14

Trust It Forward:

Tyranny of the Majority or Echo Chambers?

Paolo Massa

If reputation systems weight all perspectives similarly, they may devolve into simple majority rule. But if they give each user reputation scores that take only other similar users' opinions into account, they run the risk of becoming "echo chambers" in which like-minded people reinforce each others' views without being open to outside perspectives. Massa discusses design choices and trust metrics that may help balance these two extremes and the broader implications for our future societies.

Trust Is a Key Element for Society

Trust is a key element for society. Without trust, society could not exist (Fukuyama 1995). We rely on trust when we walk out in the street, when we talk to somebody, when we buy something—in our every action.

Even the very act of reading this contribution is based on trust: you, the reader, have some degree of trust in the editors of the book and in the authors of the contributions and their ability to collectively provide an insightful and interesting book.

The *Oxford English Dictionary* (Trust, 1990) defines trust as "the firm belief in the reliability, truth or ability of someone or something." In fact, the concept of trust is not new and has received much attention from scholars for centuries (Locke 1680). As trust is a multifaceted concept, thinkers from disciplines as diverse as economics, philosophy, psychology, sociology, anthropology, and political science have attempted to formalize it and to understand the importance trust has for our societies (De Tocqueville 1840; Putnam 1995; Stuart Mill 1859; Sunstein 1999).

Influential research has analyzed how trust correlates with basic features of communities and nations. The World Values Survey project is an ongoing research effort that is trying to assess the state of social, cultural, and moral values in different countries of the world. Every year, in each country, at least 1,000 citizens answer about 250 questions during face-to-face interviews. Some questions are related to general trust, such as, "Generally speaking, would you say that most people can be trusted?" These data give a detailed picture of values across time in the world. For example, in 1999, 61 percent of people in Norway said others in their country are trustworthy (the highest percentage in the survey), whereas only 6.6 percent of people in Brazil said others are trustworthy (the smallest). This finding suggests that different societies have a different default expectation of the trustworthiness of others.

These data have been used to demonstrate correlations between general trust and many features of societies and countries. For example, trust has been shown to be positively correlated with economic growth, well-being, and happiness and negatively correlated with crime and corruption. Greater ethnic diversity was found to be correlated with less trust by one author (Putnam 1995), though this may depend on the society in question. The influential book *Bowling Alone* (Putnam 1995) also comments on the decline of social capital and general trust in America in the past twenty-five years as something that must be fixed from a public policy point of view. One study has shown the default level of trust to increase with education (Toshio 2001). Trust can be considered and studied as a basic constituent of a human society, and it correlates with what we might have in mind as a healthy society (Fukuyama 1995).

Trust 2.0: Society Moves Online

Social network sites (SNSs) are web-based services that allow individuals to construct a public or semipublic profile. Users can generally upload their picture, add a textual description of themselves and their interests, and have this information shown in their user profile, usually with an overview of recent activity performed by the user. Users can also be "friends" with other users or express other social relationships (boyd and Ellison 2007).

Examples of this new paradigm range from entertainment-oriented sites such as Facebook or Orkut.com to e-marketplaces such as eBay. They range from opinion-sharing sites such as Epinions or Essembly.com to activity-sharing sites such as Flickr.com or Delicious.com and to business networking sites such as LinkedIn (Massa 2006).

In all these settings, the social relationships and connections users can express are different and have different meanings. As boyd and Ellison (2007) note, "The nature and nomenclature of these connections may vary from site to site." Here we want to provide a unifying view about these online social relationships: they can be considered as expressions of trust, that is, as trust statements.

We use an operational definition of trust from Massa (2006). Trust is defined as "the explicit opinion expressed by a user about a target user, regarding the perceived quality of a certain characteristic of the latter" (Massa, 2006, p. 55). The term "trust statement" is used with the same meaning. For example, in some systems quality refers to the ability to provide reliable and interesting product reviews (as in Epinions). In other systems, it refers to the ability to be a good friend for the user (as in Facebook). In yet others, it is the ability to find interesting new websites (as in Delicious.com). This is called the *trust context*, and it is the characteristic of the target user that is 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 target user. For example, the subjective trust expressed by Alice of Bob about his professional skills (the trust context of LinkedIn) may not be correlated with the trust expressed by Alice of Bob about his quality of being an honest seller online (the trust context of eBay). In this chapter, we interpret the social relationships expressed on SNSs as trust statements. It should be noted that, at present, in only a few SNSs is there explicit mention to the term "trust"—for example, on Epinions the list of favorite users is called the "Web of Trust" (Epinions, n.d.).

Bloggers commonly include a list of blogs they read in their so-called blogroll, which often represents their trust relationships. The web itself can be considered as a giant trust network. Considering links between web pages as votes or trust statements was precisely the clever intuition exploited by algorithms such as Google PageRank (Page et al. 1998) for inferring authority of web pages. Although links, citations, and other mentions can be negative as well as positive, the basic intuition works well enough in practice to be useful.

In the future, even more of our interrelationships may leave electronic trust trails, especially if current trends continue toward most people being always connected with powerful mobile devices.

Representing Trust

Trust is a relationship between two users. A formal example of a trust statement is "I, Alice, trust Bob as 0.8 in the range [0,1] about the trust context of pleasant violin playing."

Reputation is closely related to trust. In general, reputation summarizes what the community as a whole thinks about a certain user in a certain trust context. The *Oxford English Dictionary* puts it this way: "reputation is what is generally said or believed about a person's or thing's character or standing" (Reputation, 1990).

For computational purposes, both trust and reputation can be represented as normalized values in the range [0,1], with 0 as the minimum (no trust or no reputation) and 1 as the maximum (total trust or total reputation). Formally, we have $\mathit{trust}(A,B)\rightarrow[0,1]$ *and* $\mathit{reputation}(A)\rightarrow[0,1]$ meaning that trust is personalized and reputation is not. (We discuss later the key difference between trust and reputation, especially as metaphors for key design choices of SNSs.)

Frequently, SNSs don't allow users to express weights on trust relationships, mainly in order to keep the system simple for the user to understand. An exception is represented by eBay, at which it is possible to leave positive, negative and neutral feedback—this can be mapped, for example, to the values (1, 0, 0.5). Another exception is Advogato.org, an SNS for open source developers, at which users can express their trust relationships in other programmers using four textual labels—Master, Apprentice, Journeyer, and Observer—which can be mapped, for example, to the values 1.0, 0.8, 0.6, and 0.4 (Massa et al. 2008).

Asking humans to represent trust explicitly and to make trust statements visible is already changing how we humans think about and rely on trust. Zak postulates that in our society "the decision to trust another human being is largely unconscious and utilizes the 'social brain'" (2006, p. 23). Making trust explicit seems likely to change how we—as a society—use it.

Another challenge that representing trust introduces is the disproportionately large fraction of positive trust statements with respect to negative ones that is common to find on SNSs. On eBay for example, more than

99 percent of the feedback has been positive (Massa 2006). Moreover trust statements are often made public (for greater accountability and less forgeability) and this means that, especially in sensitive contexts such as job relationships, if you don't consider your boss trustworthy, you may need to have courage and a strong reason to explicitly and publicly express *trust(Me,MyBoss)=0.1*, which not many people may be willing to do.

The interface and keywords used in describing a social relationship also play a key role. In many sites, everything is conflated under the abused term "friend," while others such as Essembly allow for more contextual labels such as "ally" and "nemesis" (Brzozowski et al. 2008).

Reasoning on Trust

In all the SNSs mentioned earlier, it is possible to interact with unknown people. "Unknown" means that one does not have any firsthand idea about their reliability or about how much trust to place in them. Reputation systems (Resnick et al 2000, Ziegler and Lausen 2004) and trust metrics (Massa and Avesani 2007) are techniques for answering questions such as, "Should I trust this person?" Based on the answer, one can decide whether to interact with another user.

Once trust statements are represented electronically, it is possible to reason based on them. Reputation systems and trust metrics usually work in the following way: first they aggregate all or some of the trust statements expressed by the users in a global or partial trust network, and then they perform some computation in order to predict the reputation or the trustworthiness of all the users. The computation ranges from simple averages for computing a global reputation such as in eBay, to trust propagation over the trust network for computing a global reputation such as with PageRank (Page et al. 1998), to a personalized trust score (Golbeck, Parsia, and Hendler 2003; Massa and Avesani 2007; Ziegler and Lausen 2004), to formal trust algebra based on probability theory (Jøsang 1999). Philosopher John Locke already provided what we have called a trust metric:

> Probability then, being to supply the defect of our knowledge and to guide us where that fails, is always conversant about propositions whereof we have no certainty, but only some inducements to receive
>
> them for true. The grounds of it are, in short, these two following: First, The conformity of anything with our own knowledge, observation, and experience. Secondly, The testimony of others, vouching their observation and experience. In the testimony of others is to be considered: 1. The number. 2. The integrity. 3. The skill of the witnesses. 4. The design of the author, where it is atestimony out of a book cited. 5. The consistency of the parts, and

> circumstances of the relation. 6. Contrary testimonies. (1680, p.
> 886)

One of the main concerns about reputation systems and trust metrics is the fact that they can be attacked and gamed. What are often called “malicious users” can hijack systems in order to get a personal advantage, such as increasing the reputation of an identity (“reputation boosting”) or decreasing it (“reputation nuking”). Usually, reputation boosting is used for a personal identity or that of a friend and reputation nuking is perpetrated on the identity of a competitor or enemy. There have been different recommendations for addressing these threats and making a trust metric more attack-resistant (Levien, n.d.; Massa 2006).

A Change in Perspective: From Global to Local

We consider now a different conceptual approach and claim that a system is attackable if the system is created with the assumption of a correct value of reputation for everyone. In this case, there will be incentives to try to game the system in order to influence this unique and global reputation value. Such a system is inherently attackable. If this assumption is dropped, the threat is weakened significantly.

What we are suggesting is to move from global trust metrics to local trust metrics (Massa and Avesani 2007). Global trust metrics compute a global reputation value for every single user, coming to conclusions such as “the reputation of Carol is 0.4.” On the other hand, local trust metrics predict trustworthiness scores that are personalized from the point of view of every single user, coming to conclusions such as “Alice should trust Carol as 0.9” and “Bob should trust Carol as 0.1.” The very same user Carol can be predicted as trustworthy from the point of view of Alice and as untrustworthy from the point of view of Bob.

Local trust metrics don’t try to average differences of opinion but rather to build on them. The assumption of local trust metrics is that every opinion is worthy and there are no automatically wrong opinions. If someone happens to disagree with the large majority who think that “George is trustworthy,” it may not be useful for society at large to consider his or her opinions as wrong or malicious. Playing on “one man’s signal is another man’s noise,” we might say “one man’s trusted peer is another man’s untrusted peer” (Massa 2006).

Without moving into the contentious domain of political ideas, it is easy to provide an example from the debated domain of peer-to-peer (P2P) file sharing. In a P2P network, Alice might consider “good” a peer that shares a lot of leaked political documents or just-released copyrighted movies (i.e., trust that peer), yet Bob might consider the very same peer “bad” (i.e., distrust that peer). But there is no universally applicable trust statement. Each peer can believe what he or she prefers based on his or her own subjective belief system (though group norms and laws may constrain individual beliefs). Disagreements are a normal part of life and social groups and are often even productive. Without an overriding social concern, there may be no positive utility in trying to squash down differences of opinion.

Although the previous argument is anecdotal, we have offered in an analysis of the Epinions trust network evidence that trust statements are indeed subjective in a real world setting (Massa and Avesani 2007). Epinions is a website at which users can write reviews about products and assign them a rating. Epinions also allows users to express their Web of Trust, that is, “reviewers whose reviews and ratings they have consistently found to be valuable” and their Block list, that is, “authors whose reviews they find consistently offensive, inaccurate, or in general not valuable” (Epinions, n.d.). These expressions correspond to issuing a positive trust statement such as *trust(A,B)=1* and a negative trust statement such as *trust(A,B)=0*, respectively.

We found that on Epinions, it is common to have disagreements of opinion about the trustworthiness of other users; that is, it is common that someone places a certain user in their Web of Trust and someone else places the very same user in their Block List. Typically, these opinions are not wrong or malicious; they represent legitimate differences of evaluation. Simply put, there are users who are trusted by some and distrusted by others; we call these user *controversial users*. Surprisingly, in the Epinions dataset that we evaluated, more than 20 percent of users were controversial users (Massa and Avesani 2007).

For controversial users, global trust metrics are not effective by definition because global averages cannot predict correctly the very different trust statements received by this kind of user. However in Massa and Avesani 2007, we also performed an empirical comparison of local and global trust metrics that demonstrates our claim. Moreover local trust metrics can be attack-resistant (Golbeck, Parsia, and Hendler 2003; Levien, n.d.; Ziegler and Lausen 2004). For instance, if only the opinions of users directly trusted by the active user are considered, it is less easy for an attacker to influence the prediction that the active user gets. As long as the active user does not explicitly trust one of the bogus profiles (and the users whom he or she trusts don't do it either), the bogus profiles are not going to influence computations of trustworthiness values. The user is in control and can check which trusted users, if any, have been fooled into trusting a bogus profile.

Recently, SNSs have been moving toward emphasizing more locality. For example, Essembly, a "fiercely non-partisan social network," (Brzozowski et al. 2008, p. 1) allows members to post "resolves" reflecting controversial opinions, such as "Overall, free trade is good for American workers." Members can then vote on these resolves, using a four-point scale: Agree, Lean Agree, Lean Against, or Against. Users can vote once per resolve, and all votes are viewable by other members, forming an ideological profile. In this site, users can express three different social relationships: friend as “someone you know personally and have a friendship with in the real world,” ally as “someone who you don’t necessarily know, but . . . share a desire to make some change in the world,” and nemesis as “someone who you don’t agree with . . . their world view is just psychotically skewed.” (Brzozowski et al. 2008, p. 2). A striking pattern seems to emerge: an enemy of an enemy seems to be a friend or ally. Similarly to Essembly, Lerman and Galstyan (2008) analyze social voting patterns of Digg users; at Digg, users’ social networks are used to suggest personalized interesting stories.

Two Extremes of Possible Societies as Shaped by Trust Metrics

We would like to conclude by highlighting two extremes of society that can be induced by the basic assumptions behind the two different kinds of trust metrics: tyranny of the majority and echo chambers.

A system powered by a global trust metric, in effect, tends to assume that there are globally agreed-upon good users and that people who think differently from the average are malicious. This assumption encourages herd behavior and penalizes creative thinkers, black sheep, and original and unexpected opinions.

We underline that there is a *tyranny of the majority* risk—a term coined in 1835 by Alexis de Tocqueville in his book *Democracy in America* (1840). Nineteenth-century philosopher John Stuart Mill in his book *On Liberty* (1859) 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. On each specific issue, people will express themselves either for or against the issue, and the side with the largest number of supporters will prevail. 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.” (This quote is extracted from Wikipedia (Massa and Avesani 2007), which interestingly tries to find a balance between what different people think about every single topic by asking the contributors to adopt a “neutral point of view” (NPOV). This approach seems to work well enough in most cases at present, possibly because the people who self-elect for editing Wikipedia articles largely share a similar “culture.” However, the frequent “edit wars” 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.)

We believe that the minority's opinions should be seen as an opportunity and a point of discussion and not simply assumed to be "wrong" or "unfair" ratings, as they are often modeled in simulations in trust metrics research papers. Moreover, in digital systems such as SNSs, automatic personalization is possible, so there is in principle no need to make this assumption and try to force users to behave in the same way.

Research carried out by Salganik, Dodds and Watts (2006) and colleagues is enlightening in this regard. They created a music website at which users could rate and download songs by unknown bands. The home page also showed the top ten list, that is, the most popular songs (or, if you like, the songs currently appreciated by the majority). They divided the users into eight separated copies of the site with the same songs in them, without the users knowing it. The striking result was that in the eight separated sites, the top ten lists originated by user ratings were different. Popularity was not primarily induced by some intrinsic quality of the songs, but by aggregated ratings activity and how it was displayed in the top ten list. This result suggests that an artist like Britney Spears, who is popular in this world, might have been a nobody in some other world. Different majorities, formed based on different and seemingly random patterns, imposed on the community and its minorities a certain top artist. This fact was unavoidable in the mass media era. But now that personalization is possible, is it necessary to constrain everybody into a global “best” (such as Britney Spears) when we can end up with many different local “bests”?

We can ponder the other extreme: total personalization. But there is a risk in this extreme as well that is caused by emphasizing too much locality in trust propagation by a local trust metric. This approach consists in fact in considering, for example, only opinions of directly trusted users and not of the rest of the community constituents.

This risk has been called the *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, p.3). He warns how in this way everyone has the ability to listen to and watch just what they want to hear and see—to encounter only opinions of like-minded people and never be confronted with people with different ideas and opinions.

In this scenario, there is a risk of segmentation of society into micro groups that tend to adopt extreme views, develop their own culture, and not communicate with people outside their group. Sunstein 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 “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, 1990, p. 9). Recent research published in the *Psychological Bulletin* shows that people are about twice as likely to select information that supports their own point of view (67 percent) as to consider an opposing idea (33 percent) (Hart et al. 2009).

These considerations are not new. As cited by McPherson, Smith-Lovin, and Cook (2001), in "Rhetoric" and "Nicomachean Ethics," Aristotle noted that people "love those who are like themselves"; in addition, Plato observed in "Phaedrus" that "similarity begets friendship." McPherson, Smith-Lovin and Cook (2001) conducted a large analysis of *homophily* (the tendency to bond with others who are similar) in social networks and found that "homophily in race and ethnicity creates the strongest divides in our personal environments, with age, religion, education, occupation, and gender following in roughly that order." They conclude commenting how homophily limits people's social worlds in a way that has powerful implications for the information they receive, the attitudes they form, and the interactions they experience.

A society without some common shared culture cannot be defined as a society. The societal utility of "massification" is that we, as a society, can rely on some cultural artifacts that bond us together. Knowing we can rely on common culture is reassuring, and allows different people to feel some bonds as a group and as a society. When, or if, there are no more cultural elements able to bond us together because everybody has become a singleton with her peculiar and totally personalized culture, the very existence of our society may be at risk (Sunstein 1999).

Conclusion

As we have seen, concerns about ways of trusting and the societies they induce are not new. What is new is that information and communication technologies are playing an increasing role in shaping our future societies.

We believe that in the near future, more and more people will increasingly rely on opinions formed based on facts collected through reputation systems and social network sites (Massa 2006). The assumptions on which these systems are constructed will have a fundamental impact on the kinds of societies and cultures they shape. Here we have offered "tyranny of the majority" and "echo chambers" as a way to think about the two extremes of a range of options toward which our society might evolve.

The final and very open question is, "Will we be able to mediate between the two extremes?" This is surely not an easy task. We hope this paper can help modestly by providing some starting points for a fruitful and ongoing global discussion about these ethical issues so important for our common future.

References

boyd, d. m. and Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13: 210–230 Pl

Epinions.com. n.d. Web of Trust FAQ. Epinions. Retrieved November 7, 2010, from: <http://www.epinions.com/help/faq/?show=faq_wot>.

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

Golbeck, J., B. Parsia, and J. Hendler. 2003. Trust networks on the semantic web. *Proceedings of Cooperative Information Agents* VII:238–249.

Hart, W., A. H. Eagly, M. J. Lindberg, D. Albarraccin, I. Brechan, and L. Merrill. 2009. Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin* 135 (4): 555–588.

Brzozowski, M. J., T. Hogg, and G. Szabo. 2008. Friends and foes: ideological social networking. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*. ACM, New York, NY, USA, 817-820..

Jøsang, A. 1999. An algebra for assessing trust in certification chains. *Proceedings of the Network and Distributed Systems Security Symposium (NDSS99), San Diego, California*.

Lerman, K., and A. Galstyan. 2008. Analysis of social voting patterns on Digg. *Proceedings of the ACM SIGCOMM Workshop on Online Social Networks.*

Levien, R. n.d. Attack-resistant trust metrics. Unpublished doctoral dissertation. Retrieved from: <http://www.levien.com/thesis/thesis.pdf>.

Locke, J. 1680. *An essay concerning human understanding*. Sussex, UK: Harvester Press.

Massa, P. 2006. A survey of trust use and modeling in current real systems. In *Trust in E-services: Technologies, Practices, and Challenges*, ed. R. Song, L. Korba, and G. Yee. Idea Group. Pp. 51-83.

Massa, P., and P. Avesani. 2007. Trust metrics on controversial users: Balancing between tyranny of the majority and echo chambers. In *International Journal on Semantic Web and Information Systems*, Special Issue on Semantics of People and Culture, ed. H. Liu and P. Maes. Pp. 39–64

Massa, P., Souren, K., Salvetti, M. and Tomasoni, D. 2008. Trustlet: Open research on trust metrics. *Scientific International Journal for Parallel and Distributed Computing*, 9-4: 341–351.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27:415–444.

<bok>Mill, J. Stuart. 1859. *On liberty*. McMaster University archive for the history of economic thought. London. J.W. Parker and son editors.</bok>

<bok>*Oxford English Dictionary*. 1990. Reputation entry. New York: Oxford University Press.</bok>

<bok>*Oxford English Dictionary*. 1990. Trust entry. New York: Oxford University Press.</bok>

Page, L., S. Brin, R. Motwani, and T. Winograd. 1998. The PageRank citation ranking: Bringing order to the web. *Proceedings of ASIS98*, 161–172.

Putnam, R. D. 1995. Bowling alone: America's declining social capital. *Journal of Democracy* 6 (1): 65–78.

Resnick, P., K. Kuwabara, R. Zeckhauser, and E. Friedman. 2000. Reputation systems. *Communications of the ACM* 43 (12): 45–48.

Salganik, M. J., P. S. Dodds, and D. J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311 (5762):854–856.

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

de Tocqueville, A. [1840] 1966. *Democracy in America*, trans. G. Lawrence. New York: Doubleday.

Toshio, Y. 2001. Trust as a form of social intelligence. In *Trust in Society*, ed. C. Cook. New York: Russell Sage Foundation. Pp: 121-147

Zak, P.J. 2006. Trust. CAPCO Institute - The Journal of financial transformation, pp. 17-24.

Ziegler, C. N., and G. Lausen. 2004. Spreading activation models for trust propagation. In IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE '04. Taipei, Taiwan.), 83–97. IEEE.