, 457–460.
48. Bellini, E.; Iraqi, Y.; Damiani, E. Blockchain-Based Distributed Trust and Reputation Management Systems: A Survey. *IEEE Access* **2020**, *8*, 21127–21151. [CrossRef]
49. Alahmadi, D.H.; Zeng, X.J. ISTS: Implicit social trust and sentiment based approach to recommender systems. *Expert Syst. Appl.* **2015**, *42*, 8840–8849. [CrossRef]
50. Yan, Z.; Yan, R. Formalizing trust based on usage behaviors for mobile applications. In *Proceedings of the International Conference on Autonomic and Trusted Computing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 194–208.

51. Xiong, L.; Liu, L. A reputation-based trust model for peer-to-peer e-commerce communities. In Proceedings of the IEEE International Conference on E-Commerce (CEC 2003), Newport Beach, CA, USA, 24–27 June 2003; pp. 275–284.
52. Brickley, D.; Miller, L. *FOAF Vocabulary Specification 0.91*. 2007. Available online: http://xmlns.com/foaf/spec/20071002.html (accessed on 3 April 2020).
53. Golbeck, J.A. Computing and Applying Trust in Web-Based Social Networks. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2005.
54. Golbeck, J.; Rothstein, M. Linking Social Networks on the Web with FOAF: A Semantic Web Case Study. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*; AAAI Press: Menlo Park, CA, USA; Chicago, IL, USA; 2008; Volume 8, pp. 1138–1143.
55. Nepal, S.; Sherchan, W.; Paris, C. Strust: A trust model for social networks. In Proceedings of the 2011 IEEE 10th International Conference on Trust, security and Privacy in Computing and Communications, Changsha, China, 16–18 November 2011; pp. 841–846.
56. Hang, C.W.; Singh, M.P. Trust-based recommendation based on graph similarity. In Proceedings of the 13th International Workshop on Trust in Agent Societies (TRUST), Toronto, ON, Canada, 10 May 2010; Volume 82.
57. Liu, S.; Zhang, L.; Yan, Z. Predict pairwise trust based on machine learning in online social networks: A survey. *IEEE Access* **2018**, *6*, 51297–51318. [CrossRef]
58. Tadelis, S. Reputation and Feedback Systems in Online Platform Markets. *Annu. Rev. Econ.* **2016**, *8*, 321–340. [CrossRef]
59. Zloteanu, M.; Harvey, N.; Tuckett, D.; Livan, G. Digital Identity: The Effect of Trust and Reputation Information on User Judgement in the Sharing Economy. 2018. Available online: http://dx.doi.org/10.2139/ssrn.3136514 (accessed on 3 April 2020).
60. Slee, T. *What's Yours Is Mine: Against the Sharing Economy*; Or Books: New York, NY, USA, 2017.
61. Mauri, A.G.; Minazzi, R.; Nieto-García, M.; Viglia, G. Humanize your business. The role of personal reputation in the sharing economy. *Int. J. Hosp. Manag.* **2018**, *73*, 36–43. [CrossRef]
62. Ter Huurne, M.; Ronteltap, A.; Guo, C.; Corten, R.; Buskens, V. Reputation effects in socially driven sharing economy transactions. *Sustainability* **2018**, *10*, 2674. [CrossRef]
63. Zhang, J. Trust Transfer in the Sharing Economy-A Survey-Based Approach. *J. Manag. Sci.* **2018**, *3*, 1–32.
64. Lunden, I.; Lomas, N. Wallapop and LetGo, Two Craigslist Rivals, Merge to Take on the U.S. Market, Raise $100M More. 2016. Available online: https://techcrunch.com/2016/05/10/wallapop-and-letgo-two-craigslist-rivals-plan-merger-to-take-on-the-u-s-market/?guccounter=1 (accessed on 16 April 2020).
65. Mackin, S. Could Wallapop Be Barcelona's First Billion Dollar Startup? 2015. Available online: http://www.barcinno.com/could-wallapop-be-barcelonas-first-billion-dollar-startup/ (accessed on 16 April 2020).
66. Eagle, N.; Pentland, A.S.; Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15274–15278. [CrossRef]
67. Dong, Y.; Yang, Y.; Tang, J.; Chawla, N.V. Inferring User Demographics and Social Strategies in Mobile Social Networks. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), New York, NY, USA, 24–27 August 2014; pp. 15–24. Available online: http://keg.cs.tsinghua.edu.cn/jietang/publications/KDD14-Dong-WhoAmI.pdf (accessed on 17 April 2020).
68. Yun, G.W.; David, M.; Park, S.; Joa, C.Y.; Labbe, B.; Lim, J.; Lee, S.; Hyun, D. Social media and flu: Media Twitter accounts as agenda setters. *Int. J. Med. Inform.* **2016**, *91*, 67–73. [CrossRef]
69. Sánchez-Rada, J.F.; Torres, M.; Iglesias, C.A.; Maestre, R.; Peinado, E. A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain. In *CEUR Workshop Proceedings Joint Proceedings of the Second International Workshop on Semantic Web Enterprise Adoption and Best Practice and Second International Workshop on Finance and Economics on the Semantic Web Co-located with 11th European Semantic Web Conference, WaSABi-FEOSW@ESWC 2014, Anissaras, Greece, 26 May 2014*; CEUR: Aachen, Germany, 2014; Volume 1240, pp. 51–62.
70. Llorente, A.; Garcia-Herranz, M.; Cebrian, M.; Moro, E. Social Media Fingerprints of Unemployment. *PLoS ONE* **2015**, *10*, e0128692. [CrossRef]
71. Gosling, S.D.; Augustine, A.A.; Vazire, S.; Holtzman, N.; Gaddis, S. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychol. Behav. Soc. Netw.* **2011**, *14*, 483–488. [CrossRef]

72. Jurgens, D. That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships. In Proceedings of the 7th International AAAI Conference on ICWSM '13 Weblogs and Social Media, Cambridge, MA, USA, 8–11 July 2013; pp. 273–282.
73. Mislove, A.; Lehmann, S.; Ahn, Y.Y.; Onnela, J.P.; Rosenquist, J.N. Understanding the Demographics of Twitter Users. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011; pp. 554–557.
74. Lindamood, J.; Heatherly, R.; Kantarcioglu, M.; Thuraisingham, B. Inferring private information using social network data. In Proceedings of the 18th International Conference on World Wide Web WWW 09, Madrid, Spain, 20–24 April 2009; Volume 10, p. 1145.
75. Jain, P.; Kumaraguru, P.; Joshi, A. @i seek 'fb. me': Identifying users across multiple online social networks. In Proceedings of the Second International Workshop on Web of Linked Entities (WoLE) held in conjunction with the 22th International World Wide Web Conference, Rio de Janeiro, Brazil, 13 May 2013; pp. 1259–1268.
76. Raad, E.; Chbeir, R.; Dipanda, A. User Profile Matching in Social Networks. In Proceedings of the 2010 13th International Conference on Network-Based Information Systems, Takayama, Japan, 14–16 September 2010; NBIS '10; IEEE Computer Society: Washington, DC, USA, 2010; pp. 297–304.
77. Bennacer, N.; Jipmo, C.N.; Penta, A.; Quercini, G. Matching User Profiles Across Social Networks. In *International Conference on Advanced Information Systems Engineering*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 424–438.
78. Zhang, Y.; Tang, J.; Yang, Z.; Pei, J.; Yu, P.S. COSNET: Connecting Heterogeneous Social Networks with Local and Global Consistency. In Proceedings of the 21th ACM SIGKDD International Conference on KDD '15 Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; ACM: New York, NY, USA, 2015; pp. 1485–1494.
79. Correa, D.; Sureka, A.; Sethi, R. WhACKY!-What anyone could know about you from Twitter. In Proceedings of the 2012 Tenth Annual International Conference on IEEE Privacy, Security and Trust (PST), Paris, France, 16–18 July 2012; pp. 43–50.
80. Pennacchiotti, M.; Popescu, A.M. A Machine Learning Approach to Twitter User Classification. *ICWSM* **2011**, *11*, 281–288.
81. Zheng, R.; Li, J.; Chen, H.; Huang, Z. A Framework for Authorship Identification of Online Messages: Writing-style Features and Classification Techniques. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 378–393. [CrossRef]
82. Almendra, V. Finding the needle: A risk-based ranking of product listings at online auction sites for non-delivery fraud prediction. *Expert Syst. Appl.* **2013**, *40*, 4805–4811. [CrossRef]
83. Klassen, M. Twitter data preprocessing for spam detection. In Proceedings of the Future Computing 2013, The Fifth International Conference on Future Computational Technologies and Applications, Valencia, Spain, 27 May–1 June 2013; pp. 56–61.
84. Ahmed, C.; Elkorany, A. Enhancing link prediction in Twitter using semantic user attributes. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France, 25 August 2015; pp. 1155–1161.
85. Ceolin, D.; Potenza, S. Social network analysis for trust prediction. In *IFIP International Conference on Trust Management*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 49–56.
86. Meo, P.D.; Musial-Gabrys, K.; Rosaci, D.; Sarnè, G.M.; Aroyo, L. Using centrality measures to predict helpfulness-based reputation in trust networks. *ACM Trans. Internet Technol. TOIT* **2017**, *17*, 1–20. [CrossRef]
87. Ahn, S. An Interpretation of the Mixture of Poisson Distributions with a Gamma Distributed Parameter. In Proceedings of the Society of Korea Industrial and System Engineering (SKISE) Spring Conference, Dongan-gu, Korea, 16 May 2003; pp. 275–278.
88. Lemaire, J. *Automobile Insurance: Actuarial Models*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 4.
89. Goldburd, M.; Khare, A.; Tevet, D. *Generalized Linear Models for Insurance Rating*, 2nd ed.; Technical Report 5; Casualty Actuarial Society: Arlington, VA, USA, 2016.
90. Haberman, S.; Renshaw, A.E. Generalized linear models and actuarial science. *J. R. Stat. Soc. Ser. D Stat.* **1996**, *45*, 407–436. [CrossRef]
91. Evans, J.D. *Straightforward Statistics for the Behavioral Sciences*; Brooks/Cole: Andover, UK, 1996.

92. Barlett, C.P.; Anderson, C.A. Direct and indirect relations between the Big 5 personality traits and aggressive and violent behavior. *Personal. Individ. Differ.* **2012**, *52*, 870–875. [CrossRef]
93. Berry, J.; Hemming, G.; Matov, G.; Morris, O. Report of the model validation and monitoring in personal lines pricing working party. In *Proceedings of General Insurance Convention (GIRO 2009)*; Institute and Faculty of Actuaries: London, UK, 2009; pp. 1–54.
94. Fu, L.; Wang, H. Estimating insurance attrition using survival analysis. *Variance. Adv. Sci. Risk* **2014**, *8*, 55-82.
95. Ling, C.X.; Li, C. Data Mining for Direct Marketing: Problems and Solutions. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD'98), New York, NY, USA, 27–31 August 1998; Volume 98, pp. 73–79.
96. Mittal, S.; Gupta, P.; Jain, K. Neural network credit scoring model for micro enterprise financing in India. *Qual. Res. Financ. Mark.* **2011**, *3*, 224–242. [CrossRef]
97. Statista. Number of Facebook Users in India from 2015 to 2018 with a Forecast until 2023. 2018. Available online: https://www.statista.com/statistics/304827/number-of-facebook-users-in-india/ (accessed on 17 April 2020).
98. Kannadhasan, M. Retail investors' financial risk tolerance and their risk-taking behavior: The role of demographics as differentiating and classifying factors. *IIMB Manag. Rev.* **2015**, *27*, 175–184. [CrossRef]