Trust metrics on controversial users: balancing between tyranny of the majority and echo chambers

Paolo Massa

Paolo Avesani
ITC-iRST
Via Sommarive 14 - I-38050 Povo (TN) - Italy
{massa,avesani}@itc.it

Abstract

In today's connected world it is possible and indeed quite common to interact with unknown people, whose reliability is unknown. Trust Metrics are a technique for answering questions such as "Should I trust this person?". However, most of the current research assumes that every user has a global quality score everyone agree on and the goal of the technique is just to predict this correct value. We show, on data from a real and large user community, *Epinions.com*, that such an assumption is not realistic because there is a significant portion of what we call controversial users, users who are trusted by many and distrusted by many: a global agreement about the trustworthiness value of these users does not exist. We argue, using computational experiments, that the existence of controversial users (a normal phenomenon in complex societies) demands local trust metrics, techniques able to predict the trustworthiness of a user in a personalized way, depending on the very personal views of the judging user as opposed to most commonly used global trust metrics which assume a unique value of reputation for every single user. The implications of such an analysis deal with the very foundations of what we call society and culture and we conclude discussing this point by comparing the two extremes of culture that can be induced by the two different kinds of trust metrics: tyranny of the majority and echo chambers.

1 Introduction

In today's connected world, it is possible and common to interact with unknown people. This happens for example when contacting a stranger via her email address found on the Web, using a site that allows messaging between users or reading, on an opinions site, a review of a product written by someone we don't know.

In this *uncertain* world, it is necessary to take into account questions such as "Should I trust this person?" in order to decide to interact with or rely on her. The emerging way of dealing with this new requirement is to allow all the users to express their level of trust on other users, aggregate this information and reason about it [12]. This intuition is exploited, for instance, in modern search

engines such as *Google.com*, that considers a link from one site to another as a vote, i.e. an expression of trust [18], in e-marketplaces such as *Ebay.com*, that allows users to express their level of satisfaction after every interaction with another user and has been suggested for peer-to-peer systems where peers keep a history of interactions with other peers and share this information with the other peers [2].

A considerable amount of research has been carried on recently on these and related topics, such as trust metrics [7, 21, 11, 13, 9], reputation systems [19], and personalizing PageRank [10].

However most of the current research takes the assumption that every user[1] has an objective trustworthiness value and the goal of the techniques is just to guess this correct value. Conversely, we think that such an assumption is misleading. We argue that these techniques should take into account the fact that different users can have different opinions about a specific user.

Hence we distinguish between global and local trust metrics [13, 21]. Both try to predict the trustworthiness[2] of a given user. Global trust metrics assign to a given user a unique trust score, the same independently of the user that is evaluating the other user's trustworthiness. On the other hand, a local trust metric provides a personalized trust score that depends on the point of view of the evaluating user.

In this paper, we will devote special attention to *Controversial Users*. A controversial user is a user that is judged by other users in very diverse ways, for example, she is trusted or appreciated by many and is distrusted or negatively rated by many.

Hence, the goal of the paper is to investigate the differences between global and local trust metrics, concentrating on controversial users and considering situations where one technique is more appropriate than the other. Data from the large *Epinions.com* community confirm our hypothesis that in complex societies there is a non negligible percentage of Controversial Users. The empirical analysis we conducted shows that a local trust metric achieves higher accuracy than a global one in predicting the trust a specific user should place into a controversial user. We argue how there is an intrinsic limit in the maximum accuracy of every global trust metric that is not present for local trust metrics and this is particularly evident on controversial users.

The implications of such an analysis deal with the very foundations of what we call society and culture and we conclude the paper discussing this point by comparing the two extremes of culture that can be induced by these different kinds of trust metrics: tyranny of the majority and echo chambers.

The rest of the paper is structured as follows. First, we provide definitions for concepts such as trust and trust network. Then we explain trust metrics focusing especially on the differences between global and local trust metrics and the culture they produce in a community of people: tyranny of the majority or echo chambers. We then introduce MoleTrust, the local trust metric we chose

[1] From now on, we will use the term "user" in order to indicate an autonomous entity able to express and receive trust statements. However the same concepts apply also if considering peers in a peer-to-peer system, interacting servers, agents in a multi agents systems or communicating mobile devices such as mobiles and robots.

[2] Different authors use different terms. Others used terms are authority, reputation and reliability. The terms often represent the same concept but are used in slightly different contexts. We use trust and trustworthiness.

for the experiments. Finally we presents the experiments we carried out and discuss the results and the implications of the empirical findings.

2 Societies as Trust Networks

In the previous section we hinted how, in the so called "global village" enabled by the Internet, it is becoming common to interact with unknown people. In this new context, there is a need for new tools that can help in deciding about the trustworthiness of other people. 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, i.e. to share the burden to evaluate them. It is in fact the case that most of the community Web sites nowadays let an user express her opinions about every other user, asking how much she finds her interesting and worth her attention [12]. We call these expressed opinions, trust statements. Actually, trust has been described as the "constitutive virtue of, and the key causal precondition for the existence of any society" [5]. Other contributions from economy highlights how trust is directly correlated with the "creation of prosperity" [6] or suggest to consider trust as a commodity [3]. Reasoning on trust statements issued by other people has also been suggested as a way to balance the "information asymmetry" [1] between the two parties, buyer and seller, involved in a potential economic transaction [19].

In fact the Internet exhibit a huge advantage for information dissemination over traditional word of mouth mechanisms: all the trust statements can be made publicly and permanently visible and fetchable, possibly by everyone or only by some specific users. It is worthwhile noting that the topic of trust, its meanings and its possible uses have been a topic of interest for many different fields for centuries. This shouldn't surprise since trust is a very human and social concept and it is in the speculations of men since the first philosophers and surely before the advent of the first computers. Trust has been studied by researchers and thinkers in different fields other than computer science such as economics, scientometrics, politics, evolutionary biology, anthropology, philosophy and sociology (see [17] for a survey).

In our conceptualization of the domain, we define a *trust statement* as the explicit opinion expressed by an user about another user regarding the perceived quality of a certain characteristics of this user. For example, on a site where users contribute reviews about products, users could be asked to express a positive trust statement on a user "whose reviews and ratings they have consistently found to be valuable" and a negative trust statement on "authors whose reviews they find consistently offensive, inaccurate, or in general not valuable"[3]. We model trust as a real value in the interval $[0, 1]$, where $T(A, B) = 0$ means that A has issued a statement expressing that her degree of trust in B is the minimum, i.e. she totally distrusts B. On the other hand, $T(C, B) = 1$ means that C totally trusts B. Trust statements are subjective: a user can receive different trust scores from different users. They are also asymmetric in the sense that, if A trusts B as 0.8, this does not mean that B has to trust A as 0.8 as well, but, for example, it might be even the case that B does not know A.

[3]This is precisely what the Epinions.com Web of Trust FAQ page suggests (http://www.epinions.com/help/faq/?show=faq_wot).

In most settings, a user has a direct opinion only about a very small portion of users. The remaining users are unknown users.

Note that a number of formats for letting expressing and encoding trust statements in an explicit manner are starting to emerge. From the Semantic Web community, it is worth mentioning the trust extension of the FOAF[4] format presented in [7]. A more simple and limited by design effort is represented by the VoteLink and XFN microformats [15] that allows to add simple semantic information directly in XHTML links, expressing an opinion about or a social relationship with the linked resource. Moreover, as we already mentioned, it is becoming very common for online systems to let users express their relationships with other users of the system, even if currently most of them are not easy exportable and are meaningful only inside the community (for a longer discussion, see our survey of trust use and modelling in current real systems [12]).

By aggregating all the trust statements expressed by every user, it is possible to produce the global trust network (or social network), representing as in a snapshot the society of users with their trust relationships. An example of a simple trust network can be seen in Figure 1. As a consequence of the previously introduced properties of trust, such a network is a directed, weighted graph whose nodes are users and whose edges are trust statements.

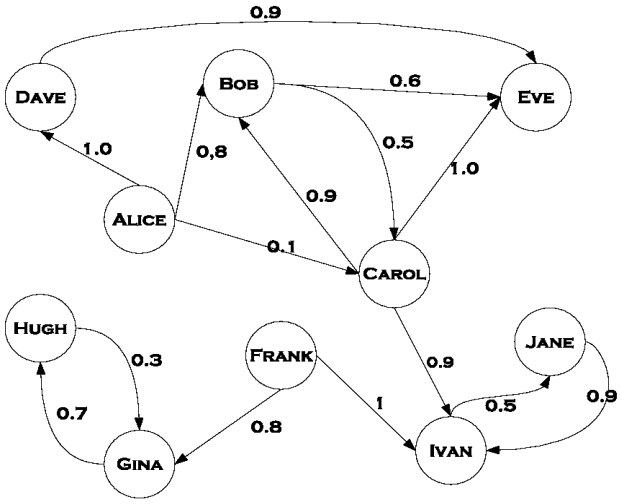

Figure 1: Trust network. Nodes are users and edges are trust statements.

3 Local vs. Global Trust Metrics

The previous section suggested that it is possible to ask users their evaluations of other users (trust statements) and to derive from the aggregated information the global trust network. However this is useful only if this trust network is then used in some way. And in fact, since in most settings, a user has a direct opinion (i.e. has issued a trust statement) only about a small portion of the other users, computational tools can be designed for predicting the trustworthiness

[4]Friend Of A Friend (FOAF) project can be found at http://www.foaf-project.com

of unknown users and, for example, providing an informed suggestion of which other users interacting with. Trust metrics [7, 11, 21, 13] have this precise aim. Given a current user A, they try to predict the trust score of the users unknown to A, by exploiting controlled trust propagation over the trust network. For example, in the social network of Figure 1, a trust metric can be used to predict the trust Alice could place in Eve. The common assumption exploited in trust metrics is that if user A trusts B at a certain level and B trusts C at another level, something can be inferred about the possible level of trust of A in C.

Trust metrics can be classified into global and local ones [13, 21]. Local trust metrics take into account the subjective opinions of the active user when predicting the trust she places in unknown users. For this reason, the trust score of a certain user can be different when predicted from the point of view of different users. Instead, global trust metrics computes a trust score that approximates how much the community as a whole trusts a specific user and is hence independent of the specific active user that is asking "How much should I trust this unknown user?". Even if there is no general agreement yet on definitions, in general, this global value is called "reputation" and "reputation systems" [19] are what we called "global trust metrics". For example, the well known auctions Web site *Ebay.com* shows for every user her "reputation" by allowing to see how many other users leaved a positive, neutral or negative feedback about that user. This "reputation" value is independent from the browsing user and hence we can say that *Ebay.com* uses a global trust metric. The formal definition of a global trust metrics is hence $T : U \to [0, 1]$ while local trust metrics are defined as $T : U \times U \to [0, 1]$.

PageRank [18], one of the algorithm behind the search engine *Google.com*, is an example of another global trust metric since the computed authority of a certain Web page is the same for every user independently of her sites preferences. In general, while local trust metrics can be more precise and tailored to the single user's peculiar views and opinions, they are also computationally more expensive, since they must be computed for every single user whereas global ones are just run once for all the community. Another interesting feature of local trust metrics is the fact they can be attack-resistant [11]: users who are considered malicious (from a certain user's point of view) are excluded from trust propagation and they don't influence the personalization of users who don't trust them explicitly. [8] shows that malicious exploitation of links is an inherent and unavoidable problem for global trust metrics. The rise of link-farms, that provide links to a certain site in order to increase its pagerank, also make evident the problem. We would like to underline once more how the majority of trust metrics proposed in literature are global, however there are few local trust metrics as well, such as the metric proposed by Golbeck [7] and the one proposed by Ziegler [21] in their PhD thesis and the metric we used in our experiments (explained in Section 4). A more complete description of trust metrics and related concepts can be found in [13].

The differences between local and global trust metrics are especially evident when considering *Controversial Users*. Controversial Users are users who are judged by other users in very diverse ways, for example they are appreciated by many users but also disliked by many other users. It is important to recall that trust statements are subjective and hence "correct" by definition. In case two users disagree about another user, it is not reasonable to make the assumption that one is a good user (expressing the "correct" trust statement)

and the other is a bad or malicious user. They simply have different but equally worthy opinions.

We will define more precisely Controversial Users in next sections, however it should be clear that, under these assumptions, global trust metrics are not suited for Controversial Users since an *average* trust score on which all the users might agree does not exist and hence cannot be found. This is an inherent limit of global trust metrics which makes them fail on Controversial Users while this theoretical limit does not exist for local trust metrics.

Summarizing, the goal of this paper is hence to provide empirical evidence that in complex societies the presence of different opinions about a specific user and so of controversial users is a normal fact. As a consequence, local trust metrics are more suited for capturing the peculiar personal opinions and views of each member of society. This is especially relevant when observing that most of the current proposals in literature are global trust metrics and are based on assumptions that there is a correct opinion about facts and those users who don't have this correct opinions are wrong and not trustworthy.

Referring to the topic of this special issue "Semantics of People and Culture", in Section 7, we compare the two extremes of culture and society induced by global and local trust metrics, respectively "tyranny of the majority" in which people who think differently from the majority are not considered and "echo chambers" in which considering only opinions of people that are very close in the social network leads to fragmentation of society and possible social conflicts.

4 A Local Trust Metric: MoleTrust

One of the important characteristics of this paper is that it is based on empirical evidence. In particular it presents a comparison of a local and global trust metric by empirically evaluating them on a large, real world dataset. These experiments are presented in Section 5 but first we need to introduce the trust metrics we used in our experiments. This section is devoted to describe the local trust metric we used. We chose to use the MoleTrust local trust metric [14]. Our choice was guided by the need of a time-efficient local trust metric, since the number of trust scores to be predicted in the experiments is very large. Another reason for choosing MoleTrust was that we wanted to analyze different levels of locality in the propagation of trust, from propagating only to the few neighbours at distance 1 in the social network (friends) to, for example, users at distance 3 (friends of friends of friends). So the presence of a tunable trust propagation horizon as input parameter of MoleTrust was important in the choice.

MoleTrust predicts the trust score of *source user* on *target user* by walking the social network starting from the source user and by propagating trust along trust edges. Intuitively the trust score of a user depends on the trust statements of other users on her (what other users think of her) weighted by the trust scores of those users who issued the trust statements. The idea is that the weight by which the opinion of a user is considered depends on how much this user is considered trustworthy. Since every trust propagation starts from a different source user, the predicted trust score of a certain user A can be different for different source users. In this sense, the predicted trust score is personalized. The pseudocode is presented in Figure 2.

Precisely, the MoleTrust trust metric can be modeled in 2 steps. Step 1 task

Step 1:
Input: $source\_user$, $trust\_network$, $trust\_propagation\_horizon$
Output: $modified\_trust\_network$
$dist = 0$; $users[dist] = source\_user$; init $modified\_trust\_network$
add node $source\_user$ to $modified\_trust\_network$
while ($dist \leq trust\_propagation\_horizon$) do
  $dist++$
  $users[dist]$=users reachable from $users[dist-1]$ and not yet visited
  add node $source\_user$ to $modified\_trust\_network$
  foreach $edge$ from $users[dist-1]$ to $users[dist]$
    add edge $edge$ to $modified\_trust\_network$

Step 2:
Input: $source\_user$, $modified\_trust\_network$, $trust\_threshold$
Output: $trust\_scores$ for users
$dist = 0$; $trust(source\_user) = 1.0$
while ($dist \leq trust\_prop\_horizon$) do
  $dist++$
  foreach u in $users[dist]$
    predecessors=users i with a trust edge in u and with $trust(i) \geq threshold$
    $trust(u) = \frac{\sum_{i \in predecessors}(trust\_edge(i,u) * trust(i))}{\sum_{i \in predecessors}(trust(i))}$

Figure 2: MoleTrust pseudocode.

is to remove cycles in the trust network and hence to transform it into a directed acyclic graph. Step 2 consists of a graph walk starting from source node with the goal of computing the trust score of visited nodes.

So the purpose of the first step is to destroy cycles in the graph. An example of cycle is the following: A trusts B as 0.6, B trusts C as 0.8, C trusts A as 0.3. The problem created by cycles is that, during the graph walk, they require visiting a node many times adjusting progressively the temporary trust value until this value converges. It is more efficient to visit each node just once and, in doing this, to compute her definitive trust value, so that time complexity is linear with the number of nodes. This is important because of the large number of trust propagations we need to performed in our experiments.

Step 1 orders users based on shortest-path distance from the source user.

There is an important input parameter of MoleTrust, the *trust propagation horizon*. This value specifies the maximum distance from the source user to which trust is propagated. For example, if the *trust propagation horizon* is set to 2, trust is propagated only up to users at distance 2 while users at distance greater than 2 are not reached and hence their trust score is not predicted (for example, user *Jane* in Figure 3). The intuition is that the reliability of the propagated trust decreases with every new trust propagation hop. Moreover, this parameter allows to reduces the number of visited users and hence to achieve shorter computational time. So step 1 modifies the social network by ordering users based on distance from the source user, keeping only users who are at a distance less or equal the trust propagation horizon. Only trust edges that

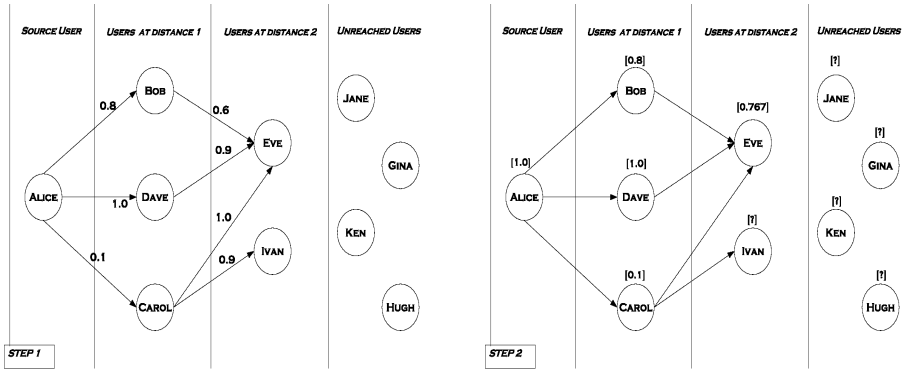

Figure 3: Graphical representation of steps 1 and 2, with *trust horizon* 2 and *source user* Alice.

goes from users at distance k to users at distance $k+1$ are kept. Left part of Figure 3 shows how the trust network of Figure 1 is modified after step 1, when considering Alice as source user. We are aware this step removes trust statements that can be informative but, because of the requirement of time-efficiency, we believe this is an acceptable trade off. After step 1, the modified social network is a reduced directed acyclic graph, with trust flowing away from the source user and never flowing back.

Step 2 is a simple graph walk over the modified social network, starting from source user. The initial trust score of the source user is set to 1. MoleTrust computes first the trust score of all the users at distance 1 (Bob, Carol and Dave in Figure 3), who are users on which the source user, Alice, directly provided a trust statement. MoleTrust then moves to users at distance 2 (Eve and Ivan) and so on. Note that, in virtue of step 1, the trust score of a user at distance k only depends on trust scores of users at distance $k-1$, that are already computed and definitive. In this way, there is no need to walk over one user more than once, for example until her predicted trust score converges.

For predicting the trust score of a user, MoleTrust analyzes all the incoming trust edges (representing the trust statements remaining from step 1) and accepts only the ones coming from users with a predicted trust score greater or equal than a certain threshold. Let us analyze the example of Figure 3. When predicting the trust score of user Eve as seen by Alice, MoleTrust only accepts the opinions of Bob and Dave about Eve and does not accept the trust statement issued by Carol. The reason is that the predicted trust score of Carol is 0.1, less than the threshold (0.6 in this example).

The predicted trust score of a user is the average of all the accepted incoming trust edge values (representing the subjective judgments), weighted by the trust score of the user who has issued the trust statement. The formula is hence

$$trust(u) = \frac{\sum_{i \in predecessors}(trust(i) * trust\_edge(i,u))}{\sum_{i \in predecessors}(trust(i))}$$

For example, in Figure 3, the trust score of Eve only depends on trust statements (trust edges) issued by Bob and Dave and their trust statements are weighted according to their predicted trust score (respectively, 0.8 and 1.0). The predicted trust score of Eve is $\frac{0.8*0.6+1.0*0.9}{0.8+1.0} = 0.767$.

The reason for accepting only trust statements from users whose predicted trust is greater or equal than a certain threshold is the following. Users who have a predicted trust score below the threshold are users that MoleTrust predicted as untrustworthy (from the point of view of current source user). So their opinions should not influence the predictions about the trust score of other users and the best possible action is simply to not consider their trust statements. In fact, this precaution avoids situations in which the trust score of an unknown user depends only on statements issued by untrustworthy users. In the example of Figure 3, Carol is the only one to have expressed a trust statement on Ivan. If MoleTrust was to consider the trust statement expressed by user Carol, it would predict a trust score of 1.0 for Ivan but this predicted value would have been derived only from the opinion of a user, Carol, with very low predicted trust and hence from the opinion of an untrustworthy user. Moreover, if Carol knows or guesses its predicted trust score from the point of view of user Alice, she has incentives into providing trust statements with the unique goal of influencing the trust scores predicted by MoleTrust on behalf of user Alice. For example, user Carol could express a trust statement in all the other users with a value of 0.01, in order to nuke the reputation of all the unknown users, or could boost just the reputation of chosen users. In short, an untrustworthy user would be able to influence the predicted trust score of other users, a situation a trust metric should be able to prevent.

Since the first step retained only users inside the trust propagation horizon, MoleTrust trust metric is able to compute trust values only in those users, that are users reachable from the the source user in less steps than the trust propagation horizon. This is the coverage of the trust metric.

As already stated, there are other recently proposed local trust metrics [7, 21]. In particular, the metric proposed by Golbeck, TidalTrust [7], is similar to MoleTrust [14], since they both work in a breadth first search fashion. One difference is that in MoleTrust the trust propagation horizon is an input parameter and hence it is tunable so that it is possible to test how different levels of locality in the trust propagation affect accuracy and coverage of the trust metric (see Section 6). For instance, setting a trust propagation horizon at 1 means that users are just influenced by their direct friends (users at distance 1) and this makes very strong echo chambers, while larger values allows more opinions to be taken into account. On the other hand, in TidalTrust, the maximum depth is inferred from the network and corresponds to the length of the first path found from the source to the sink. The threshold is also computed in a different way in the two metrics. Another very interesting local trust metric is Ziegler's AppleSeed [21] which is based upon spreading activation models, a concept borrowed from cognitive psychology. Some clever adaptations are made in order to handle distrust and sinks such as trust decay and normalization.

In this section, we presented MoleTrust, the local trust metric we used in our experiments. The global trust metric we chose for comparison instead is introduced in the next section where the dataset used in our experiments is described since the choice of the global trust metric was guided by the characteristics of the available data.

5 Experiments about controversiality on Epinions.com

We conducted the experiments on data of the community of users of the popular Web site *Epinions.com*. *Epinions.com* is a Web site where users can write reviews about products and assign them a rating. *Epinions.com* also allows the users to express their *Web of Trust*, i.e. "reviewers whose reviews and ratings they have consistently found to be valuable" and their *Block list*, i.e. a list of authors whose reviews they find consistently offensive, inaccurate, or in general not valuable. Inserting a user in the Web of Trust equals to issuing a trust statement in her ($T(A, B) = 1$) while inserting her in the Block List equals to issuing a distrust statement in her ($T(A, B) = 0$). Intermediate values such as 0.7 are not expressible on *Epinions.com* and hence not available in our experiments.

The Epinions dataset we used contains $\sim$132,000 users, who issued $\sim$841,000 trust statements. Of these, $\sim$717,000 were positive trust statements with value 1 (representing in fact trust) and $\sim$124,000 were negative trust statements with value 0 (representing in fact distrusts). $\sim$85,000 users received at least one trust statement. Some statistics on distributions of data are presented in [9].

Based on the actual characteristics of available data, particularly the fact that statements values are just 1 (total trust) and 0 (total distrust) and not any real in the interval $[0, 1]$, we now define some quantities. The *Controversiality Level* of a user is the number of users who disagree with the majority in issuing a statement about that user. For example, a user who received 21 distrust statements and 14 trust statements has a Controversiality Level of 14. Formally,

$$controversiality\_level(u) = min(\#received\_trust(u), \#received\_distrust(u))$$

A user who has a Controversiality Level of x is called *x-controversial*. *0-controversial users* received only trust or distrust statements and they are non controversial. We define a user who has a Controversiality Level not less than x as a *at least x-controversial* user.

As one might expect, most of the users are non controversial, in the sense that all the users judging them share the same opinion. This happens especially in societies where there is an agreed concept of what is good and what is bad. For instance this is probably more likely to happen on *Ebay.com* than on *Epinions.com*. On Ebay almost all the users agree about the fact that a good buyer is one that pays timely and a good seller is one that sends a product similar to the one described and does it quickly, while on Epinions the central contributed information is subjective opinions about items and hence a user might be considered a good user by someone and a terrible one by someone else based on their different inherently subjective tastes.

In the Epinions dataset we used, out of the 84,601 users who received at least one statement, 67,511 are 0-controversial, 17,090 (more than 20%) are at least 1-controversial (at least one user disagrees with the others), 1,247 are at least 10-controversial, 144 are at least 40-controversial and one user is 212-controversial. Figure 5 shows the number of users with a certain level of controversiality.

However, a user with 100 received trusts and 5 received distrust and another user with 5 received trusts and 5 received distrusts have the same controversiality level, even if the first one can be seen as much less controversial. For this reason

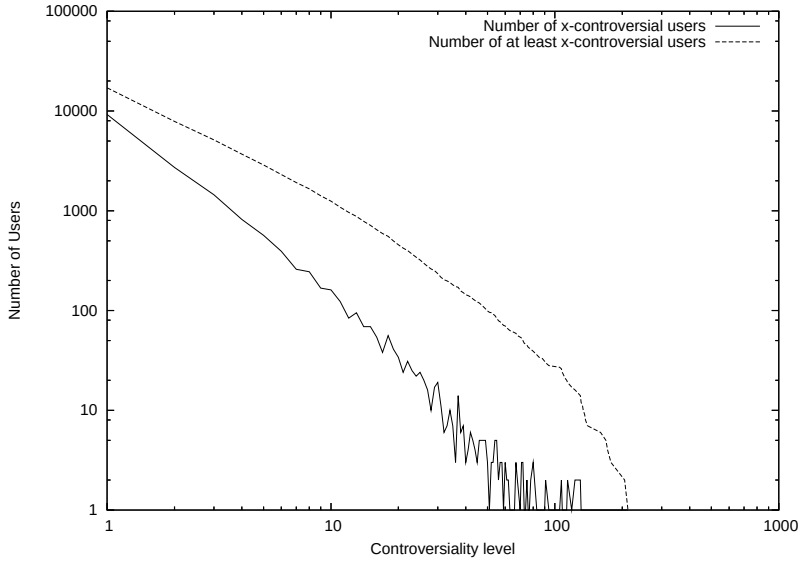

Figure 4: Number of Epinions users with a certain controversiality level.

we define another quantity, *Controversiality Percentage* as

$$controversiality\_percentage(u) = \frac{\#received\_trust(u) - \#received\_distrust(u)}{\#received\_trust(u) + \#received\_distrust(u)}$$

A user with 1 as controversiality percentage is trusted by all her judgers, while a user with -1 is distrusted by all her judgers. Users at these two extremes are non-controversial since all the other users agree on their opinions about them. Viceversa, a user with 0 as controversiality percentage is highly controversial since other users split into two same-sized opinions group about this user, i.e. $n/2$ users trust her and $n/2$ users distrust her[5].

We now describe the global trust metric we compared in our experiments with the local trust metric, MoleTrust, we already introduced in Section 4. As already defined, a trust metric is an algorithm that, given a source user and a target user, returns the trust score that the source user could place in the target user. The choice for the global trust metric was guided by the available data, mainly by the fact that statements are binary (1 and 0) and not continuous. So we chose to use a very simple metric that is similar to the one used by the online auctions site *Ebay.com*. In order to predict the trust score of one user, the global trust metric (in the following called *ebay*) simply computes the fraction of received trust statements over all the received statements. Formally,

$$trust_{ebay}(u) = \frac{\#received\_trust(u)}{\#received\_trust(u) + \#received\_distrust(u)}$$

The trust score of a user with 0 received statements is not predictable.

[5]Strictly speaking, this is not a percentage since it is defined in [-1,1]. However it is useful to keep separated mainly trusted users ($percentage > 0$) and mainly distrusted users ($percentage < 0$) in order to evaluate the different trust metrics separately on them. An actual percentage of controversiality would be $1 - |controversiality\_percentage|$.

There have been proposals for more advanced global trust metrics, for example some adaptations of PageRank [18] able to handle also continuous trust scores in input. We decided to use the simple global trust metric we explained before since in this paper the point we want to make is that every global trust metric (either simple or more complex ones) have an intrinsic limit on Controversial Users. In fact, since it is not reasonable to assume that some users are malicious or bad, every trust statement has the same weight and importance and hence, if a user received x trust and x distrust (controversiality percentage equal to 0), no global trust metric is able to achieve an error that is smaller than 0.5, notwithstanding the complexity of the global trust metric. This is an intrinsic limit of every global trust metric. We discuss this point by commenting the real data in the following sections (see Figure 8). We preferred to use a simple global trust metric also because the input data are just binary, either total trust (1) or total distrust (0). Actually note that the theoretically best global trust metric is the one that always returns what the majority thinks, so that it is able to incur in the smallest possible error for the largest possible number of predicted trust statements. We decided nevertheless to use the ebay-like global trust metric we described before since it is intuitively more meaningful to show to the user the average of the ratings received by a user and not total trust or distrust depending what the majority expressed. The point of the paper anyway is not to show that a certain global trust metric performs worst than a certain local trust metric, but to discuss the intrinsic limits in accuracy of every global trust metric and to show how local trust metrics are not affected by these and could, at least theoretically, achieve total accuracy, i.e. an error equal to 0.

As Local Trust Metric, we choose MoleTrust, the metric introduced in Section 4. We run different tests with different trust propagation horizons, precisely 2, 3 and 4. The trust threshold was set to 0.6. It should be noted that, even if the input trust statement values are just 1 and 0, both the local and global metric predict trust scores in the interval $[0, 1]$.

In order to test the performances of the different Trust Metrics, we use a standard evaluation technique in machine learning: leave-one-out. Taken one trust statement from user A to user B, we remove it from the trust network and try then to predict it using the local trust metric. We then compare the predicted trust score against the original trust statement and compute the prediction error. For the global trust metric, we compare the predicted global trust score of B against the statement issued by A on B. Note that in the Epinions dataset there are $\sim 841,000$ trust statements and the entire leave-one-out process has to be evaluated for every single trust statement. This is the reason behind the choice of time-efficient local trust metric. Two measures are derived from this evaluation technique: accuracy and coverage. Accuracy represents the error produced when predicting a score. We use Mean Absolute Error that consists in computing the absolute value of the difference between the real score and the predicted one and averaging the errors over all the predictions. Coverage refers to the ability of the algorithms to provide a prediction. In this case we compute the percentage of trust statements that are predictable.

6 Results

Figure 5 and Figure 6 show the prediction errors for the two compared trust metrics, *moletrust2*[6] and *ebay*, respectively. The x axis represents the *controversiality level* of users, as defined in the previous section. The plotted value is the Mean Absolute Error over the statements received by users who are at least x-controversial. We distinguish the error over only trust statements (bottom line), over only distrust statements (top line) and over all the statements (central line). The graphs show how predicting distrust statements is more difficult than predicting trust statements: the error on distrusts is greater than that on trusts in both cases. However while for ebay the error on distrust statements is higher than 0.6, for moletrust2 it is around 0.4.

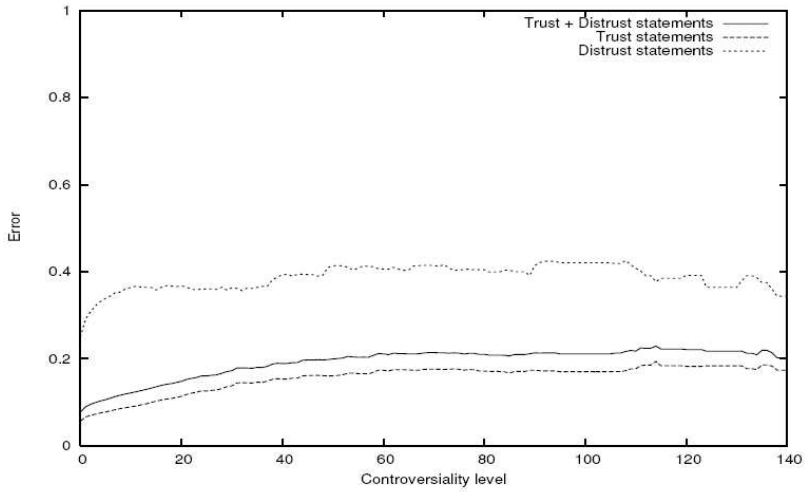

Figure 5: Moletrust2 prediction error on different kinds of statements (trust, distrust and both) for users who are *at least x-controversial*.

On the other hand, the error over trust statements is very similar for the two different techniques. This is because the number of trust statements is much larger than the number of distrust statements: $\sim 717,000$ against $\sim 124,000$. This fact does not allow to clearly distinguish how much trust metrics are effective: a technique predicting almost always a trust score close to 1 (as ebay does) produces a very small error over trust statements.

So we concentrate on *controversiality percentage*. Figure 7 and Figure 8 show the accuracy of moletrust2 and ebay over buckets grouping users based on their controversiality percentage. The mean absolute error close to the borders (1 and -1 that represents users who are not controversial since they are mainly trusted or distrusted) is very small and very similar for both algorithms, this indicates that it is easy to spot out highly trusted or highly distrusted users and that there is no strong requirement for particularly clever trust metrics. However, if we concentrate our attention on controversial users in Figure 7 and

[6]Moletrust2 refers to the local trust metric MoleTrust run with trust propagation horizon set to 2. The results obtained with different trust propagation horizon are very similar and hence we just report the performances of moletrust2.

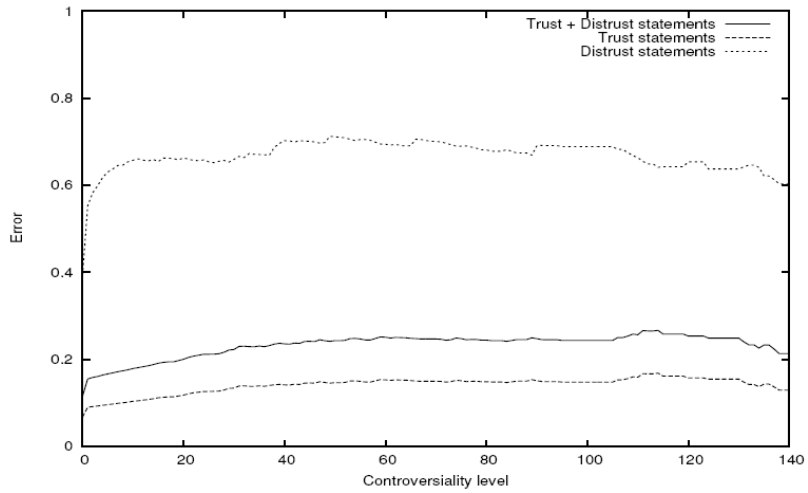

Figure 6: Ebay prediction error on different kinds of statements (trust, distrust and both) for users who are *at least x-controversial*.

Figure 8, we see a different story. Controversial users fall into buckets close to the centre of x axis, for example a user trusted by k users and distrusted by k other users would fall into the controversiality percentage bucket of value 0. For controversial users, moletrust2 is able to significantly reduce the prediction error. As expected, the error produced by ebay on users with controversiality percentage 0 is 0.5 since these users received k trusts and k distrusts and the metric predicts 0.5 as the global trust score for them so that, when compared to the real trust score (either 1 for trust or 0 for distrust), it encounters an error of 0.5 in every single case. This is an inherent limit for global trust metrics, i.e. on users with controversiality percentage of 0 they cannot achieve an error smaller than 0.5 (see also Figure 8).

It should be clearly noted, however, that the majority of users fall into the buckets near the borders. In fact, while the total number of statements is $\sim$841000. $\sim 440,000$ (more than 50%) of them go into (i.e. are issued about) users who are in the 1 bucket and $\sim 206,000$ for the 0.9 bucket. These users are either totally trusted or almost totally trusted and hence not or little controversial. $\sim 41,000$ statements go into the -1 bucket, that contains users who received only distrust statements and are not controversial as well. Only $1,972$ statements go into users falling into the 0 bucket and $1,013$ into the -0.1 bucket, the least populated. Once more, this data shows that most of the users are not (or little) controversial. However, the fraction of controversial users is not negligible and it can be argued that these users are the ones on which a trust metric is more required and its performances matter the most. It is also interesting to note that users who received dozens of statements (i.e. popular users) tend to have higher controversial percentage that little known users who received just 2 or 3 statements: this means that the more a user is popular, the more likely is that someone will disagree with the majority. Concluding, we can say that Figure 7 and Figure 8 provide empirical evidence of how, for controversial users of a real world society, the chosen local trust metric significantly reduces error

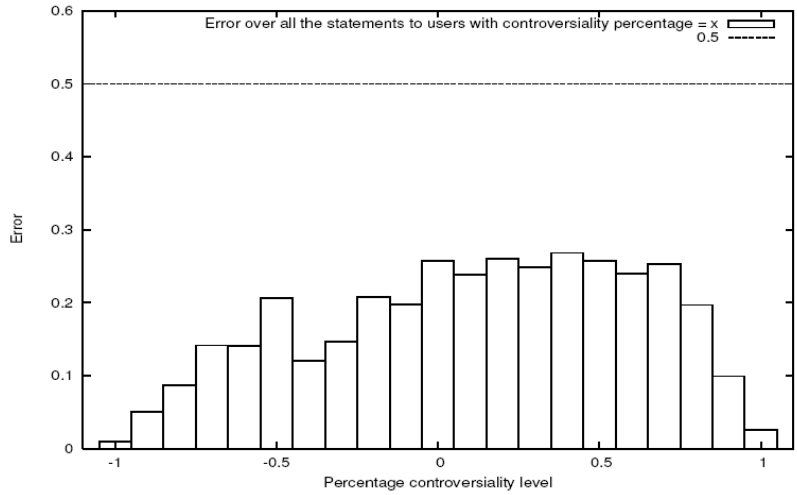

Figure 7: Moletrust2 prediction error for users grouped by controversiality percentage. Users at the borders are not controversial while users at the centre are highly controversial.

when compared to a global trust metric.

It is interesting to separate the error generated in predicting trust and distrust statements and analyze them independently (see Figure 9). As expected, for users close to -1 bucket (distrusted by almost all their judgers), it is easy to correctly predict (the many) distrust statements and it is hard to predict (the few) trust statements. The opposite is true for users with controversiality percentage close to 1. If we compare the error over distrust statements for moletrust2 and ebay, we observe how much the local metric is able to reduce the error. A similar observation can be made for trust statements as well. It is important to note that the difference in error between the local and global trust metric is larger on predictions for distrust statements than for trust metrics and this fact once more confirms that predicting distrusted users is harder and a local trust metric is particularly needed and useful in order to identify these rare but crucial situations. However, it is not clear why the error for moletrust2 on distrust statements is smaller than the one for trust statements also for users who are mostly trusted by others, such as users in the 0.4 bucket. Since most of the users (more than 75%) fall into bucket 1 and 0.9, these results are not in contrast with the previous ones showing that the error is greater on distrust than trust statements.

Another important evaluation quantity, besides accuracy, is coverage, i.e. the percentage of statements that are predictable by the algorithms. The coverage of a global trust metric is very close to 1, since a single statements received by a user is enough for predicting a trust score for that user. Instead, MoleTrust, the chosen local trust metric is able to predict a trust score only if there is at least one trust path from source user to target user, shorter than the trust propagation horizon. The coverage of moletrust2 is around 0.8 for distrust statements, 0.88 for trust statements and 0.86 on both. These percentages are stable across all the controversiality levels.

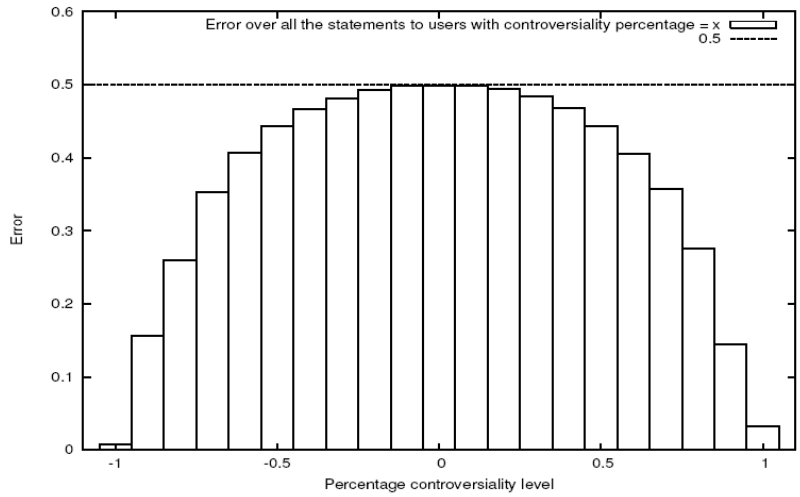

Figure 8: Ebay prediction error for users grouped by controversiality percentage.

Running MoleTrust with different trust propagation horizons (2, 3 and 4) produce very similar results. The accuracy is very similar for the 3 configurations across all the levels and percentages of controversiality. The coverage is, as expected, higher for moletrust4 that for moletrust2 since it is possible to propagate further the trust and hence to reach more users, who might have expressed a statement about the target user. However, a larger trust propagation horizon also means greater computational time and this can be an issue if time requirements are important. In this paper we concentrate more on the differences between one local trust metric (moletrust2) and one global trust metric (ebay) and so we don't analyze the differences produced on MoleTrust performances by different trust propagation horizon values. As already stated, these differences are not very significant, at least with respect to accuracy.

7 Discussion of Results

Since MoleTrust is a local trust metric, it provides personalized trust scores that depend on the active user. For this reason, it is expected to work better than a global one, especially when predicting the trust score of controversial users that, by definition, don't have a *global* trustworthiness value on which all the other peers can agree. It is worth recalling that on *Epinions.com community*, controversial users are a non negligible fraction: 20% of the users who received a statement are at least 1-controversial.

Figure 8 shows that the error of a global metric for highly controversial users is, as expected, 0.5. As we already argued, this is an intrinsic limit of every global trust metric, from simpler to more complex one. This is so because every trust statement has the same weight and importance (it is not reasonable to label some users as malicious and hence to not consider their trust statements). As a consequence, if we concentrate on predicting the trust scores for users with a controversiality percentage equal to 0, every global trust metric incur in an

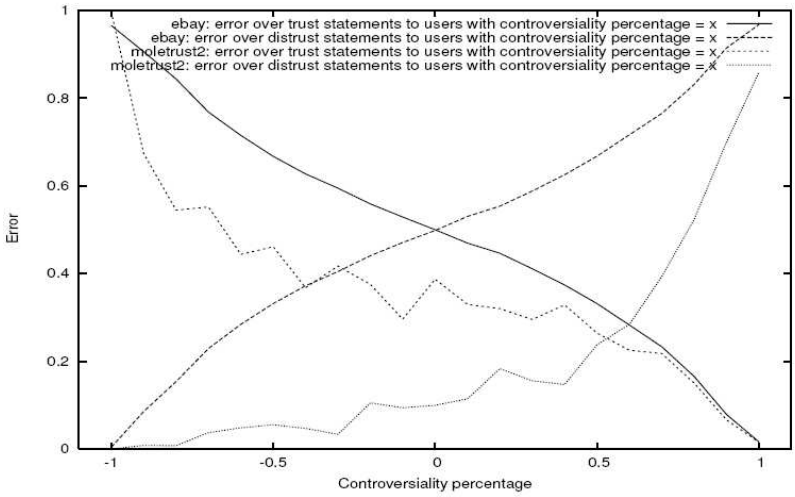

Figure 9: Prediction error of moletrust2 and ebay, considered separately for trust and distrust statements.

error that is at least 0.5. It is not possible to do any better than this.

Instead, there is no theoretical limit for local trust metrics that can, by definition, returns a different and personalized trust score in the target user for every single source user. However MoleTrust does not perform as well as one could have desired, reducing the error close to 0 for every kind of users. One reason could be that the *Epinions.com* dataset we used is too sparse (on average a user has expressed statements on very few users) and trust propagation often cannot reach a user from many different trust paths that would allow more accurate predictions in all the cases. It remains an open point to verify if, on a more connected community, a local trust metric can decrease the error close to 0 also for controversial users.

The experiments clearly show that correctly predicting a distrust statement is harder than predicting a trust statement. However, it is very important to inform the user about other users who should not be trusted, such as a malicious user that is trying to fool the active user for her personal benefit: for example by rating highly her own book on an opinions site in order to have it recommended to every other user. Correctly predicting distrust is hence an important research challenge [9]. Since detecting attacks [11] is one of the most important tasks for trust metrics, we think this fact asks for more research into designing metrics particularly tailored for predicting distrust. This should be carried on without assuming that the society is formed by "good" users and "bad" users and that the goal is to spot the "bad" users but keeping as granted that users have a subjective notion of who to trust or not.

Speaking of complexity, a global trust metric can be run once for the entire community, while a local one must be run once for every single user, in order to predict the trust scores of other users from her personal point of view. This fact makes local trust metrics difficult to integrate into a centralized service such as *Google.com* because of computational time and storage problems. The more reasonable setting for a local trust metric is the one in which every user runs it

from her personal point of view, possibly in her mobile device or in her browser.

Another weak point of local trust metrics is the reduced coverage: while global trust metrics coverage is usually close to 1, this is not always the case for local ones.

We have seen how on the *Epinions.com* community, moletrust2 is able to reach on average a good coverage (almost 0.8). A possible improvement would be to integrate the two techniques, for example, by using a global metric when the local one is not able to predict a trust score but also making this fact known to the user using the system.

Eventually, in non controversial domains, global metrics can be more suited because they guarantee greater coverage, smaller computational time with similar accuracy. For example, on *Ebay.com*, the notion of good seller is a shared concept agreed by most of the users. Maybe some users will give more importance in timeliness in sending the goods while others will give more importance in correspondence between photo and the shipped good but what makes a good seller is a little ambiguous and largely agreed concept. When we move into more subjective domains, such as evaluating music tracks or movies (or even discussing political ideas), it is reasonable to accept significant disagreement between users. In this contexts, a local trust metric can be more effective and appropriate.

Actually, the basic assumptions behind a choice for a global or a local trust metric influence significantly the kind of society they can induce. A global trust metric assumes that there are globally agreed good peers and that peers who think different from the average are malicious. This assumption encourages herd behavior and penalizes creative thinkers, black sheep and original, unexpected opinions. What we would like to underline is that there is a "tyranny of the majority" risk, a term coined in 1835 by Alexis de Tocqueville in his book, Democracy in America [4]. The 19th century philosopher John Stuart Mill in his book "On Liberty" [16] also analyzes this concept, with respect to social conformity. The term "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, so that on a particular issue people will align themselves either for or against this issue and the side of greatest volume prevails. 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" [16][7]. 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 simulations in research papers. Moreover, on digital systems, such as online communities on the Web, automatic personalization is possible and so there is no need to make this assumption and try to force all the users to behave and think in the same way or be considered "unfair" or "wrong".

However there is a risk on the opposite extreme as well that is caused by

[7]This definition is extracted from Wikipedia (http://en.wikipedia.org/wiki/On_Liberty) which interestingly tries to find a balance between what different people think about every single topic, by asking to the contributors to adopt a neutral point of view (NPOV). This seems to work well enough for now, possibly also because the people who self-elect for editing Wikipedia articles largely share a similar "culture". However the frequent "edit wars" (http://en.wikipedia.org/wiki/Wikipedia:Edit_war), particularly evident on highly sensitive and controversial topics, show that it is and will be hard to keep this global and theoretically unbiased point of view.

emphasizing too much locality in trust propagation by a local trust metric. This means considering, for example, only opinions of directly trusted users (friends) stopping the propagation at distance 1. This risk is called "echo chamber" or "daily me" [20]. Sunstein, in the book Republic.com [20], notes how "technology has greatly increased people's ability to "filter" what they want to read, see, and hear". He warns how in this way everyone has the ability to just listen and watch what 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 in micro groups who tend to extremize their views, develop their own culture and not being able to communicate with people outside their group anymore. He argues that, in order to avoid these risks, "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" [20].

Global trust metrics, considering all the opinions expressed by members of the society, have the advantage of being able to reach a large coverage, i.e. being able to make an informed decision on almost every topic or other user. However they tend to propose to everyone the same views, for example, in the case of Epinions, adopting a global trust metric would result in all the users seeing the same unpersonalized Top 10 lists. On the other hand, local trust metrics promised to deliver personalized results but have some disadvantages: from a purely opportunistic point of view, they suffer of reduced coverage since they choose to consider not all the opinions but just the opinions of a portion of the society (the friends of the active user) and from a collective point of view they risk to fragment the society in a bunch of non communicating clusters. In the case of Epinions it is evident that this risk is not an important one since people use Epinions for getting informations about commercial products and hence a very local trust metric able to deliver highly personalized recommendations is the best choice. But it would be surely different when pondering about a community in which political ideas were discussed.

It is easy to foresee that in the short future more and more people will increasingly rely on opinions formed based on facts collected through online systems such as opinions and reviews sites, mailing list, fora and bulletin boards, matchmaking sites of every kind, aggregators and in general social software sites [12]. The assumptions on which these systems are constructed have and will have a fundamental impact on the kinds of societies and cultures they will shape. Finding the correct balance between the two described extremes, "tyranny of the majority" and "echo chambers", is surely not an easy task but something that must be taken into account both for systems designers and researchers. We hope this paper can help a bit with regard to these important topics, by presenting some empirical evidence about what is reasonable to expect from a real world community and their social relationships.

8 Conclusions

In this paper we compared local and global trust metrics by analyzing the differences in accuracy and coverage of two chosen representative examples of these

trust metrics in the task of predicting trust scores of unknown users. We focused our attention particularly on Controversial Users, defined as users that are judged in very different ways by other users and we have shown that Controversial Users are a non negligible portion of the users on the large *Epinions.com* community. We have argued how there is an intrinsic limit for global trust metrics on Controversial Users and how local trust metrics are instead more suited in contexts in which opinions are subjective and especially for Controversial Users. The empirical results demonstrates that the chosen local trust metric is able to significantly reduce the prediction error for Controversial Users, while retaining a good coverage. We believe the evidence presented in this paper is relevant in the discussion about the different kinds of societies induced by different trust metrics, tyranny of the majority for the global trust metrics and echo chambers for the local trust metrics. We hence concluded by discussing the risks represented by both these extremes.

9 ACKNOWLEDGMENTS

The authors would like to thank *Epinions.com* for kindly making available the data for this study. We would in particular like to thank also R. Guha, Ravi Kumar, Prabhakar Raghavan and Andrew Tomkins for sharing these data.

References

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

[2] F. Cornelli, E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. Implementing a reputation-aware gnutella servent. In *International Workshop on Peer-to-Peer Computing*, May 2002.

[3] P Dasgupta. Trust as a commodity. In Diego Gambetta, editor, *Trust: making and breaking cooperative relations*. Basil Blackwell, Oxford, 1990.

[4] Alexis de Tocqueville. *Democracy in America*. Doubleday, New York, 1840. The 1966 translation by George Lawrence.

[5] John Dunn. The Concept of Trust in the Politics of John Locke, 1984. In, Philosophy in History, R. Rorty, J. B. Schneewind and Q. Skinner (eds.). Cambridge University Press, Cambridge.

[6] Francis Fukuyama. *Trust: the Social Virtues and the Creation of Prosperity*. Free Press Paperbacks, New York, 1995.

[7] Jennifer Golbeck. *Computing and Applying Trust in Web-based Social Networks*. PhD thesis, University of Maryland, 2005.

[8] Marco Gori and Ian Witten. The bubble of web visibility. *Commun. ACM*, 48(3):115–117, 2005.

[9] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *WWW '04: Proc. of the 13th int. conf. on World Wide Web*, pages 403–412. ACM Press, 2004.

[10] T. Haveliwala, S. Kamvar, and G. Jeh. An analytical comparison of approaches to personalizing pagerank, 2003. Technical report, Stanford University, 2003.

[11] R. Levien. *Advogato Trust Metric*. PhD thesis, UC Berkeley, USA, 2003.

[12] P. Massa. A survey of trust use and modeling in current real systems, 2006. Under publication in "Trust in E-Services: Technologies, Practices and Challenges", Idea Group, Inc.

[13] P. Massa and P. Avesani. Trust-aware collaborative filtering for recommender systems. In *Proc. of Federated Int. Conference On The Move to Meaningful Internet: CoopIS, DOA, ODBASE*, 2004.

[14] P. Massa, P. Avesani, and R. Tiella. A Trust-enhanced Recommender System application: Moleskiing. In *Proceedings of ACM SAC TRECK Track*, 2005.

[15] Paolo Massa and Conor Hayes. Page-rerank: using trusted links to re-rank authority. In *Proc. of the Web Intelligence Conference*, 2005.

[16] John Stuart Mill. *On Liberty*. History of Economic Thought Books. McMaster University Archive for the History of Economic Thought, 1859.

[17] L. Mui. *Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks*. PhD thesis, Massachusetts Institute of Technology, 20 December 2002.

[18] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford, USA, 1998.

[19] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation Systems. *Communication of the ACM*, 43(12), December 2000.

[20] Cass Sunstein. *Republic.com*. Princeton University Press, 1999.

[21] Cai-Nicolas Ziegler. *Towards Decentralized Recommender Systems*. PhD thesis, Albert-Ludwigs-Universität Freiburg, Freiburg i.Br., Germany, June 2005.