Chapter III

A Survey of Trust Use and Modeling in Real Online Systems

Paolo Massa, ITC-IRST, Italy

Abstract

This chapter discusses the concept of trust and how trust is used and modeled in online systems currently available on the Web or on the Internet. It starts by describing the concept of information overload and introducing trust as a possible and powerful way to deal with it. It then provides a classification of the systems that currently use trust and, for each category, presents the most representative examples. In these systems, trust is considered as the judgment expressed by one user about another user, often directly and explicitly, sometimes indirectly through an evaluation of the artifacts produced by that user or his/her activity on the system. We hence use the term "trust" to indicate different types of social relationships between two users, such as friendship, appreciation, and interest. These trust relationships are used by the systems in order to infer some measure of importance about the different users and influence their visibility on the system. We conclude with an overview of the open and interesting challenges for online systems that use and model trust information.

Introduction

The Internet and the Web are pretty new creations in human history, but they have already produced a lot of changes in the lives of people who use them. One of the most visible effects of these two artifacts is that nowadays everyone with an Internet connection has the possibility to easily create content, put it online, and make it available to everyone else, possibly forever. If we are to compare this with the situation of some dozens of years ago, the difference is striking. In fact, until recently, only a tiny fraction of the world population had the possibility to "publish" content and distribute it to the public: for instance, few were the authors of books and few the musicians able to publish their music. Conversely, now everyone with an Internet connection can easily publish his/her thoughts on the Web: opening and keeping a blog, for instance, is both very easy and cheap today (actually it is offered for free by many Web sites, for example, blogger.com). Likewise, any band can record its songs in a garage, convert them to MP3 format, and create a Web site for the band to place their song files for the global audience. Moreover, in the future, we can only expect to have these capabilities extended, both on the axis of types of content that can easily be created and shared, and in terms of the range of people that are currently excluded for different reasons, such as location (many countries in the world still have to get the benefit of reliable and cheap Internet connections), age, education level, income.

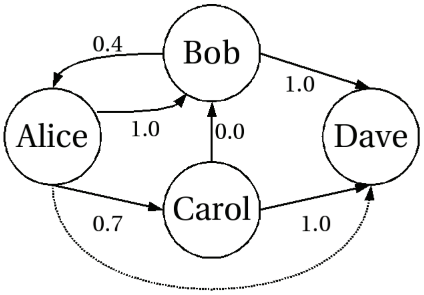
This phenomenon has been described as the "The Mass Amateurisation of Everything" (Coates, 2003), and we believe this term describes effectively the new situation. However, the easy publishing situation creates a problem, namely "information overload," a term coined in 1970 by Alvin Toffler in his book *Future Shock*. Information overload refers to the state of having too much information to make a decision or keep up to date about a topic. In fact, while it is good to have as many points of view as possible on any topic, it is impossible for a single human being to check them all. So we are faced with the challenge of filtering out the vast majority of the flow of daily created information and experience just the small portion that our limited daily attention and time can manage.

At the present time, it is unreasonable (and luckily almost impossible) to have a centralized quality control authority that decides what is good content, and thus worth our attention, and what instead must be ignored. But of course not all the content has the same degree of worthiness and interestingness for a specific person. What can be done is to infer the quality and value of the content from the "quality" of the content creator. However there is a problem: it is impossible for anyone to have a first-hand opinion about every other single creator of content. Until a few years ago, before the widespread availability of Internet, it was normal for most of the people to interact just with the people who were living physically close by. Geography was used to shape communities, and a person was able to decide about the neighbors trustworthiness in a lifelong ongoing process based on direct evidence and judgments and opinions shared by trusted people, for example by parents. Physical clues like the dress or the perceived sincerity of the eyes were also used to make decisions about trusting someone or not. Moreover, local authorities had some real power to enforce law in case of unacceptable and illegal behavior.

Instead, nowadays, as an example, it is a realistic possibility for a man in Italy to buy a used guitar from a woman in Taiwan and they will never see each other in the eyes, nor even talk. Also, the fact they live in different countries with different law systems makes it very difficult to enter into a legal litigation unless for really huge problems. Thanks to the Internet, we live in the so-called "global village," and in this new and totally different context we need new tools. To date, the most promising solution to this new situation is to a have a decentralized collaborative assessment of the quality of the other unknown people, that is, to share the burden to evaluate them. It is in fact the case that most of the community Web sites nowadays let a user express her opinions about every other user, asking how much she finds her interesting and worth her attention. We call these expressed opinions trust statements. For example, on Epinions (http://epinions.com), a site where users can review products, users can also specify which other users they trust, that is, "reviewers whose reviews and ratings they have consistently found to be valuable" (Epinions.com Web of Trust FAQ, n.d.) and which ones they do not. Similar patterns can be found in online news communities (for example, on slashdot.org, on which millions of users post news and comments daily), in peer-to-peer networks (where peers can enter corrupted items), in e-marketplace sites (such as eBay.com) and in general, in many open publishing communities (Guha, 2003). Usually judgments entered about other users (trust statements) are used to personalize a specific user's experience on the system, for example by giving more prominence to content created by trusted users. These approaches mimic real-life situations in which it is common habit to rely on opinions of people we trust and value: for instance, it is pretty common to ask like-minded friends their opinions about a new movie while considering if it is worth to go watching it or not. But the Web and the Internet exhibit a huge advantage for information dissemination on a global scale: all the trust statements can be made publicly and permanently visible and fetchable, possibly by everyone or only by some specific users. In this way, it is possible to aggregate all the trust statements in one single graph. A trust network is the graph obtained by aggregating all the trust statements issued by users. Figure 1 shows an example of a simple trust network. In it, all the trust statements are aggregated, such as the one issued by Alice expressing she trusts Bob as 1.0. In fact, trust statements can be weighted so that it is possible to express different levels of trust on another user. We assume the range of trust weights is [0,1]: small values represent low trust expressed by the issuing user on the target user while large values represent high trust. Precisely, the extremes, 0 and 1, represent respectively total distrust and total trust. Modeling distrust and making explicit its meaning is undoubtedly an open point and will be discussed in Section 3. Note that the trust network is, by definition, directed, and hence not necessarily symmetric. It is totally normal that a user expresses a trust statement on another user and that this user does not reciprocate: for example, Bob Dylan will hardly reciprocate a trust statement by a fan of his on a music site. In the simple trust network of Figure 1, for example, Bob totally trusts Dave but Dave did not express a trust statement on Bob. We say that Bob is unknown to Dave. Even when users know each other, it can be that their subjective trust statements exhibit different scores, as in the case of Alice and Bob in Figure 1.

On a trust network, it is possible to run a trust metric (Golbeck, Hendler, & Parsia, 2003; Levien, n.d.; Massa & Avesani, 2004; Ziegler & Lausen, 2004). A trust metric is an algorithm that uses the information of the trust network in order to predict the trustworthiness of unknown users. Coming back to Figure 1, since user Alice does not have a direct opinion about user David, a trust metric can be used to predict how much Alice could trust David (represented by the dotted edge in Figure 1). Let us suppose that in Figure 1 the trust statements are expressed by the source user based on perceived reliability of the target user as seller of used products, and let us suppose that Alice wants to buy a used camera. She finds

Figure 1: Trust network. Nodes are users and edges are trust statements. The dotted edge is one of the undefined and predictable trust statements.



out that David is selling a camera but she does not know David and is not sure about his trustworthiness and reliability. However, Alice knows Bob and Carol, who both know and trust David. In this case, a trust metric can happen to suggest to Alice to trust (or not) David and, as a consequence, to buy (or not) the camera from him. Of course, more complex reasoning involving more users and longer trust paths can happen in more realistic examples. By using trust metrics, even if the users known on a first-hand basis are a small fraction, it is possible to exploit the judgments of other users and figure out how much a certain user (and indirectly the content she creates) is interesting for the active user. Trust Metrics' common assumption is that trust can be propagated in some way, that is, if user *A* trusts user *B* and user *B* trusts user *C*, something can be said about the level of trust *A* could place in *C*.

The remaining of this chapter is organized as follows: Section 2 presents a classification of systems in which trust is modeled and used, along with a description of the most representative examples of real systems. For each of them, it describes what are the entities in the system (source and target of trust statement), which social and trust relationships they can express in the system, and how. It also analyzes how trust is used by the system for giving a better experience to the user. In Section 3, we discuss the challenges faced by online systems that model and use trust relationships.

Categories of Online Systems in which Trust is Modeled and Used

This section presents a classification of the online systems in which the concept of trust is modeled and used, and some examples of online systems that fit into the different categories. Even if the listed systems span a large spectrum of purposes and designs, it is possible to recognize some common features, and these drove our classification. In these online systems, visitors are invited to create a user profile so that their online persona is made visible, in general within a "user profile" Web page. Usually this page shows a picture of the user and some information entered by himself/herself. Often it also shows a summary of the user's activity on the system. This page is very important in these systems since it completely represents the human being behind the online identity, and often users form their opinions about other users only based on this page. These systems also allow one user to express some level of relationship with the other users, for example, concerning friendship, professional appreciation, commercial satisfaction, or level of acquaintance. We use the term ”trust” to represent many slightly different social relationships. Usually the list of expressed relationships with other users is public, but it might also be secret, as in the case of the distrust (or block) list in Epinions.

Trust is a very broad concept that has been investigated for centuries in fields as diverse as sociology, psychology, economics, philosophy, politics, and now computer science (see Mui, 2002 for a detailed summary of contributions from different research fields), and there are no commonly agreed definitions that fit all the purposes and all the investigation lines.

For the purpose of this chapter, we are going to provide an operational definition of trust.

Trust is defined as “the explicit opinion expressed by a user about another user regarding the perceived quality of a certain characteristic of this user.” The term “trust statement” will be used as well with the same intended meaning. The user expressing trust (i.e., issuing the trust statement) is the “source user,” while the evaluated user is the “target user.” We will see in the following how trust is represented and modeled in different ways in the online systems we will explore. For example in some systems, quality refers to the ability to provide reliable and interesting product reviews (as in Epinions). In others systems, it refers to the ability of being a good friend for the user (as in Friendster.com), while in others tit is the ability to find interesting new Web sites (as in Del.icio.us). This is called ”trust context,” and it is the characteristic of the target user evaluated by the user who emits the trust statement. Of course, in different trust contexts, a user can express different trust statements about the same user. For example, the subjective trust expressed by Alice on Bob about his ability of writing an interesting story about computers (the trust context of Slashdot.org) is in general not correlated with the trust expressed by Alice on Bob about his quality of being an honest seller online (the trust context of eBay.com). In the following, we describe different online systems that use and model trust.

We have identified few different categories in which the systems can be grouped based on the common features and properties they share. The categories we define are:

- E-marketplaces
- Opinions and activity sharing sites
- Business/job networking sites
- Social/entertainment sites
- News sites
- The Web, the Blogosphere, and the Semantic Web
- Peer-to-Peer (P2P) networks

E-marketplaces are online systems in which a user can sell items owned and can buy offered items. In such a context, typically the buyer does not know the seller, and vice versa.

So, in order to decide about a possible commercial exchange that involves the risk of not being paid or of not receiving the products already paid for, it is very important to be able to decide in a quick, reliable, and easy to understand way about the trustworthiness of the possible commercial partner. The success of eBay (http://ebay.com) is largely due to the fact it provides an easy way to do this.

Opinions and activity sharing sites on the other hand are Web sites where users can share with the other users their opinions on items and, in general, make their activities and preferences visible to and usable by other users. The best example of an opinions site is Epinions, in which users can write reviews about products. In activity sharing sites, the user activity is made visible to the other users who can in some way take advantage of it. Two examples of activity sharing sites are Del.icio.us, in which users can bookmark URLs they consider interesting, and Last.fm, in which users make visible which songs they listen to. Bookmarking a URL and listening to a song can be considered as the elicitation of a positive opinion about the considered item. However, a user might be more interested in following the reviews and activity of a certain other user, and trust statements can be used exactly for this purpose.

Business/job networking sites are Web sites where users post information about their job skills and ambitions so that other people can find them when they are looking for someone to hire for a specific job. Lately, many systems started to exploit the social (trust) network between users: users can explicitly state their connections, that is, professionals they have already worked with and found reliable and trustworthy. In this way, using the system, a user can enter in contact with the connections of his/her connections and discover potentially interesting new business partners. Linkedin.com and Ryze.com are two examples of such sites.

The idea behind the *social/entertainment sites* is similar to business/job networking sites. However, in this case, the context is more relaxed and informal, and sometimes involves dating and partner search. Here users are, in general, requested to list their friends so that, by browsing the social networks of them, it is possible to discover some friends of friends that might become friends. The first successful example was Friendster.com, soon followed by many other attempts.

News sites are centralized Web sites where users can submit news and stories and comment on them freely. The challenge is to keep the signal noise ratio high. Usually, the users can rate other users' activities (posted news and comments) and these ratings are used to give more visibility to posts and comments the other users appreciate and value. Slashdot.org and Kuro5hin.org are two examples of this category.

The Web, the Blogosphere, and the Semantic Web can be viewed as decentralized news sites. They are systems in which anyone is free to publish whatever content at whatever time in whatever form. Different from the previous examples, in these systems, there is not a single, central point where content is submitted and stored, but the content is published in a decentralized way; for example, it is stored on different Web servers. The challenge in this case is to design a system able to collect all this content and find a suitable algorithm to quickly decide about its importance and value. Google.com was the first company able to achieve this and in fact, a large part of its initial success over search engines of the time is due to the PageRank algorithm (Brin & Page, 1998). PageRank's assumption is to consider a link from page *A* to page *B* as a vote of *A* to *B*, or, in our jargon, a trust statement. The number of received trust statements influences the authority value of every Web page. In the following, we will review also how the concept of trust can be used in the Blogosphere (the collections of all the Web logs) and research efforts for introducing and exploiting trust in the Semantic Web.

The last category of systems that use and model trust is *Peer-to-Peer (P2P) networks*. P2P networks can be used to share files. P2P networks are, in fact, a controversial technology: large copyright holders claim they are used mainly to violate the copyright of works they own and are fighting to shut down this technology all together. We will not comment on this issue, but present the technological challenges faced by P2P networks. The open, autonomous, and uncontrollable nature of P2P networks in fact opens new challenges: for example, there are peers that enter poisoned content (e.g., corrupted files, songs with annoying noise in the middle) into the network. It has been suggested that a possible way to spot these malicious peers is to let every peer client express their opinions about other peers and share this information using the P2P network in order to isolate them. On a more positive take, a peer can mark as interesting (i.e., trust) another peer when it makes available for download many files that are considered interesting by its human user, or P2P networks can be used to share files only with a controlled and limited community of trusted friends.

Before going on with the discussion, it is worth mentioning that one of the first uses of the concept of trust in computational settings was in PGP (Pretty Good Privacy), a public key encryption program originally written by Phil Zimmermann in 1991. In fact, in order to communicate securely with someone using PGP, the sender needs to be sure that a certain cryptographic key really belongs to the human being he/she wants to communicate with. The sender can verify this by receiving it physically from that person's hands but sometimes this is hard, for example if they live in different continents. The idea of PGP for overcoming this problem was to build a "Web of Trust": the sender can ask someone whose key she already knows to send her a certificate confirming that the signed key belongs to that person. In this way, it is possible to validate keys based on the Web of Trust. The Web of Trust of course can be longer than two hops in the sense the sender can rely on the certificate received by someone who received it as well as from someone else, and so on. However, in this chapter, we are interested in the concept of trust from a more sociological point of view: trust here represents a social relationship between two entities, usually two users of an online system.

In the following, we present in more details different examples of online systems and, for each of them, what are the entities of the system and which social and trust relationships they can express in the system. We also analyze how trust is used by the system for providing a better experience to the user.

E-Marketplaces

E-marketplaces are Web sites in which users can buy and sell items. The more widely adopted models for arranging a deal are the fixed price (first in first out) and the auction. While the fixed price sell is a fairly straightforward model we are all acquainted with, auctions can take several forms and indulge in a few variants, depending on variables such as visibility of the offer, duration in time, stock availability, start bid, and so on. In an auction, buyers will compete, over the stated period, to put the best bid and win the deal. For the purposes of this chapter, we do not need to go any further in detailing the difference between the types of deals that can be conducted on an e-marketplace, but rather focus on the trust issues between the two roles that users play in this environment: the buyer and the seller. The main complication in conducting a deal in a virtual marketplace is that, in general, the buyer and seller do not know each other, and they only know the information that the Web site is showing about the other user. It is clear that there is a risk involved in a commercial transaction with a total stranger and, in fact, it is not common to give our money to a stranger in the street who promises to send us, days later, a certain product. Akerlof, Nobel Prize in Economy, formalized this idea in his "The market for lemons: Quality uncertainty and the market mechanism" (Akerlof, 1970). He analyzes markets with asymmetry of information, that is, markets in which the seller knows the real quality of the goods for sale but the buyer does not have this information. Using the example of the market of used cars, he argues that people buying used cars do not know whether they are "lemons" (bad cars) or "cherries" (good ones), so they will be willing to pay a price that lies in between the price for lemons and cherries, a willingness based on the probability that a given car is a lemon or a cherry. The seller has incentives to sell bad cars since he/she gets a good price for them, and not good cars since he/she gets a too low price for them. But soon the buyer realizes this situation and that the seller is actually selling only or mainly bad cars. So the price will lower and even less good cars and more bad ones will be put for sale. In the extreme, the sellers of good cars are driven out of the market and only lemons are sold. This effect is the opposite of what a free market should achieve and the only reason for this is asymmetry of information: the buyer has less information about the quality of the goods than the seller. Here, Trust Metrics (Golbeck et al, 2003; Levien, n.d.; Massa & Avesani, 2004; Ziegler & Lausen, 2004) and Reputation Systems (Resnick, Zeckhauser, Friedman, & Kuwabara, 2000) come to provide an escape to this vicious circle by the means of removing or at least reducing asymmetry of information. Giving users the chance to declare their degree of trust in other users makes it possible for future interactions to be influenced by past misbehaviors or good conduct. From this point of view, we can thus say that Trust Metrics and Reputation Systems promise to "unsqueeze the bitter lemon" (Resnick et al, 2000) and are a means to even the "risk of prior performance" (Jøsang, 2005).

The prototype of e-marketplace is eBay (http://www.ebay.com), at present the most known and successful example. Let us go through a typical use case of an e-marketplace. Alice found a user whose nickname is CoolJohn12 who accepts bids for his own used guitar. Let us suppose the bid price is fine with Alice. How can Alice be sure that, after she sends the money, CoolJohn12 is going to send her the guitar? How can Alice be sure that the picture on the site is really the picture of the guitar for sale and she is not going to receive another, possibly older, guitar? Unless she finds some evidence reassuring her about these questions, she is probably not going to take the risk and start the commercial exchange. This phenomenon reduces the quantity of commercial exchanges and hence the creation of prosperity (Fukuyama, 1995). But what if the e-marketplace Web site shows Alice that the guitar seller has already sold 187 guitars and banjos, and 100% of the buyers were satisfied by the received product? Or vice versa, what if the site tells Alice that many of the buyers were reporting that seller did not ship the guitar? A simple bit of information shown on the site can make the difference between "It is too risky to buy the guitar" and "I'm going to buy it." This is precisely what eBay does and this is the reason for its worldwide huge success. On eBay, users are allowed to rate other users after every transaction (provide "feedback" in eBay jargon or express trust statements in our jargon). The feedback can be positive, neutral, or negative (1, 0, -1). The main reason for the great success of eBay is due precisely to the idea of assigning a reputation score to every user and showing it. This simple bit of information is shown on the profile page of every eBay user and it is a summary of the judgments of the users who had in past a commercial transaction with that user. It represents what the whole community thinks about the specific user and corresponds to the definition of "reputation" found in Oram (2001). Thanks to this information, everyone can quickly form an opinion about every other user and decide if the risk of conducting a commercial exchange with this user is acceptable. Precisely, on eBay, the reputation score is the sum of positive ratings minus negative ratings. Moreover, the eBay user profile page also shows the total number of positive, negative, and neutral ratings for different time windows: past month, past 6 months, and past 12 months. The purpose is to show the evolution in time of the user's behavior, especially the most recent one.

EBay's feedback ecology is a large and realistic example of a technology-mediated market. The advantage of this is that a large amount of data about users' interactions and behaviors can be recorded in a digital format and can be studied. In fact, there have been many studies on eBay and in particular on how the feedback system influences the market (see for example Resnick & Zeckhauser, 2002). A very interesting observation is related to the distribution of feedback values: "Of feedback provided by buyers, 0.6% of comments were negative, 0.3% were neutral, and 99.1% were positive" (Resnick & Zeckhauser, 2002). This disproportion of positive feedback suggests two considerations: the first is actually a challenge and consists of verifying if these opinions are to be considered realistic or distorted by the interaction with the media and the interface. We will discuss this point later in Section 3. The second is about possible weaknesses of the eBay model. The main weakness of this approach is that it considers the feedback of every user with the same weight, and this could be exploited by the malicious user. Since on eBay there are so few negative feedbacks, a user with just a few negative feedbacks is seen as highly suspicious, and it is very likely nobody will risk engaging in a commercial transaction with him/her. Moreover, having an established and reputable identity helps the business activity. A controlled experiment on eBay (Resnick, Zeckhauser, Swanson, & Lockwood, 2003) found that a high reputation identity is able to get a selling price 7.6% higher than a newcomer identity with little reputation. For this reason, there are users who threaten to leave negative feedback (and therefore destroy the other user's reputation) unless they get a discount on their purchase. This activity is called "feedback extortion" on eBay's help pages (EBay help: Feedback extortion, n.d.) and in a November 2004 survey (Steiner, 2004), 38% of the total respondents stated that they had "received retaliatory feedback within the prior 6 months, had been victimized by feedback extortion, or both."

These users are "attacking" the system: as eBay's help page puts it "Feedback is the foundation of trust on eBay. Using eBay feedback to attempt to extort goods or services from another member undermines the integrity of the feedback system" (EBay help: Feedback extortion, n.d.). The system could defend itself by weighting, in different ways, the feedback of different users. For example, if Alice has been directly threatened by CoolJohn12 and thinks the feedback provided by him is not reliable, his feedback about other users should not be taken into account when computing the trust Alice could place in the other users. In fact, a possible way to overcome this problem is to use Local Trust Metric (Massa & Avesani, 2005; Ziegler & Lausen, 2004), which considers only (or mainly) trust statements given by users trusted by the active user and not all the trust statements with the same, undifferentiated weight. In this way, receiving negative feedback from CoolJohn12 does not influence reputations as seen by the active user if the active user does not trust explicitly CoolJohn12. For a short discussion of Global and Local Trust Metrics, see Section 3. However, eBay at the moment uses the Global Trust Metric we described before, which is very simple. This simplicity is surely an advantage because it is easy for users to understand it, and the big success of eBay is also due to the fact users easily understand how the system works and hence trust it (note that the meaning of “to trust” here means “to consider reliable and predictable an artifact” and not, as elsewhere in this chapter, ”to put some degree of trust in another user”). Nevertheless, in November 2004, a survey on eBay’s feedback system (Steiner, 2004) found that only 3% of the respondents found it excellent, 19% felt the system was very good, 39% thought it was adequate, and 39% thought eBay’s feedback system was fair or poor. These results are even more interesting when compared with numbers from a January 2003 identical survey. The portion of “excellent” went from 7% to 3%, the “very good” from 29% to 19%, the “adequate” from 35% to 39%, the “fair or poor” from 29% to 39%. Moreover, the portion of total respondents who stated that they had received retaliatory feedback within the prior 6 months passed from 27% of the 2003 survey to 38% of the 2004 survey. These shifts seem to suggest that the time might have come for more sophisticated (and, as a consequence, more complicated to understand) Trust Metrics. Following the success of eBay, many other online communities spawned their e-marketplaces; notable examples are Amazon Auctions and Yahoo! Auctions.

Opinions and Activity Sharing Sites

Opinions and activity sharing sites are Web sites where users can share with the other users their opinions on items and, in general, make their activities and preferences visible to and usable by the other users. The best example of an opinion site is Epinions (http://epinions.com). On it, users can write reviews about products (such as books, movies, electronic appliances, and restaurants) and assign them a numeric rating from 1 to 5. The idea behind opinions sites is that every user can check, on the site, what are the opinions of other users about a certain product. In this way, he/she can form an informed opinion about the product in order to decide about buying it or not. However, different reviews have different degrees of interest and reliability for the active user. Reviews are based on subjective tastes and hence, what is judged a good review by a user might be judged as unuseful by another user. So, one goal of Epinions is to differentiate the importance and weight assigned to every single review based on the active user currently served. Epinions reaches this objective by letting users express which other users they trust and which they do not. Epinions’ FAQ (Epinions.com Web of Trust FAQ, n.d.) suggests to place in a user’s Web of Trust “reviewers whose reviews and ratings that user has consistently found to be valuable” and in its Block List “authors whose reviews they find consistently offensive, inaccurate, or in general not valuable.” Inserting a user in the Web of Trust is equal to issuing a trust statement in him/her while inserting a user in the Block List equals to issuing a distrust statement in him/her. Note that Epinions is one of the few systems that model distrust explicitly. Trust information entered by the users is used to give more visibility and weight to reviews written by trusted users. Reviewers are paid royalties based on how many times their reviews are read. This gives a strong incentive to game the system and this is a serious challenge for Epinions use of trust. Challenges will be analyzed in Section 3. Epinions' use of trust has been analyzed in Guha (2003). Far from being the only example, other sites implementing metaphors very similar to Epinions are Dooyoo.com and Ciao.com. Note that their business models, heavily based on reviews generated by users, can be threatened by a controversial patent recently acquired by Amazon (Kuchinskas, 2005).

We decided to also place in this category those Web sites where users do not explicitly provide reviews and opinions, but their activity is made visible to the other users who can then take advantage of it. In fact, the activity performed by a user on a system can be seen as an expression of the opinions of that user on what are the most interesting actions to perform on the system, according to his/her personal tastes. Examples of these sites are Del.icio.us (http://del.icio.us), in which users can bookmark URLs they consider interesting, and Last.fm (http://last.fm), in which users make visible which songs they listen to. Bookmarking a URL and listening to a song can be considered as positive opinions about the considered item. On Del.icio.us, the act of trusting another user takes the form of subscribing to the feed of the URLs bookmarked by that user. In this way, it is possible for the active user to follow in a timely manner which URLs the trusted user considers interesting. Flickr (http://flickr.com) is defined by its founders as being part of "massive sharing of what we used to think of as private data" (Koman, 2005). In this scenario, of course, trust is something that really matters. On Flickr, users can upload their photos and comment on those uploaded by other users. Flickr users can then declare their relationship with other users on a role-based taxonomy as friend, family, or contact. Eventually, they can choose to make some photos only visible to or commentable by users of one of these categories. Similarly, Flickr makes use of this information by letting you see the pictures uploaded by your friends in a timely manner. Similar patterns can be seen in the realm of events sharing as well: Web sites such as Upcoming (http://upcoming.org), Rsscalendar (http://rsscalendar.com), and Evdb (http://evdb.org) allow one to submit to the system events the user considers interesting. It is also possible to add other users as friends (i.e., trusted users) in order to see all the public events they have entered. Then, if a user adds another user as a friend, the second also sees the private events entered by the first. In the domain of music, we already mentioned Last.fm: here the users can declare their friendship to other users by means of a free text sentence connecting user *A* with *B* (for example, Alice "goes to concerts with" Bob). Friends are then available to the user who can peek at their recently played tracks, or send and receive recommendations. On the other hand, Last.fm users played tracks are recorded in their profiles along with" (track the users especially likes) and "bans" (tracks the users does not want to listen to anymore). These are used by the system to evaluate a user's musical tastes and to identify his/her "neighbors" (members with interests in similar groups or musical genres). Last.fm members can then exploit their neighborhood by eavesdropping on casual or specific neighbor playlists. Many of these sites, including Del.icio.us, Last.fm, and Flickr, Upcoming, expose very useful Application Programming Interfaces (API) so that the precious data users entered into the system can be used by independent programs and systems (see Section 3 on problems related to walled gardens). Obviously, by combining two or more dimensions of activity of a specific user, it is possible to aggregate a profile spanning more than one facet about activities of one identity, as the Firefox extension IdentityBurro tries to do (Massa, 2005). The challenge of keeping a single identity under which all users' activities can be tied is briefly addressed in Section 3. Interestingly enough, Flickr, Del.icio.us, and Upcoming were recently bought by Yahoo!, whose interest into this so-called "social software" seems huge.

Business/Job Networking Sites

On business/job networking sites, users can register and post information about their job skills and ambitions so that other people can find them when they are looking for someone to hire for a specific job. Lately, many systems started to exploit the social (trust) networks between users: users can explicitly state their connections, that is, professionals they have already worked with and found reliable and trustworthy. Notable examples of these sites are LinkedIn (http://linkedin.com) and Ryze (http://ryze.com). On these sites, a user can discover new possible business partners or employees, for example, by entering in contact with the connections of his/her connections. These sites invite users to keep their connections list very realistic and to add as connections only people they really have worked with and deem reliable and recommendable. In order to achieve this purpose, business/job networking sites rely on the fact that user's connections are shown in the profile page and that other users will judge on the basis of the connections (Donath & Boyd, 2004). It is intuitive to say that a user will be better judged as IT consultant if reciprocated connections include Richard Stallman and Steve Jobs than if they contain many random users.

A similar but more playful site is Advogato (http://www.advogato.org). Advogato is a community site of freesoftware developers. The site was designed by Raph Levien, who planned to use it for studying and evaluating his trust metric (Levien, n.d.). On Advogato, users can keep their journal and indicate which free-software projects they are contributing to. A user can also express judgment on every other user based on their hacking skills by certifying him/her on a three-level basis: Master, Journeyer, and Apprentice. The Advogato trust metric is used to assign to every user a trust level. The trust metric is run once for every level on a trust network consisting only of the certificates not less than that level. Thus, Journeyer certification is computed using Master and Journeyer trust statements (certificates). The computation of the trust metric is based on a network flow operation, also called trust propagation. The trust flow starts from a "seed" of users representing the founders of the site, who are considered reliable ex-ante, and flows across the network with a decay factor at every hop. The computed certification level of a user (i.e., trust score) is the highest level of certification for which there was a flow who reached him/her; for example, if a user was reached both when propagating trust at level Journeyer and Apprentice, the certification is Journeyer. The trust metric is claimed to be attack-resistant, that is, malicious nodes are not reached by trust propagation (the topic is discussed in Section 3). Some other community sites use Advogato's code and hence show similar features. Something notable about Advogato is that it is one of the few sites that let users express a relationship with other users on a weighted base, in this case 3 levels. As a consequence, it is one of the few trust networks with weighted edges. From a research point of view, the availability of the trust network data (at http://www.advogato.org/person/graph.dot) is surely a relevant fact.

On a similar line, Affero (http://www.affero.org) is a peer-based reputation system, combined with a commerce system. It enables individuals to rate other individuals (i.e., express trust statements) and make payments on their behalf. Its goal is to exploit trust elicitation in order to democratically and distributedly decide which projects and foundations are more promising for the community and worth funding. Also, the system does not come bundled with any particular forum or community platform, so any independent community host can integrate the services and individuals can share reputation across various communities. One possible use case is the following: messages written by a user on an independent forum (or via e-mail) are signed with a message such as "Was I helpful? Rate me on Affero." Any individual reading the message and feeling he/she was helped can click on this Affero link and express gratitude by offering ratings, comments, and financial gifts to worthy causes chosen by the helping user on his/her behalf. Affero did not seem to have gotten momentum and is currently used by very few users.

Social/Entertainment Sites

Friendster (http://friendster.com) (Boyd, 2004), founded in 2002, was the first successful site to reach a critical mass of users among the social networking sites. On these sites, every user can create an online identity by filling out a profile form and uploading a their picture, and can then express a list of friends. The friends list, along with the user's picture and details, is shown on the user profile page. The idea is that other users can search through the friends lists of their friends and, in this way, discover and be introduced to new people that might be more interesting than a random stranger. We have already called this intuition "trust propagation." In December 2005, Friendster homepage claimed that there were more than 21 millions users using the system; however, this is not verifiable. A similar system was Club Nexus (Adamic, Buyukkoten, & Adar, 2003), an online community site introduced at Stanford University in the fall of 2001. Creators were able to use the system to study the real world community structure of the student body. They observed and measured social network phenomena such as the small world effect, clustering, and the strength of weak ties (Adamic et al., 2003). A very interesting and almost unique aspect of Club Nexus was the ability of users to rate other users (express trust statements) on a number of different axis: precisely, based on how "trusty," "nice," "cool," and "sexy" they find their connections (called buddies). Instead, current real online systems in general let users express just a single kind of trust statement and not many facets of it, and we believe this is a strong limitation.

Social sites (and also the previously analyzed business/job networking sites) usually enjoy a rapid growth of their user base due to the viral nature of the invitation process. We have seen that when users register, they can express their trust statements, that is, indicate other users they are connected to. If those users are not on the system they usually receive an e-mail from the system containing an invitation to join the network. This viral invitation strategy is able to rapidly bootstrap the user base. However, one risk is that the social network quickly becomes not representative of the real world because users tend to compete in the game of having more connections than others. Moreover, since everyone can create an identity, fake identities, sometimes called "fakester" (Boyd, 2004), start to emerge and lead the online system even further from a representation of real-world relationships. We will discuss this challenge in Section 3; however, let us briefly note how the creator of Club Nexus, Orkut Buyukkokten, later created Orkut (http://www.orkut.com), the social network of Google, and took a different approach. In fact, on Orkut site, it is not possible to create an identity without receiving an invitation from a user who already has an identity in the system. In this way, Orkut staff were able to control Orkut social network's growth, to keep it closer to reality and, as a by-product, to create a desire for users to be inside the system. In fact, the number of social networking sites counts at least in the hundreds, and there are less and less incentives for users to join and reenter their information in YASN (Yet Another Social Network).

Trust statements can be used also for making secure an otherwise risky activity such as hosting unknown people in one's personal house. CouchSurfing (http://couchsurfing.com) and HospitalityClub (http://hospitalityclub.org) are two Web sites in which registered users offer hospitality in their houses to other users, for free, with the goal to make their trips more enjoyable. In order to reduce the risk of unpleasant experiences, there is a trust system in place by which users can express their level of trust in other users (notably, on CouchSurfing the scale is based on 10 different levels ranging from "Only met online and exchanged emails" to "I would trust this person with my life"). The functioning is very similar to the other sites: users can create their profiles, filling in personal details and uploading photos of them. The system shows in the user profile the activity history (who that user hosted, by whom he/she was hosted, how the experiences were in the words and trust statements of the other users) so that, when receiving a request for hospitality from a user, anyone can check his/her history and what other users think about him/her and decide about hosting or denying the request. Additional security mechanisms are possible as well: for example, on CouchSurfing, a user can ask to have his/her physical location address certified by the system by a simple exchange of standard mail with the administrators of the site, and it is also possible to ask administrators to verify personal identity via a small credit card payment. In December 2005, CouchSurfing declared to have almost 44,000 users and HospitalityClub almost 98,000 users.

News Sites

News sites are Web sites where users can write and submit stories and news they want to share with the community. Two notable examples of News sites are Slashdot (http://slashdot.org) and Kuro5hin (http://kuro5hin.org). The most important requirement for such systems is the ability to keep the signal-to-noise ratio high. Slashdot was created in 1997 as a "news for nerds" site. It was a sort of forum in which users could freely post stories and comment on those stories. With popularity and an increased number of users, spam and low-quality stories started to appear and destroy the value of Slashdot. The first countermeasure was to introduce moderation: members of the staff had to approve every single comment before it was displayed. However, the number of users kept increasing and this moderation strategy did not scale. The next phase was the introduction of mass moderation: everyone could act as moderator of other users' posted stories. But in this way there was less control over unfair moderators and hence, metamoderation was introduced.

In December 2005, moderation on Slashdot consists of two levels: M1, moderation, serves for moderating comments, and M2, metamoderation, serves for moderating M1moderators. Note that moderation is used only for comments; in fact, it is the staff of Slashdot editors who decide which stories appear on the homepage. Then, once a story is visible, anyone can comment on it. Every comment has an integer comment score from -1 to +5, and Slashdot users can set a personal threshold where no comments with a lesser score are displayed. M1 moderators can increase or decrease the score of a comment depending on the fact they appreciate it or not. Periodically, the system chooses some users among longtime regular logged-in ones and gives them some moderation points, at the moment 5. A moderation point can be spent (during the next 3 days) for increasing the score of a comment by 1 point, choosing from a list of positive adjectives (insightful, interesting, informative, funny, underrated), for decreasing the score of a comment by 1 point, or choosing from a list of negative adjectives (offtopic, flamebait, troll, redundant, overrated). Moderation points added or subtracted to a comment are also added or subtracted to the reputation of the user who submitted it. User reputation on Slashdot is called Karma and assumes one of the following values: Terrible, Bad, Neutral, Positive, Good, and Excellent. Karma is important on Slashdot since a comment initial score depends on the Karma of its creator. Slashdot editors can moderate comments with no limits at all in order to cope with attacks or malfunctions in a timely fashion. So at M1 level, users rate other users (i.e., express trust statements on them) by rating their comments based on the perceived and subjectively judged ability to provide useful and interesting comments (the trust context of Slashdot M1 level). In fact, the ratings received by a comment directly influence its creator Karma.

Level M2 (called metamoderation) has the purpose to moderate M1 moderators and to help contain abuses by malicious or unreliable moderators. At M2 level, the trust context is related to how good a job a moderator did in moderating a specific comment. Only users whose account is one of the oldest 92.5% of accounts on the system can metamoderate, so that it is ineffective to create a new account just in order to metamoderate and possibly attack the system. Users can volunteer to metamoderate several times per day. They are then taken to a page that shows 10 randomly selected comments on posts along with the rating previously assigned by the M1 moderator. The metamoderator's task is to decide if the moderator's rating was fair, unfair, or neither. M2 is used to remove bad moderators from the M1 eligibility pool and reward good moderators with more moderation points. On Slashdot, there is also the possibility of expressing an explicit trust statement by indicating another user as friend (positive trust statement) or foe (negative trust statement). For every user in the system, it is possible to see friends, and foes (users at distance 1 in the trust network), friends of friends, and foes of friends (distance 2). For every user in the system it is also possible to see which users consider that user a friend (they are called fans) or a foe (called freaks). Every user can specify a comment score for every one of these categories so that, for example, he/she can increase the comment score of friends and be able to place their comments over the threshold, notwithstanding the comment score they received because of moderation.

Kuro5hin is a very similar system, but in December 2005, had a smaller community. However on Kuro5hin, users can directly rate stories and not only comments like on Slashdot. In this way, they influence which stories appear on the homepage, while on Slashdot this is done by editors. Kuro5hin users have the following options for rating a submitted story: "Post it to the Front Page! (+1)," "Post it to the Section Page Only (+1)," "I Don't Care (0)," "Dump It! (-1)." User reputation on Kuro5hin is called Mojo.

The goal of these systems is to keep the signal-to-noise ratio very high, even in the presence of thousands of daily comments, and in order to achieve this they rely on all the users rating other users' contributions and hence, indirectly expressing trust statements on them. The code running Slashdot and Kuro5hin is available as free software (GPL license). We will discuss, in Section 3, how the fact everyone can analyze and study the code is a positive fact for the overall security of the system and for the ability of the system to evolve continuously and to adapt to new situations and challenges.

The Web, The Blogosphere, and The Semantic Web

Different from the previous examples, in the systems presented in this section, there is not one single central point where content is submitted and stored, but the content is published in a decentralized way; for example, it is stored on different Web servers. The challenge here is to design a system able to collect this vast amount of content, and algorithms able to quickly decide about its relative importance and value. This section is about how the concept of trust can be used and modeled in the Web, the Blogosphere, and the Semantic Web, in order to make them more useful, so that, for example, users can search them and be informed about the quality of the different published information. This might mean either exploiting existing information such as the link structure of the Web or proposing new ways to represent trust information, for example, on the Semantic Web.

The World Wide Web (WWW, or in short simply the Web) is the collection of all the Web pages (or Web resources). It is an artifact created in a decentralized way by millions of different people who decided to publish a Web page written in HTML (hypertext markup language). Web pages are published on billions of different Web servers and are tied together into a giant network by HTML links. Hence, the Web is not controlled in a single point by anybody: the Web can be considered as a giant, decentralized online system. Search engines try to index all the information published on the Web and make it available for searching. Typically, search engines return a list of Web page references that match a user query containing one or more words. Early search engines were using information retrieval (Salton & McGill, 1986) techniques and were considering all the pages published on the Web as equally relevant and worth. However, since search engines are the most used way to locate a page, Webmasters wanted to have their pages on top of the list returned by a search engine for specific keywords. This gave the rise in the mid-1990s to a practice called Spamdexing: some Webmasters were inserting into their Web pages chosen keywords in small-point font face the same color as the page background so that they are invisible to humans but not to search engine Web crawlers. In this way, performances of early search engines quickly degraded to very low levels since the returned pages were no more the most relevant ones but just the better manipulated. Note that there is not a single entity with control over the content published on the Web and hence, it was not possible to block this behavior. In 1998, Sergey Brin and Larry Page, at that time students at Stanford University, introduced a new algorithm called PageRank (Brin & Page, 1998) that was very successful for combating spamdexing and producing better search results. They founded a new search engine company, Google (http://google.com), that, thanks to PageRank, was able to quickly become the most used search engine. The simple and genial intuition of PageRank is the following: not all the pages have the same level of authority, and their level of authority can be derived by analyzing the link structure. PageRank assumption is that a link from page *A* to page *B* is a "vote" of *A* on *B* and that authoritative pages either received many incoming links (votes) or even few incoming links but from authoritative pages. As an example, it seems reasonable to assume that a page that received no links is a nonauthoritative page. Based on an iterative algorithm, PageRank is able to assign to every page a score value that represents its predicted authority. This score value can be used by the search engine in order to give more prominence to more authoritative pages in the results list. PageRank is reported in this chapter because links are essentially what we call "trust statements," and PageRank is performing what we called "trust propagation" over the link network representing the Web.

Instead of asking trust statements in order to form the network as the previously introduced online systems did, PageRank's great intuition was to exploit a great amount of information that was already present, the links between Web pages, in a new and effective way.

Other even more explicit trust statements already available on the Web are represented by so-called blogrolls. Web logs (often contracted in blogs) are a very interesting recent phenomenon of the Web. A blog is a sort of online diary, a frequently updated Web page arranged chronologically, that is very easy to create and maintain and does not require knowing HTML or programming. It provides a very low barrier entry for personal Web publishing and so many millions of people in the world maintain their own blog and post on it daily thoughts (Coates, 2003). They pose new challenges and new opportunities for search engines and aggregators due to their continuously changing nature. In general, blogs contain a blogroll: a list of the blogs the blogger usually reads. It is self-evident that with the blogroll, the blogger is stating: "I trust these other blogs, so, if you like what I write, you will like what they write." What is relevant is that today there are millions of daily updated blogs and that blogs represent, in some sense, a human being identity. So the network of blogrolls really represents an updated, evolving social network of human beings who express their trust relationships via their blogrolls. There is an attempt to add some semantics to blogrolls: XFN (XHTML Friends Network, n.d.) is a microformat that allows representation of human relationships using hyperlinks. XFN enables Web authors to indicate their relationships to the people in their blogrolls simply by adding an rel attribute to their <a> tags. For example, <a href="http://alice.example.org" rel="met friend"> means that the author of the Web page in which the link is contained has met the person "represented" by http://alice.example.org and considers her a friend. There are also some Semantic Web proposals for expressing, in a semantic format, social relationships. FOAF (Friend-Of-A-Friend) (Golbeck et al., 2003) is an RDF format that allows anyone to express social relationships and place this file on the Web. There is also a trust extension that allows enrichment of an FOAF file by expressing a trust statement in other people on a 10 level basis (Golbeck et al., 2003). While preliminary research in this field hints its usefulness, the adoption of these semantic formats is slow and not straightforward.

Peer-to-Peer (P2P) Networks

Peer-to-peer (P2P) is "a class of applications that takes advantage of resources (storage, CPU cycles, content, human presence) available at the edges of the Internet" (Shirky, 2000), and has been defined as a disruptive technology (Oram, 2001). Three primary classes of P2P applications have emerged: distributed computing, content sharing, and collaboration. P2P networks are based on decentralized architectures and are composed by a large number of autonomous peers (nobody has control over the overall network) that join and leave the network continuously. In this sense, their open, autonomous, and evolving nature pushes the challenges of the Web to new and harder levels. Just as with Web pages, the reliability of other peers is not uniform. For example, in content-sharing networks, there are peers who insert poisoned content, such as songs with annoying noise in the middle, or files not corresponding to the textual description. And there are peers who share copyrighted content violating the law of some countries. The human controlling the peer, based on his/her subjective judgments, might not want to download files from peers of one of the two categories, that is, she distrusts them. So one possibility is that peers are allowed to express trust statements in other peers in order to communicate their level of desire to interact in future with those peers. By sharing these trust statements (expressing both trust for appreciated peers and distrust for disliked peers), it is possible to use a Trust Metric to predict a trust score in unknown peers, before starting to download content from them or upload it to them. Trust Metrics can also be used for individuating a close community of friends and share private files just with them.

There are some attempts to build trust-aware systems on top of current P2P networks: on the eDonkey network, it is possible for every peer to mark other peers as friends who are given more priority in downloading files, and a protocol for sharing trust statements has been proposed for the Gnutella P2P network (Cornelli, Damiani, DeCapitani di Vimercati, Paraboschi, & Samarati, 2002). A trust model called Poblano (Chen & Yeager, 2001) was introduced in JXTA, a Java-based P2P framework, and mechanisms based on trust and reputation are present in Free Haven (Oram, 2001) and in BitTorrent (Cohen, 2003).

There is also evidence that P2P networks suffer free riding (Adar & Huberman, 2000), that is, some peers only download files without letting their files available for downloads and in this way, they reduce the value of the entire network. The same trust-aware techniques can be used to share information about which peers allow or not to download files and give priority to nonfree-riding peers.

Research on reputation and trust in P2P networks is an ongoing effort and many proposals have been made lately. However, due to the autonomous and inherently uncontrollable nature of P2P networks, most of the research papers present results validated with simulations (Kamvar, Schlosser, & Garcia-Molina, 2003; Lee, Sherwood, & Bhattacharjee, 2003), while it is difficult to evaluate the real impact of these strategies on real and running systems.

Open Challenges

In this section, we will introduce what are the most interesting challenges related to the use and modeling of trust in online systems. They are divided into three subsections analyzing respectively: (1) differences in how trust relationships are modeled in real and virtual worlds, (2) how trust can be exploited in online systems, and (3) identity, privacy, and attacks in online systems.

Differences in how Trust Relationships are Modeled in Real and Virtual Worlds

It should come as no surprise that social relationships (particularly trust relationships) are different in the “real” world and in the “virtual” world. However, this fact is particularly relevant if the online systems designers want the trust statements expressed in their environment to resemble the real ones. The differences are especially evident with respect to the following issues: how trust relationships can be represented, how they begin and develop over time, and how their representation is perceived by the humans involved. What follows is a list of the most relevant issues.

Explicitness and visibility of trust statements. In a virtual environment, for example, on a community Web site, often trust relationships are explicit. And they are often publicly visible. This means that a user is in general able to check if there is a relationship between two users and, in this case, to see it and refer to it (Donath & Boyd, 2004).

Trust statements realism and social spam. There is also a risk of creating what has been named "social spam" (Shirky, 2004). This happens when a new user in a social network site is allowed to invite, by e-mail, a large number of people into the system, for example, by uploading his/her entire address book. This has happened with at least two social network sites, ZeroDegrees (http://zerodegrees.com) and Multiply (http://multiply.com), and has generated a large vent of protests (Shirky, 2004). New users used this feature and, with a single click, sent an invitation e-mail to all the e-mail addresses in their uploaded contact list, often without realizing this would have resulted in thousands of sent e-mails. Exploiting the viral nature of social networks can be used for passing from zero members to millions, but designers should ask themselves if it is worthwhile to annoy so many users in the hope of retaining a small portion of them, or if this feature is just creating annoying "social spam." Instead, since the beginning, Orkut tried to exploit the same viral nature of sending invitations into the system but in an opposite and more creative way: it was possible to register on the Web site only by being invited by someone already inside the system. By manipulating the number of invitations members could use to invite other people who were still outside the system, Orkut staff was able to create a lot of expectation and a lot of requests for joining the network. In this way, they were also able to control the growth of the network, in order to check if their servers and code were able to handle the load. And another good side effect of this was that, at least at the beginning, the social network was resembling real-world relationships, since every user had a limited number of invitations he/she could use and could not easily engage in the activity of adding as many friends as possible, even if they are not real-world friends. The optimal situation would be the one in which the user remains the owner of hiss/her social network and trust statements and can export them to every social site instead of having to re-express them every time. We will comment on interoperability at the end of subsection 3.3.

Disproportion in positive trust. The explicitness and visibility of social relationships represents a huge challenge especially for e-marketplaces. Some reports (see for example Resnick & Zeckhauser, 2002) have found there is high correlation between buyer and seller ratings on eBay, meaning that there is a degree of reciprocation of positive ratings and retaliation of negative ratings. This fact is probably a by-product of the site design and does not closely represent real-world opinions. We have also already commented on how feedbacks on eBay are disproportionately and unrealistically positive (almost 98% of feedback is positive) (Resnick & Zeckhauser, 2002). One explanation of this fact is that, for fear of retaliation, negative feedback is simply not provided in case of a negative experience. Gross and Acquisti (2003) suggest that "no feedback is a proxy for bad feedback," and one solution the authors propose is that the seller gives feedback first and then the buyer is free to give the "real" feedback without fear of retaliation. Anyway, it is easy to argue how often online trust statements do not represent real-world relationships; for example, on many social sites there is a run to have as many friends as possible, and on many e-marketplaces there is an incentive for not providing negative ratings. Psychological and sociological considerations must be taken into account when making available a system that allows one to express relationships online.

Modeling negative trust. Modeling negative relationships (i.e., distrust) is another serious challenge. Few systems attempt to do it: eBay allows users to give negative feedback, but we have seen how this is problematic and seldom used; Epinions allows the active user to place another user in the Block List in order to communicate to the system his/her reviews are considered unreliable and should be hidden and not considered. However, Epinions clearly states that "the distrust list is kept private and only you can see your Block List. Unlike the Web of Trust, there is no way you can display your Block List to others. This feature was designed to prevent hard feelings or retaliation when you add members to your Block List" (Epinions.com Web of Trust FAQ, n.d.). In a similar way, while in the real world it can happen, for example, that someone expresses, in private, doubts about the skills of the boss, it is very unlikely that he/she will state a negative trust statement on the boss on a professional site, if this is publicly visible. So surely, the visibility of trust statements changes how users express them and this is something that must be taken into account when designing an online system that models trust. Moreover, on a social site (like Friendster), there are few reasons for entering a negative concept like distrust, since people engaging in a community of friends are there for sharing experiences with people they like and not to punish people they do not appreciate. On the other hand, on P2P systems, trust statements are used both in a positive way in order to keep track of peers whose shared content is reliable and appreciated, but also in a regulative way in order to keep track of peers whose shared content is considered inappropriate and undesirable (for example, depending on the subjective desires, it is poisoned or it is illegally shared). So, in those systems, explicit modeling of negative relationships is necessary since one of the goals is to spot out what are the peers the active peer considers malicious and to warn other peers about them. In short, modeling negative trust statements must be dealt with even more care than positive trust statements, because of the perception humans can have of it and for its great potential of destroying the feeling of community users often look for in online systems. How to exploit negative trust statements, in case they are modeled, will be analyzed in subsection 3.2.

Rigidity of language for expressing trust statements. We have also seen in the examples of the previous subsection that online relationships are represented in a rigid way. For example, it is common to represent friendship as a binary relationship between two users: either it is there or not. Even the richest representations are just a set of numeric values or of predetermined text labels. Anyway, they are rigid. For instance, the evolution in time of a relationship in real life follows a smooth growth or decay and it is often unconscious, or at least not continuously explicitly represented and polled. On the other hand, in virtual environments, the representation is always explicit, and it grows or decays in discrete steps: a possible event on an online community is, for example: "today, I downgrade my friendship to you from level 7 to level 6" and this discreteness hardly models any real-world relationship evolutions. This is surely a challenge for a system that wants to model real-world trust relationships in a reasonable, human-acceptable way. On the other hand, it is possible to keep relationships implicit: for example, the strength of a relationship can be derived on the fly from the number of messages two users have exchanged and hence, this value would closely model changes in the relationship patterns. While this option partially solves the aforementioned issue, in this way the system would become a black box for the user who is not in the condition to know what the system thinks is his/her relationship with the other users and possibly to change it.

Keywords for trust statements conveying undesired meanings. Keywords used in the GUI (graphical user interface) of the system are very important as well. If the system uses the term "friend" for defining a relationship on the system, where a friend is someone who provides timely links to interesting pages (e.g., Del.icio.us), that could be misleading, since the term "friend" in real life means something else. For example, a non-Web savvy but real friend could be unhappy with not seeing himself/herself on the friend list. A reasonable suggestion is to avoid the term "friend" in online systems unless the social relationship really represents friendship, and to use less emotional terms such as "connection" (as LinkedIn and Ryze do) or to use a unique, made-up word with no predefined meaning in the real world (Allen, 2004). We believe this is a key issue for the representativeness of issued trust statements, but we are not aware of research analyzing, with controlled experiments, the impact of different chosen terms in the trust elicitation patterns.

Single-trust context. Moreover, real-world relationships are not embeddable into a single-trust context. A user might appreciate another user for his/her discussions on computer-related topics, but less for his/her always being late or for his/her political ideas. At the moment, it seems very unlikely that an online system that asks users to state trust statements for more than one trust context will be successful; the previously described Club Nexus (Adamic et al., 2003) was an exception in this sense. Even in this case, it is not easy to find the "right" categories for defining a relationship and, as already stated, rigid predefined categories are surely not optimal for representing ongoing real-world situations.

Incentives mechanism for trust elicitation. Another challenge is to find the correct incentives for providing trust statements. The basic assumption of economy, rationality, would suggest that users have the incentive to free ride. In this context, free riding means not providing trust statements and just relying on the trust statements provided by the other users. However, contrary to the basic assumption of economy, many eBay users do provide feedback after a transaction: Resnick and Zeckhauser (2002) found that on average 60.7% of the buyers and 51.7% of the sellers on eBay provided feedback about each other. However, in general, incentives must be envisioned by online-systems designers. On eBay, providing (positive) feedback after a transaction might be seen as an exchange of courtesies (Resnick & Zeckhauser, 2002), or it might be that users perceive the global value of the feedback system and that in order to keep the community healthy, they think they should contribute to it when they can by providing feedback.

On social and activity sharing sites, expressing a trust statement provides a direct benefit to the user since he/she is then able to spend more time on content created by trusted users and less time on not interesting content. In fact, the system in general gives more visibility to trusted users and the content they created. For example, Flickr shows to logged-in users the pictures uploaded by their friends and contacts in a timely manner, and the same happens on Upcoming, which gives visibility to events entered by friends; on Last.fm for songs recently played by friends and on Del.icio.us for URLs recently bookmarked by subscribed users. In a similar way, the Epinions "Web of Trust" and "Block List" give an immediate benefit to the user: when an offensive or unreliable review is found, the user can simply add the reviewer into his/her Block List, telling the system he/she does not want to see his/her reviews again. On the opposite side, users who create interesting, useful reviews can be placed in the "Web of Trust" so that their reviews are given more prominence. An alternative way for using trust statements would be to exploit already existing information instead of asking it directly of the user. This was the path Google's founders followed when they created PageRank (Brin & Page, 1998). Links between Web pages were already there and PageRank intuition was to consider a link from page *A* to page *B* as a vote of *A* on *B*, or as a trust statement in our jargon.

How to Exploit Trust in Online Systems

In the previous subsection, we discussed challenges in modeling trust in online systems. In this one we assume the trust relationships information is available and concentrate on ways of exploiting it. Based on the subjective trust statements provided by users, we can in fact aggregate the complete trust network (see Figure 1). Trust Metrics (Golbeck et al., 2003; Levien, n.d.; Massa & Avesani, 2004; Ziegler & Lausen, 2004) and Reputation Systems (Resnick et al., 2000) can then be used in order to predict a trust value for every other user based on the opinions of all the users. An important classification of Trust Metrics (TM) is in local and global (Massa & Avesani, 2004; Ziegler & Lausen, 2004). Global TMs predict the same value of trustworthiness of *A* for every user. PageRank (Brin & Page, 1998) is a global Trust Metric: the PageRank of the Web page Microsoft.com is, for example, 9/10 for everyone, notwithstanding what the active user querying Google likes and dislikes. Sometimes this identical value for all the members of the community is called reputation: ”reputation is what is generally said or believed about a person's or thing's character or standing” (Oxford Dictionary). On the other hand, local TMs are personalized: they predict the trustworthiness of other users from the point of view of every single different user. A personalized PageRank (Haveliwala, Kamvar, & Jeh, 2003) would predict different trust values for the Web page Microsoft.com for a user who appreciates (trusts) GNU/Linux and a user who appreciates Windows. In fact, a trust statement is personal and subjective: it is absolutely possible for user Bob to be trusted by user Alice and distrusted by Carol, as it is the case in the simple trust network depicted in Figure 1. Actually, it is normal to have, in a real community, controversial users: users trusted by many other users and distrusted by many other ones (Massa & Avesani, 2005). We have seen that reputation is a global, collective measure of trustworthiness (in the sense of reliability) based on trust statements from all the members of the community. Surely, in many situations, it is important to compute an average value. For example, different people can have different opinions on who is the best physicist of the year and nominate different ones. However, only one physicist can get the Nobel Prize and it should be the one that is more appreciated by the community as a whole. The same happens when someone might be elected president of a country: different people would have different preferences but they must be averaged using a global metric in order to identify just one person who will become president. Actually, most of the system we reviewed in section 1 uses a global metric: eBay, Slashdot (on which reputation is called Karma), Kuro5hin (on which it is called Mojo), PageRank, and many others. Reputation is a sort of status that gives additional powers and capabilities in the online system, and it can even be considered a sort of currency. In fact, Cory Doctorow, in his science-fiction novel *Down and Out in the Magic Kingdom* (Doctorow, 2003) already envisioned a postscarcity economy in which all the necessities of life are free for the taking, and what people compete for is whuffie, an ephemeral, reputation-based currency. A person's current whuffie is instantly viewable to anyone, as everybody has a brain-implant giving them an interface with the Net. The usual economic incentives have disappeared from the book's world. Whuffie has replaced money, providing a motivation for people to do useful and creative things. A person's whuffie is a general measurement of his or her overall reputation, and whuffie is lost and gained according to a person's favorable or unfavorable actions. Note that he/she also acknowledges that a personalized and subjective whuffie can be useful as well, weighting opinions of other people differently depending on one's subjective trust in them. Even if this does not refer to how trust is used and modeled by current online systems, it is an interesting speculation into one of the possible futures and the central role trust would play in it. Coming back to current online systems, Epinions provides personalized results and filtering and hence, exploits a local Trust Metric, even if precise details about how it is implemented and used are not public (Guha, 2003).

It is worthwhile noting that the largest portion of research papers studying reputation and trust often run simulations on synthesized data representing online systems, and often assume that there are "malicious" peers and "good" peers in the system, and that the goal of the system is just to allow good peers to spot out and isolate malicious peers. In this sense, they also often assume there are "correct" trust statements (a good peer must be trusted) and "wrong" or "unfair" trust statements (if a peer does not trust a good peer, he/she is providing a wrong rating). We would like to point out how these synthesized communities are unrealistic and how reality is more complicated than this; see, for example, a study on controversial users on Epinions (Massa & Avesani, 2005). Assuming that there are globally agreed good peers and that peers who think differently from the average are malicious encourages herd behavior and penalizes creative thinkers, black sheep, and original, unexpected opinions. This is in essence the "tyranny of the majority" risk, a term coined by Alexis de Tocqueville in his book, *Democracy in America* (1835) (de Tocqueville, 1840). The 19th century philosopher John Stuart Mill, in his philosophical book *On Liberty* (Mill, 1859), analyzes this concept with respect to social conformity. Tyranny of the majority refers to the fact that the opinions of the majority within society are the basis of all rules of conduct within that society. On a particular issue, people will align themselves either for or against this issue; the side of greatest volume will prevail, but this does not mean the other side is wrong. So for one minority, which by definition has opinions that are different from the ones of the majority, there is no way to be protected "against the tyranny of the prevailing opinion and feeling" (Mill, 1859). However, we believe the minority's opinions should be seen as an opportunity and as a point of discussion and not as "wrong" or "unfair" ratings, as often they are modeled in research simulations. However, there is a risk on the opposite extreme as well and it is called "echo chamber" or "daily me" (Sunstein, 1999). Sunstein notes how "technology has greatly increased people's ability to filter what they want to read, see, and hear" (Sunstein, 1999). He warns how in this way, everyone has the ability to just listen and watch what he/she wants to hear and see, to encounter only opinions of like-minded people and never again be confronted with people with different ideas and opinions. In this way, there is a risk of segmentation of society into microgroups who tend to extremize their views, develop their own culture and language, and not be able to communicate with people outside their group anymore. He argues that "people should be exposed to materials that they would not have chosen in advance. Unplanned, unanticipated encounters are central to democracy itself ," and that "many or most citizens should have a range of common experiences. Without shared experiences, ... people may even find it hard to understand one another" (Sunstein, 1999). Finding the correct balance between these two extremes is surely not an easy task, but something that must be taken into account both for systems designers and researchers.

Creating scalable Trust Metrics. A challenge for local Trust Metrics is to be time efficient, that is, to predict trustworthiness of unknown peers in a short time. In fact, in general, local Trust Metrics must be run one time for every single user propagating trust from his/her point of view, while global ones are just run once for the entire community. In this sense, the load placed on a centralized system (for example, on Google) for predicting the trust scores for every user as seen by every other user seems to be too large to be handled. We believe a much more meaningful situation is the following: every single user predicts the trust scores he/she should place in other users from his/her personal point of view and on his/her behalf. In this way, every user is in charge of aggregating all the trust statements he/she deems relevant (and in this way, he/she can, for example, limit himself/herself to just fetch information expressed by friends of friends) and run the local Trust Metric on this data just for himself/herself on his/her local device, his/her server or, in the short future, his/her mobile.

Exploiting negative trust statements. We mentioned earlier the challenges in modeling negative trust statements and how few systems attempt to do it. For this reason, research about how to exploit distrust statements is really in its infancy. The lines of early inquiry at the moment are limited to the already cited studies on eBay's feedback system (Resnick & Zeckhauser, 2002), to propagation of distrust (Guha, Kumar, Raghavan, & Tomkins, 2004), and analysis on controversial users (Massa & Avesani, 2005).

Visualization of trust network for explanation. Another open challenge is related to visualization and explanation of how the system used trust information, especially if this affects the user experience. For example, it is important that the user is aware of the reason a certain review is shown, especially if the system's goal is to let to the user be able to master and guide the process and provide additional information. Visualizing the social network, for example showing to the user a picture similar to Figure 1, might be a powerful option to give awareness to the user of his/her position in the network, and to let him/her navigate it. Surely this kind of interface promises to be useful and enjoyable (see for example, a study on visualization of Friendster network in Heer & Boyd, 2005). However, we note that none of the online systems we introduced earlier use them: the reasons might be that these interfaces are not easily doable with standard HTML, but at the moment require the browser to use external plugins (for example, supporting Java applets, Flash, or SVG), and in this way they also break standard browsing metaphors and linking patterns. Moreover, creating a visualization tool easily understandable and usable is a very difficult task.

Public details of the used algorithms. Another challenge we think should be overcome is related to "security through obscurity" principle. Security through obscurity refers to a situation in which the internal functioning of an artifact is kept secret with the goal to ensure its security. However, if a system's security depends mainly on keeping an exploitable weakness hidden, then, when that weakness is discovered, the security of the system is compromised. In some sense, the system is not secure; it is just temporarily not compromised. This flaw is well acknowledged in cryptography: Kerckoffs' law states that: "a cryptosystem should be secure even if everything about the system, except the key, is public knowledge." Most of the systems we reviewed adopt the security through obscurity principle in the sense that the precise details of how they exploit trust information are kept secret. For example, PageRank is left intentionally obscure. There are early reports about its functioning (Brin & Page, 1998), but Google does not disclose the used algorithm (probably different from the original one) and in particular the parameters used to fine-tune it. Epinions follows the same "security through obscurity" principle: "Income Share is determined by an objective formula that automatically distributes the bonuses. The exact details of the formula must remain vague, however, in order to limit gaming or other attempts to defraud the system" (Epinions.com earnings FAQ, n.d.). Interesting exceptions are Slashdot and Kuro5hin, whose code is free software released under the GNU General Public License and available respectively at SlashCode (http://slashcode.org) and Scoop (http://scoop.kuro5hin.org). In a similar way, Advogato Trust Metric is described in detail (Levien, n.d.), and the code is available on the Advogato Web site as well. Of course, one problem is related to the fact that commercial companies do not want to disclose their secret algorithms because this would allow any competing company to copy them. Luckily this is not a problem for non-commercial online systems and for systems that do not rely on a central server. However, we believe that a user should be able to know how recommendations are generated (for example, for checking if the system introduces undesired biases) and, in case she desires it, to use trust information as she prefers. We will touch this topic briefly in the following section on walled gardens.

Identities, Privacy, and Attacks

Identity, privacy, and attacks are huge topics by themselves and in this subsection, we are just going to scrape the surface and touch on challenges related to online systems that model and use trust. In general, on these systems, users act under pseudonyms (also called nicknames or usernames). Seldom, the real-world identity of the person using the online system is verified by the system because this would create a huge access barrier, cause great costs, and slow down the process of creating an identity, in a significant way. Unless there is a great need for the user to enter the system, this will drive him/her away and to the next easier-to-enter online system.

As long as a user has some way to decide if another user (as represented by their nickname) is trustworthy, this is often enough. A partial exception in this is represented by eBay. An eBay user can enter credit card details and in this way, eBay can tie the pseudonym with that credit card so that it can be possible to find the person in the real world in case this is needed for some reason, for example, an accusation of fraud or a law suit.

Pseudonymity. Pseudonymity is of course a situation that marks a striking difference between online systems and real world. In real-world interactions, almost always the identity of the other person is known, while this is really the exception in online systems, where it is often possible to interact, communicate, and make business with other users who will never be met in person. In general, users can enter some details that describe themselves, and the system shows this information in their profile page. Note that users can lie in providing this information; for example, a survey found that 24% of interviewed teens that have used instant messaging services and e-mail or been to chat rooms have pretended to be a different person when they were communicating online (Lenhart, Rainie, & Lewis, 2001). The profile page of a user often shows a summary of recent activity in the system and social relationships with other users (Donath & Boyd, 2004), and usually this is the only information available and other users will form an opinion of that user based on this information. The effect of pseudonymity is well captured in the popular cartoon depicting two dogs in front of a computer with one dog saying to the other dog "On the Internet, nobody knows you're a dog." Of course, this situation works until there are no problems, but in case something goes wrong (accusation of fraud, molestation, or any accusation of illegal activity), it is required to identify the real-world identity, and this is not always easy. Moreover, different legal systems make it hard to have justice for crimes perpetuated in the virtual world.

Multiple identities. It is also common for a person to have more than one identity in an online system (Friedman & Resnick, 2001). A recent survey (Aschbacher, 2003) found that users of an informal science learning Web Site have more than one identity (60% of girls vs. 41% of boys) on the site. Respondents gave various reasons for the multiple identities including sharing them with school friends, using them to earn more points on the site, and just trying out different identities from day to day. This behavior is quite common in social sites.

Fake identities and attacks. Moreover, sometimes humans create fake identities (also known as fakester) (Boyd, 2004) such as identities representing famous people. However, besides playful reasons, often these multiple identities in control of a single human being are used to game the system. This behavior is often called "pseudospoofing" (a term first coined by L. Detweiler on the Cypherpunks mailing list) or "sybil attack" (Douceur, 2002). Usually these fake identities are used in a concerted way and collaborate with each other in order to achieve a certain result on the system. For example, a person might use them to submit many positive ratings for a friend in order to boost his/her reputation (positive shilling), or negative ratings for the competition in order to nuke his/her reputation (negative shilling) (Oram, 2001).

These multiple identities can also be used, for example, by a book's author for writing many positive reviews of his/her own book. At least an occurrence of this behavior has been revealed publicly because of a computer "glitch" that occurred in February 2004 on the Canadian Amazon site. This mistake revealed for several days the real names of thousands of people who had posted customer reviews of books under pseudonyms (Amazon glitch out, 2004), and it was possible to note that many reviews made about a certain book were in reality created by the author of that book using many different pseudonyms. This possibility seriously mines at the basis the functioning of opinion-sharing sites. Another similar attack occurs on the Web: a link farm is a large group of Web pages that contain hyperlinks to one another or a specific other page. Link farms are normally created by programs, rather than by human beings, and are controlled by a single principal. The purpose of a link farm is to fool search engines into believing that the promoted page is hugely popular, and hence the goal is to maliciously increase its PageRank. A considerable amount of research is devoted to designing methods to spot out these attacks. For example, TrustRank (Gyongyi, Molina, & Pedersen, 2004) is a technique proposed by researchers from Stanford University and Yahoo! to semiautomatically separate reputable, good pages from spam.

Another possible way to deal with link farms is to enrich the language for expressing links, that is HTML (hypertext markup language). In fact, a common practice for increasing the score of a certain page is to use programs that automatically insert links to that page on blogs (in the form of comments) and wikis available on the Web. In order to counter this practice, in early 2005, Google proposed a new solution suggesting that blog and wiki engines should add to every link not directly created by the blog and wiki author a rel="nofollow" attribute. This attribute of the <a> HTML element is a explicit way to tell search engines that the corresponding link should not be considered as a "vote" for the linked page, or a trust statement in our jargon. A related initiative is VoteLinks Microformat (Technorati.com, n.d.), which enriches HTML by proposing a set of three new values for the rev attribute of the <a> HTML element. The values are vote-for, vote-abstain, and vote-against, and represent agreement, abstention or indifference, and disagreement respectively. In fact, as already noted, PageRank's assumption is that a link from page *A* to page *B* is a vote of *A* on *B*. However, this means that a link created with the purpose of critiquing the linked resource is increasing its PageRank score, and this might induce the author to not link to the criticized page. In short, attention is not necessarily appreciation (Massa & Hayes, 2005). With VoteLinks, it would be possible to tell search engines the reason behind a link so that they could create more useful services. Considering these proposals from a trust point of view, nofollow would express "this is not a trust statement, do not consider it" and VoteLinks would allow authors to express weighted trust statements in a linked page: vote-for is trust, vote-against is distrust, vote-abstain is similar to nofollow. It is interesting to note that Google's nofollow proposal was adopted by most search engines and blog and wiki programs in a few weeks, while VoteLinks proposal seems very little used. This has to do a lot with the authority of the proponent and a little with the proposal itself.

As we already said, Local Trust Metrics can be effective in not letting untrusted nodes influence trust scores (Levien, n.d.; Massa & Avesani, 2004; Ziegler & Lausen, 2004) and in fact, there is research into personalizing PageRank (Haveliwala, Kamvar, & Jeh, 2003) as well. OutFoxed (James, 2005) is exploring ways for a user to use his/her network of trusted friends to determine what is good, bad, and dangerous on the Web. This is done by adding functionality to the Firefox Web browser who is able to predict the trust score of Web pages based on opinions of trusted friends.

Another possible attack is the following. A user could "play" the good behaviored role for a while with an identity and gain a good trust and reputation through a series of perfectly good deals, then try to complete a fraud and eventually drop the identity to start again with a new one. This has been reported at least once in a mid-2000 eBay fraud in which the user "turned evil and cashed out" (Wolverton, 2000). Friedman and Resnick (2001) analyze the phenomenon of multiple pseudonyms and conclude that, in systems in which new pseudonyms can be acquired for free, since new logins could be malicious users who just dropped an identity, the starting reputation of newcomers should be as low as possible. They prove that "there is no way to achieve substantially more cooperation in equilibrium than that achieved by distrusting all newcomers. Thus, the distrust of newcomers is an inherent social cost of easy identity changes."

Local Trust Metrics (Massa & Avesani, 2004'Ziegler & Lausen, 2004) can solve the problem introduced by multiple identities. Since with local Trust Metrics only trusted users (or users trusted by trusted users) are considered, the activity of fake identities not reached by the trust propagation does not influence the active user. In fact, attack resistance of Trust Metrics and Reputation Systems is a very important topic that is starting to receive great attention only recently, probably because of the complexity of the problem itself. Some Trust Metrics are claimed to be resistant to some attacks, for example Advogato (Levien, n.d.): "If a bunch of attackers were to create lots of accounts and mutually certify each other, only a very few would be accepted by the Trust Metric, assuming there were only a few certificates from legitimate members to the hackers." On the other hand, eBay metric (Resnick et al., 2003) is a very simple one, and we have seen that many attacks can be easily mounted against it.

However, it seems to work well in practice, and surely one of the reasons is that, because of its simplicity, every user can understand how it works and get some confidence in the functioning of the system: more complicated metrics would be harder to understand and the user would probably lose confidence in the system altogether. In fact, Resnick and Zeckhauser (2002) consider two explanations related to the success of eBay's feedback system: (1) "The system may still work, even if it is unreliable or unsound, if its participants think it is working. (...) It is the perception of how the system operates, not the facts, that matters" and (2) "Even though the system may not work well in the statistical tabulation sense, it may function successfully if it swiftly turns against undesirable sellers (...), and if it imposes costs for a seller to get established." They also argue that: "on the other hand, making dissatisfaction more visible might destroy people's overall faith in eBay as a generally safe marketplace." This seems confirmed by a message posted on eBay by its founder in 1996: "Most people are honest. And they mean well. Some people go out of their way to make things right. I've heard great stories about the honesty of people here. But some people are dishonest: or deceptive. This is true here, in the newsgroups, in the classifieds, and right next door. It's a fact of life. But here, those people can't hide. We'll drive them away. Protect others from them. This grand hope depends on your active participation" (Omidyar, 1996). On eBay, whose goal, after all, is to allow a large number of commercial transactions to happen, it seems that positive feelings and perceptions can create a successful and active community more than a sound Trust Metric and reputation system. This means that the fact that a Trust Metric or reputation system is proved to be attack resistant does not have an immediate effect on how users perceive it and hence, on how this helps in keeping the community healthy and working.

Another problem with online identities is represented by "identity theft." This refers to the ability of someone to get in control of someone elses identity on an online system. We have seen already how a reputable identity on eBay is valuable by an average 7.6% increase of selling price (Resnick et al,, 2003), and this gives a reason for trying to get into control of them. This phenomenon is also called "account hijacking" and usually happens by phishing or by password guessing. Since online identities have an economic value, they are also sold for real money, often on e-marketplaces.

Privacy. Privacy is another huge issue for online systems, and here we are just going to discuss its main implications. Who can access information users express in an online system undoubtedly modifies which kind of information they will be willing to express and how they express it. As we have already seen, fear of retaliation for a negative trust statement has the consequence of very few negative ratings on eBay and, for this reason for example, Epinions distrust list (Block List) is kept secret and not visible. Moreover, trust statements can also be used to model access permission to published information. For example, on Flickr it is possible to make some photo visible only to contacts, friends, or family members. The topic of privacy is very large and has huge psychological implications we cannot address here for reasons of space. Also note that private information a user expresses in an online system can be disclosed by error, as the previously cited example of Amazon Canada showed. The best possible situation for users would be to remain in total control and possession of their information (not only trust statements), and to upload it and show it to who the user wants, when he/she wants.

Portability and interoperability. And in fact, the next challenge we are going to comment about is related to portability of trust and reputation scores across walled gardens. Let us consider the following situation. A person utilizes eBay for some years, provides a lot of trust statements and, even more importantly from his/her point of view, receives a lot of trust statements: he/she built up a good reputation and is recognized by the community. If then, for some reason, he/she would like to change e-marketplaces (for example, eBay could close its operations or the user could prefer a new system that applies smaller fees), he/she has no choice but to start from scratch: there is no way he/she can migrate with his/her activity history (the information he/she entered in the system) and his/her reputation. This is because his/her information is not under his/her control, but under the control of the online system: he/she does not own the information. Clearly, the value of an online system is in the network of users that are using it, and companies prefer to not allow interoperability because competitors would use this information to bootstrap their networks. For example, eBay started a law suit against another e-marketplace who was copying the information about users and their feedback from theeBay Web site or that was, according to eBay, "engaging in the wholesale copying of our content and then using that content without our permission" (Sandoval & Wolverton, 1999). We believe that the content is the users' content and not eBay's content and in fact, users would have all the advantages letting different online systems compete to provide useful and cheap services with the information they expressed. We already discussed about semantic formats for letting users express, on their servers and under their control in a decentralized way, information (for example about the people they know using FOAF or XFN). These attempts have still to gain momentum, and it is surely easier to manage information about millions of users on a centralized server (as eBay, Epinions, Amazon, and almost all the systems we reviewed do at the moment) because there are no problems with formats, retrieval, and update. An attempt to achieve portability of trust and reputation across communities was Affero (see description in Section 2), but it seems it did not reach a critical mass and has very few users. However, we note how many of the online systems we reviewed in the previous section are starting to expose application programming interfaces (API) so that the precious data users entered into the system can be extracted by them for backups and for migration and, even more interestingly, can be used by independent programs and systems. Flickr, Del.icio.us, Upcoming, and many more systems have already done this or are in the process of doing it and this fact, instead of endangering their existence, has favored a plethora of independent services that are adding value to the original systems.

We believe users are starting to understand that the data they inserted into an online system really belong to them and not to the system, and they will be requiring more and more possibility of directly managing these data and getting them back. When all the systems will export this information, it will be possible to aggregate all the different domains in which a user is acting and get an overall perception of his/her activity in the online world, as the Firefox extension IdentityBurro tries to do (Massa, 2005).

Conclusion

In this chapter, we have presented a classification and prototypical examples of online systems that model and use trust information. We have also discussed what are the most important challenges for these systems and some possible solutions. This domain is very active and new initiatives, both commercial startups and research studies, are proposed continuously. New service metaphors and algorithms are invented daily, also based on feedback from users who are becoming more and more aware of their needs. It seems unreasonable to claim that a single approach might fit all the different scenarios we presented in this chapter, and the ones that will emerge in future. Instead, the designers of the online communities will have to continuously rethink basic mechanisms and readapt them to the different needs that emerge. Nevertheless, learning from past experiences, successes, and failures is an important activity, and this is what we tried to do with this chapter. Modeling and exploiting trust in online systems is and will remain an exciting, ongoing challenge.

Acknowledgments

We would like to thank Riccardo Cambiassi for helpful feedback on this chapter and discussions. We would like to thank Jennifer Golbeck for compiling and making public a list of trust-enabled platforms (Golbeck, 2005) that was a good starting point for this chapter. We would also like to thank the creators of the online systems and Web sites we reviewed and commented. Without them, this chapter would have not been, and the Web would have been a less interesting place.

References

Adamic, L. A., Buyukkokten, O., & Adar, E. (2003). A social network caught in the web. *First Monday, 8*(6).

Adar, E., & Huberman, B. (2000). *Free riding on Gnutella*. Technical report, Xerox PARC.

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*(3), 488-500.

Allen, C. (2004). *My advice to social networking services*. Retrieved December 28, 2005, from http://www.lifewithalacrity.com/2004/02/my_advice_to_so.html

Amazon glitch outs authors reviewing own books. (2004). Retrieved December 28, 2005, from http://www.ctv.ca/servlet/ArticleNews/story/CTVNews/1076990577460_35

Aschbacher, P. R. (2003). Gender differences in the perception and use of an informal science learning Web site. *Journal of the Learning Sciences, 9*.

Boyd, D. (2004). Friendster and publicly articulated social networking. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems* (pp. 1279-1282). New York, NY: ACM Press.

Brin, S., & Page, L. (1998). The anatomy of a largescale hypertextual Web search engine. In *WWW7: Proceedings of the Seventh International Conference on World Wide Web 7*. Elsevier Science Publishers.

Chen, R., & Yeager, W. (2001). *Poblano: A distributed trust model for peer-to-peer networks*. Technical report, Sun Microsystems.

Coates, T. (2003). *(Weblogs and) the mass amateurisation of (nearly) everything*. Retrieved December 28, 2005, from http://www.plasticbag.org/archives/2003/09/weblogs_and_the_mass_amateurisation_of_nearly_everything.shtml

Cohen, B. (2003). Incentives build robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, USA.

Cornelli, F., Damiani, E., De Capitani di Vimercati S., Paraboschi, S., & Samarati, P. (2002). Implementing a reputation-aware Gnutella servent. In *International Workshop on Peer-to-Peer Computing.*

de Tocqueville, A. (1840). *Democracy in America*. (G. Lawrence, Trans). New York: Doubleday.

Doctorow, C. (2003). *Down and out in the Magic Kingdom*. Tor Books.

Donath, J. & Boyd, D. (2004). Public displays of connection. *BT Technology Journal, 22*(4).

Douceur, J. (2002). The Sybil attack. In *Proceedings of the 1st International Peer-To-Peer Systems Workshop (IPTPS).*

eBay help: Feedback extortion. (n.d.). Retrieved December 28, 2005, from http://pages.ebay.co.uk/help/policies/feedback-extortion.html

Epinions.com earnings FAQ. (n.d.). Retrieved December 28, 2005, from http://www.epinions.com/help/faq/show_faq_earnings

Epinions.com Web of Trust FAQ. (n.d.). Retrieved December 28, 2005, from http://www.epinions.com/help/faq/?show=faq_wot

Friedman, E. J., & Resnick, P. (2001). The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy, 10*(2), 173-199.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Free Press Paperbacks.

Golbeck, J. (2005). *Web-based social network survey*. Retrieved December 28, 2005, from http://trust.mindswap.org/cgibin/relationshipTable.cgi

Golbeck, J., Hendler, J., & Parsia, B. (2003). Trust networks on the semantic Web. In *Proceedings of Cooperative Intelligent Agents.*

Gross, B., & Acquisti, A. (2003). *Balances of power on eBay: Peers or unequals? The Berkeley Workshop on Economics of Peer-to-Peer Systems*. Berkeley, CA.

Guha, R. (2003). *Open rating systems*. Technical report, Stanford University, CA, USA.

Guha, R., Kumar, R., Raghavan, P., & Tomkins, A. (2004). Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th Int. Conf. on World Wide Web* (pp. 403-412). ACM Press.

Gyongyi, Z., Molina, H. G., & Pedersen, J. (2004). Combating Web spam with TrustRank. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases (VLDB)* (pp. 576-587). Toronto, Canada: Morgan Kaufmann.

Haveliwala, T., Kamvar, S., & Jeh, G. (2003). An analytical comparison of approaches to personalizing PageRank. In *WWW '02: Proceedings of the 11th Int. Conference on World Wide Web*. ACM Press.

Heer, J., & Boyd, D. (2005). Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization (InfoViz).*

James, S. (2005). *Outfoxed: Trusted metadata distribution using social networks*. Retrieved December 28, 2005, from http://getoutfoxed.com/about

Jøsang, A., Ismail, R., & Boyd, C. (2005). A survey of trust and reputation Systems for online service provision. In *Decision Support Systems*.

Kamvar, S. D., Schlosser, M. T., & Garcia-Molina, H. (2003). The Eigentrust algorithm for reputation management in P2P Networks. In *WWW'03 Conference*.

Koman, R. (2005). *Stewart Butterfield on Flickr.* Retrieved December 28, 2005, from www.oreillynet.com/pub/a/network/2005/02/04/sb_flckr.html

Kuchinskas, S. (2005). *Amazon gets patents on consumer reviews.* Retrieved December 28, 2005, from http://www.internetnews.com/bus-news/article.php/3563396

Lee, S., Sherwood, R., & Bhattacharjee, B. (2003). Cooperative peer groups in NICE. In *IEEE Infocom.*

Lenhart, A., Rainie, L., & Lewis, O. (2001). *Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships.* Retrieved December 28, 2005, from http://www.pewinternet.org/report_display.asp?r=36

Levien R. (n.d.). *Attack resistant trust metrics*. Retrieved December 28, 2005, from http://www.advogato.org/trust-metric.html

Massa, P. (2005). *Identity burro: Making social sites more social*. Retrieved December 28, 2005, from http://moloko.itc.it/paoloblog/archives/2005/07/17/identity_burro_greasemonkey_extension_for_social_sites.html

Massa, P., & Avesani, P. (2004). Trust-aware collaborativefiltering for recommender systems. In *Proceedings of Federated Int. Conference On The Move to Meaningful Internet: CoopIS, DOA, ODBASE*.

Massa, P., & Avesani, P. (2005). Controversial users demand local trust metrics: An experimental study on Epinions.com community. In *Proceedings of 25th AAAI Conference*.

Massa, P., & Hayes, C. (2005). Page-rerank: Using trusted links to re-rank authority. In *Proceedings of Web Intelligence Conference.*

Mill, J. S. (1859). *On Liberty. History of Economic Thought Books*. McMaster University Archive for the History of Economic Thought.

Mui, L. (2002). *Computational models of trust and reputation: Agents, evolutionary games, and social networks.* PhD thesis, Massachusetts Institute of Technology.

Omidyar, P. (1996). *Ebay founders letter to eBay community*. Retrieved December 28, 2005, from http://pages.ebay.com/services/forum/feedback-foundersnote.html

Oram, A., editor (2001). *Peer-to-peer: Harnessing the power of disruptive technologies*. O'Reilly and Associates.

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. *The Economics of the Internet and Ecommerce. Advances in Applied Microeconomics, 11*.

Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. (2000). Reputation systems. *Communication of the ACM, 43*(12).

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2003). *The value of reputation on eBay: A controlled experiment.*

Salton, G., & McGill, M. J. (1986). *Introduction to modern information retrieval*. NY: McGraw-Hill, Inc.

Sandoval, G. & Wolverton, T. (1999). *EBay files suit against auction site bidder's edge*. Retrieved December 28, 2005, from http://news.com.com/2100-1017-234462.html

Shirky, C. (2000). *What is P2P ... and what isn't?* Retrieved December 28, 2005 from http://www.openp2p.com/pub/a/p2p/2000/11/24/shirky1-whatisp2p.html

Shirky, C. (2004). *Multiply and social spam: Time for a boycott*. Retrieved December 28, 2005, from http://many.corante.com/archives/2004/08/20/multiply_and_social_spam_time_for_a_boycott.php

Steiner, D. (2004). *Auctionbytes survey results: Your feedback on eBay's feedback system*. Retrieved December 28, 2005, from http://www.auctionbytes.com/cab/abu/y204/m11/abu0131/s02

Sunstein, C. (1999). *Republic.com*. Princeton, NJ: Princeton University Press.

Technorati.com. (n.d.). *VoteLinks*. Retrieved December 28, 2005, from http://microformats.org/wiki/votelinks

XHTML Friends Network. (n.d.). Retrieved December 28, 2005, from http://gmpg.org/xfn/

Wolverton, T. (2000). *EBay, authorities probe fraud allegations*. Retrieved December 28, 2005, from http://news.com.com/2100-1017_3-238489.html

Ziegler, C., & Lausen, G. (2004). Spreading activation models for trust propagation. In *IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE'04).*