

# 1 Designing Reputation Systems for the Social Web

Chrysanthos Dellarocas

*How can online reputation systems be designed to meet the needs of both users and system designers? Chrysanthos Dellarocas looks at the promises and challenges of online reputation mechanism design and how to implement incentives for good behavior, truthful reporting, and resistance to manipulation.*

Reputation systems are arguably the unsung heroes of the social web. In some form or another, they are an integral part of most of today's social web applications. Yet they usually play a supporting role and have thus received less attention than some of their higher-profile cousins such as social networks, recommender systems, consumer review sites, and crowdsourcing communities, whose success would often have been unsustainable without the quiet but effective support of reputation systems (Dellarocas 2003). Two examples include eBay, a trading community of virtual strangers, which would have been impossible without the reputation system that builds trust by enabling buyers and sellers to rate each other; as well as Amazon, where product reviews are rendered both more useful and more credible by the presence of a scoring system that allows readers to vote on a review's usefulness. Amazon employs these votes to point customers to the most useful reviews and socially reward the most successful reviewers. Other examples are Yelp.com, a youthful urban review community that similarly uses what is essentially a reputation mechanism to help users better interpret opinions that are inherently subjective, and Xbox Live, a gaming site that makes widespread use of scoreboards to reward skilled gamers, help gamers find opponents of similar skill level, and provide incentives for people of all skill levels to keep playing.

This chapter provides an introduction to what we know about the practical design of these ubiquitous but undervalued systems. It begins by outlining the surprisingly diverse set of business objectives that reputation mechanisms satisfy in today's social web. It then highlights four key decisions related to their design and what we know about each one. It concludes with a brief case study of a real-life reputation system.

## What Is a Reputation System?

There are no widely agreed-upon definitions of reputation and reputation systems. Even within this volume, different authors define reputation using slightly different terms. Given the many reasons that websites employ reputation systems, this chapter uses the following correspondingly broad pair of definitions. *Reputation* is a summary of one's relevant past actions within the context of a specific community, presented in a manner that can help other community members make decisions with respect to whether and how to relate to that individual (and/or to the individual's works). Accordingly, a *reputation system* is an information system that mediates and facilitates the process of assessing reputations within the context of a specific community.

## Why Do Social Web Applications Use Reputation Systems?

The concept of reputation is most often associated with the notion of trust. Although trust building has historically been an important role of reputation, today's web-based reputation systems serve a much broader set of objectives for users and system designers. These objectives include the following:

- *Trust building* This goal is achieved by encouraging "good" and discouraging "bad" social behaviors within the context of a site. This is perhaps the most obvious and widely discussed role of reputation systems, best exemplified by eBay.
- *Filtering* Most web-based systems are characterized by an abundance of information that can quickly overwhelm their users. In the majority of settings, the quality of contributed content tends to be very uneven. Reputation systems can improve the attractiveness of the systems they support by helping users easily identify the highest-quality contributions.
- *Matching* This objective is important in settings where members vary widely in interests and tastes and where the quality of contributions is characterized by a high degree of subjectivity. In such settings, reputation systems ought to summarize the relevant attributes of a member in a neutral manner so that others can assess the degree to which they can relate to him or her.
- *User lock-in* Because reputation is usually specific to a single community or web system, it constitutes a powerful form of lock-in and can be used strategically by system designers to increase user loyalty and decrease attrition. Once a user has built a reputation on a site, that user will be reluctant to defect to a competitor because he or she would then have to build his or her reputation from scratch. In an increasingly competitive environment, this dimension of reputation mechanisms must not be overlooked. It poses some interesting questions (e.g., who owns a user's reputation?) and may at times pit user interests against system designer interests.

All four objectives are relevant to the design of most reputation systems. Different systems, however, assign different priorities to each of these objectives. For example, building enough trust so that buyers can feel comfortable enough to send their money to sellers they have never met—and, very likely, will never buy from again—is arguably the primary objective of eBay’s reputation mechanism. On the other hand, the most important objective of Amazon’s reviewer reputation mechanism is to induce members to contribute well thought out, high-quality reviews and to identify (i.e., filter) the highest-quality reviews.

Inducing high-quality contributions is high on Yelp’s agenda as well. However, because the majority of Yelp reviews are subjective, it is usually difficult to reliably assess a review’s quality. What is more important is to provide tools that can help readers make a personal assessment of a review’s credibility and compatibility with their view of the world. Finally, Xbox Live is an example of a system in which user loyalty and retention is the overarching (business) objective of its reputation feature.

### What Are the Key Design Decisions of a Reputation System?

According to the definition stated previously, reputation is a summary of relevant past actions. Designing an effective reputation system correspondingly entails four key decisions:

1. What actions are relevant to include in one’s reputation profile?
2. How to obtain information about these actions
3. How to aggregate and display reputation information
4. How to deal with manipulation and gaming

#### Key Decision 1: Which Actions Are Relevant to Include in One’s Reputation Profile?

Users typically engage in rich behaviors within the context of a site. Effective reputation systems must carefully choose the aspects of user behavior to track and report on. Making this decision correctly requires a clear understanding and prioritization of the reputation system’s business objectives.

There are two questions to be considered in this decision:

- *What actions are most relevant to the reputation system’s users?* For example, if the key objective of the system is to help users decide whether a seller is honest, keeping track of the seller’s percentage of completed transactions is a very relevant indicator, whereas keeping track of a seller’s own purchase history is less relevant. On the other hand, if a system’s key objective is to help users determine whether a reviewer has similar tastes to their own, keeping track of a reviewer’s purchase history might be a very relevant indicator.

- *What user behaviors are desirable to encourage or discourage?* The mere act of publicly keeping track of someone's actions can encourage or discourage the incidence of those actions (depending on whether the action in question has a good or bad connotation). Therefore, if a site wants to encourage the volume of contributions, its designers might consider keeping track of the number of reviews posted or the number of comments posted on somebody's reputation profile. On the other hand, if a site wants to encourage the quality of contributions, it might want to hide information about contribution volume and to keep track instead of how other people rated a particular contribution.

### **Key Decision 2: How to Obtain Information about Relevant User Actions**

The main choice here is between internally generated (firsthand) information and (secondhand) feedback provided by others. For example, Yelp's counts of the volume of a user's posted reviews or of the size of one's friend network are examples of firsthand information. In contrast, eBay's reliance on asking the buyer and the seller what they thought about a transaction is an example of secondhand feedback.

Firsthand information about user actions is generally preferable, as it is accurate and unbiased. Unfortunately, it is not always available, or its collection might require the development of costly additional infrastructure. For this reason, reputation systems must often rely on secondhand feedback provided by others. Secondhand feedback is subject to two important shortcomings that must be addressed by the designer: untruthful reporting (i.e., lying) and reporting bias. *Untruthful reporting* can often be kept in check by using some form of community policing. "Rate-the-rater" schemes whereby users rate the "usefulness" or "credibility" of feedback postings appear to work well in many settings, even though in theory such schemes are not difficult for determined users to game (Lampe and Resnick 2004). Some researchers have proposed payment mechanisms that induce truthful reporting. Such mechanisms pay reviewers an amount that is based on how well their ratings of an individual can predict future ratings posted by others about the same individual (Miller, Resnick, and Zeckhauser 2005). A schedule of payments that makes it optimal for all reviewers to report the truth can theoretically be devised. However, such mechanisms break down in the presence of collusion and have not been widely used in practical systems.

*Reporting bias* is a situation in which users are systematically more likely to post feedback when certain types of outcomes occur and to remain silent otherwise. For example, it has been documented that users are generally more likely to post feedback when they have had extreme (either very good or very bad) experiences than when they have had average experiences (Dellarocas and Narayan 2006). Reporting bias can reduce the reliability of the reputational information. For example, reporting bias is largely responsible for the improbably high percentage (approximately 99 percent) of positive feedback that was the norm on eBay until 2009 (when, as I discuss in the next paragraph, eBay made some important changes to its mechanism). It appears that in

a lot of middle-of-the-road transactions in which things don't go great but which are still not a disaster, eBay users have preferred to stay silent rather than to post neutral or negative feedback (Dellarocas and Wood 2008).

Preventing reporting bias is tricky, as it is not usually practical or desirable to force members to post feedback. In some cases, one can use domain knowledge and statistical methods to filter out reporting bias when aggregating feedback. In other cases, one might be able to eliminate some of the sources of this bias by redesigning the mechanism. For example, it appears that the fear of retaliatory negative feedback from sellers discourages many buyers from posting neutral or negative feedback when they are unhappy with a transaction. Recognizing this, eBay recently changed its mechanism so that it no longer gives sellers the option to post unfavorable feedback for buyers.

### **Key Decision 3: How to Aggregate and Display Reputation Information**

Reputation mechanisms employ a variety of methods for displaying outputs. These include:

- Simple statistics (e.g., number of reviews posted, number of transactions completed)
- Star ratings (e.g., Amazon reviews)
- Numerical score (e.g., eBay's reputation score, and the technical news site Slashdot's karma score—see chapter 7 in this volume)
- Numbered tiers (e.g., World of Warcraft's player levels)
- Achievement badges (e.g., eBay Power Seller, Yelp Elite Reviewer)
- Leaderboards (lists where users are ranked relative to each other and top users are highlighted, e.g., list of top Amazon reviewers)

The choice of a display method is very important because it can determine: (a) the extent to which the reputation mechanism makes a judgment versus allowing users to make their own judgments, and (b) the extent to which the presence of the reputation system can create competition among users.

Displaying simple statistics of a person's activity within a community is perhaps the most neutral method of summarizing one's reputation. Social network statistics in particular have been rapidly rising in popularity in many contexts. The advantage of using simple statistics is that the reputation system makes minimal judgments and allows users to draw their own conclusions. The disadvantage is that the burden for interpreting these quantities falls on the shoulders of the user, who must be familiar enough with the environment to draw the proper conclusions.

Star ratings and achievement badges solve this problem by immediately communicating to users whether somebody's performance along some dimension is "good" or "bad." This method helps users more easily digest information, but also makes an explicit judgment of quality. It is therefore best applicable in settings in which there is a commonly agreed-upon notion of what "quality" means.

Whereas star ratings and achievement badges highlight individual achievement, numbered tiers and leaderboards go a step further, as they not only confer a judgment but also indicate a user's standing in relation to others. They thus introduce a pecking order among users, which can have a very strong impact on a community's culture and behavior (see chapter 2 in this volume). Supporting direct comparison of users against one another often increases incentives to contribute and enhances the filtering role of a reputation system. On the other hand, such direct comparison instills a culture of competition, which might end up being disruptive in a number of ways. First, obsession with rankings might lead some users to manipulative and counterproductive behavior. Second, users who don't make it to the top ranks might feel resentful and be induced to exit.

The story of Digg.com provides an example of the dangers of introducing too much competition in a community whose culture is otherwise meant to be cooperative. Digg is a system that allows users to recommend news articles that piqued their interest. Other users vote recommended articles up or down. An article's number of votes determines its relative visibility on the site. The site was meant to be a grassroots method of determining what is interesting to read about today.

Digg implemented a reputation mechanism that originally included a leaderboard listing the most "influential" Diggers—members whose recommended articles received the most votes. Some of the leading Digg members became so obsessed with their rankings that they formed a collusion ring for the purpose of ensuring that their recommendations always made it to the top, which ended up eroding the culture and dynamics of the site. Amid controversy, in January 2007 Digg decided to abolish the Digg leaderboard. As of June 2010, Digg's user profiles consisted of various "simple statistics" only.

#### **Key Decision 4: How to Deal with Manipulation and Gaming**

The sensitive and personal nature of reputational information makes reputation systems a natural target for manipulation and gaming. Anticipating and minimizing the impact of such behavior ought to be high on every site designer's agenda. Interestingly, making appropriate choices in the preceding three key decisions can often go a long way toward making a reputation system less prone to manipulation.

For example, before deciding to include an item of information on a reputation profile, designers should carefully think through the types of undesirable behaviors that users might engage in to boost their score on that dimension. Also, before asking users for feedback, the designer should carefully consider whether other aspects of the system induce reporting bias or give incentives to lie. Finally, a designer should be careful to not instill an unnecessarily competitive spirit by overemphasizing comparisons across users, which might tempt some users to engage in manipulative behavior.

In addition to these considerations, the following two design dimensions can influence the manipulation resistance of a reputation system:

- *Identity of feedback sources* The degree to which the true identity of the sources of feedback is known to the system and/or the community is a key variable. Full anonymity and easy creation of virtual identities makes it easy for determined users to flood a reputation system with fake ratings. Mapping virtual identities to real identities (e.g., by asking for credit card numbers during registration) can help keep such behavior in check. At the same time, however, controlling identities has drawbacks of its own. First, it might discourage some users from joining the system. Second, it might discourage users from posting anything but positive feedback, increasing reporting bias and reducing the usefulness and credibility of the system. Sometimes a middle-of-the-road approach whereby a reviewer's identity is fully known to the system but not disclosed to other members can work better than either of the two extremes.
- *Transparency of aggregation rules* Concealing the details of the algorithms used to aggregate feedback is another way of resisting manipulation and gaming. Amazon and Google.com have gone down that path. The former does not disclose the precise formula they use to rank-order reviewers, and the latter does not disclose all the details of rank-ordering search results. Lack of transparency, however, has a downside, as it can lessen a site's credibility and the users' trust in the reputation system's fairness.

In general, it is impossible to design a totally manipulation-resistant reputation system. No matter what mechanisms one puts in place, creative and determined users are bound to find a way around them. For that reason, community administrators must constantly monitor such systems, organically evolving their designs.

### A Case Study: Yelp

Yelp, as mentioned earlier, is an online review community that specializes in reviews of restaurants and entertainment establishments in major urban centers. At the same time, Yelp is a community of mostly young users that relate to each other through similar tastes and interests. Yelp maintains a profile page for each of its members. The profile page is essentially a reputation mechanism.

An important objective of Yelp's profile feature is to help other users decide how to interpret a review by finding out more about a reviewer's personality and tastes. At the same time, the profile allows a member to showcase his or her status and contributions to the system, building site loyalty and providing incentives for continued contributions.

Because Yelp's domain (restaurants, entertainment, and other cultural establishments) is highly dependent on personality and taste, Yelp's profile feature wisely avoids the use of numerical scores and star ratings. Instead, a reviewer's profile primarily

consists of a series of simple statistics that summarize salient aspects of the user's activity. In addition to simple statistics, Yelp supports secondhand feedback in the form of "compliments" received from other users. However, these compliments are displayed below the neutral statistics and are thus not highlighted as prominently. There are also several different types of compliments ("good writer," "funny," "you're cool," etc.) underlining the plurality of ways in which a user can make valuable contributions and making it difficult to compare users against each other.

Although Yelp supports a leaderboard, it is interesting that the ranking feature is not easily accessible from the home page and that Yelp offers multiple dimensions across which users can be rank-ordered—once again emphasizing that one can stand out in this community for multiple reasons and making it difficult to compare users against each other.

In summary, Yelp's reputation system design is well suited to its business goals of inducing participation and valuable contributions and facilitating the matching of like users in a domain where the notion of "value" has a highly subjective definition.

## Conclusions

Reputation systems are ubiquitous in the social web, and their design principles surprisingly subtle. For a more comprehensive introduction to reputation system design theory, see Dellarocas (2006) and Dellarocas, Dini, and Spagnolo (2007). For a practitioner-oriented perspective, see Dellarocas (2010) and Farmer and Glass (2010).

In closing, it's worth reiterating the following points that the author feels have not received their fair share of attention in previous literature:

- The design choices of a reputation system can profoundly affect a community's culture, making an otherwise collaborative and cordial community into a competitive and even combative space. Users often obsess over rankings and scores even when they do not receive any economic benefit from them. Although such features can increase incentives to contribute, they can also provoke manipulative and antisocial behaviors.
- A reputation system can be an important source of user loyalty and a powerful mechanism for user retention. This aspect of reputation must be factored into any proposed design change.
- Reputation mechanisms serve a variety of objectives apart from trust building. These include filtering content, matching users, and building user loyalty. Understanding and prioritizing a system's business objectives should be a designer's first task.

## References

Dellarocas, C. 2003. The digitization of word-of-mouth: Promise and challenges of online reputation systems. *Management Science* 49 (10): 1407–1424.

Dellarocas, C. 2006. Reputation mechanisms. In *Handbook of economics and information systems*, ed. T. Hendershott, 629–660. Amsterdam, The Netherlands: Elsevier Publishing.

Dellarocas, C. 2010. Online reputation systems: How to design one that does what you need. *Sloan Management Review* 51 (3): 33–38.

Dellarocas, C., F. Dini, and G. Spagnolo. 2007. Designing reputation mechanisms. In *Handbook of Procurement*, ed. N. Dimitri, G. Piga, and G. Spagnolo, 446–482. Cambridge: Cambridge University Press.

Dellarocas, C., and R. Narayan. 2006. A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science* 21 (2): 277–285.

Dellarocas, C., and C. A. Wood. 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* 54 (3): 460–476.

Farmer, R., and B. Glass. 2010. *Building web reputation systems*. Sebastopol, CA: O'Reilly.

Lampe, C., and P. Resnick. 2004. Slash(dot) and burn: Distributed moderation in a large online conversation space. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, 543–550.

Miller, N., P. Resnick, and R. Zeckhauser. 2005. Eliciting honest feedback: The peer prediction model. *Management Science* 51 (9): 1359–1373.