Otello: A Next-Generation Reputation System
For Humans and NPCs

Michael Sellers

Online Alchemy, Inc.
mike@onlinealchemy.com

Abstract

This paper introduces Online Alchemy's Otello technology as a way to enable reputational capabilities beyond any found in games or other online social contexts today. This technology allows participants to quickly and easily assess another's reputation in ways meaningful to them, and enables individuals – both players and non-player characters (NPCs) – to contribute to an individual's reputation in unique and novel ways. Otello also enables new forms of "relational gameplay" that feature social management, effectively an extension of resource management into the social realm. The player's actions and opinions affect others, including how they see the player, and how ideas and opinions propagate through a population.

Introduction

As more games merge into online social spaces, reputation systems are becoming increasingly important to players looking for gaming partners. Existing reputation systems are based on diffuse global scores, and do not give players the information they need – they do not provide either subjective reputation based on the social graph or significant social gameplay capabilities. Online Alchemy's Otello reputation engine is designed as the next step in reputation technology. It provides both meaningful subjective reputation information and enables entirely new forms of reputational and relational gameplay.

Opinions and Reputation

In any multi-actor context, each actor (player or non-player character (NPC)) may form opinions about others based on directly observing their actions (this assumes a certain amount of situational awareness for NPCs). An individual may like, respect, fear, or disapprove of another based on how they act, and this is the observer's opinion based on direct experience.

By contrast, someone's *reputation* is based on indirect experience. That is, if Anne knows Bob but not Carl, Anne may at most be informed of Carl's actions by Bob (or she

may remain entirely ignorant of Carl). If Bob tells Anne that Carl is a good guy or has done good things, Anne is likely to think well of Carl – this is his reputation with her, assuming for the moment that Anne also thinks well of Bob (see Figure 1).
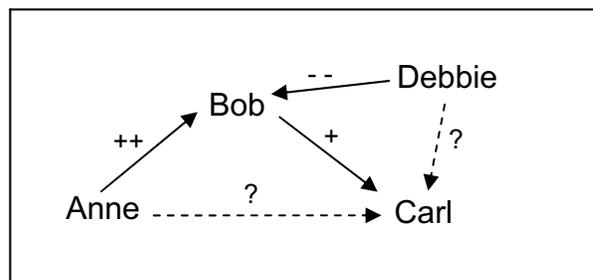

Figure 1: Anne does not know Carl, so his reputation with her comes through Bob. Debbie does not trust Bob, so his opinion of Carl does not matter to her.

In short, reputation is an opinion based on *indirect experience related to an individual by trusted sources.* Abdul-Rahman and Hailes (2000) define reputation as "an expectation about an agent's behavior based on information about or observations of its past behavior." This captures the important aspect of future expectation based on past actions, but not the indirect aspect of reputation. Glass (2008) says that in current online communities reputation "is equal to the sum of your past actions" and is "determined by the community" in which the reputation resides. This is consistent with current systems implementing global reputation, but this type of design negates much of the power of reputation found in everyday experience.

Unlike opinion from observation, reputation is indirect. How an individual interprets reputation commentary from another person depends on the trust assigned by one person to another. As such, this assigned trust is always personal and subjective, not global or objective. For example, Anne trusts Bob, but Debbie does not. As such, Bob's opinion of Carl is of little use to Debbie, and so does not affect Carl's reputation with Debbie (see Figure 1). By contrast, saying that reputation is global or community-based implicitly

assigns equal trust to all actors in the community, which in most cases is socially an unwarranted position.

While trust is personal and subjective, it can be inherited transitively with some loss due to "friction" of indirect social contact. Even if Anne trusts Bob, she will be less confident of Bob's opinion, and thus of Carl's reputation, than if she knew Carl directly. Carl's opinion matters less to Anne that does Bob's, but more to her than someone with whom she has no relationship. This transitive decrease in trust is known as a continuous (rather than binary) *discounting factor* in subjective logic (Jøsang 1997), and allows opinions and reputation to spread throughout a social network in proportion to their relevance, without doing so infinitely.

Since trust is always personal and imperfectly transitive, reputation is also personal and subjective. This is an important aspect of social relationships that we each encounter every day, and yet which is not included in existing reputation systems. An individual like Carl in the previous example may be well thought of in some circles and very poorly thought of in others. Given this, it is not sufficient in creating a useful social model to assign a single, objective, global number to Carl and call that his reputation. His reputation is inherent in and subjective to those considering him, not contained in or attached to himself. Enabling this subjective view technologically opens the door to having, for example, Montagues and Capulets, where one individual may be seen differently by two (or more) formal or ad hoc groups. This non-monotonic nature of reputation is something not found in existing systems and yet is crucial for effective relationship modeling in social and game contexts. It enables both more informative matching systems as well as entirely new forms of gameplay, as discussed below.

Current Reputation Systems

Existing reputation and rating systems fall into one of two types: either they create global reputations based on the averaged opinions of all commentators, or they use centralized authorities to create global reputations inherited by all observers. In either case, reputation is currently seen as a single quantity or score inherent in the targeted individual, and that is seen the same by all observers.

Examples of global-average reputation and rating systems include those used by online sites like eBay, Slashdot, iKarma, Digg, Orkut, and Xbox Live. eBay is in some ways the canonical global reputation system, in that an individual's reputation there depends on the averaged opinions of everyone who chooses to submit a rating. In such systems each individual typically has one numerical rating (though sometimes this number is hidden behind text covering an interval of ratings), and the rating is based on a simple average of submitted values. In some systems, notably Slashdot's, ratings are culled from a subset of the entire community, but even here these are taken from an essentially random grouping without reference to the relationships between individuals.

The second primary type of reputation systems are authority-based, such as Advogato or Venyo. In these systems there are authoritative individuals who act as vetting nodes in a social graph: the originators are granted "most trusted" status, and assign levels of trust to others. Each individual can assign to others levels of trust less than or equal to their own, thereby excluding bad actors, assuming no mistakes are made. This type of system is analogous to the credit reporting system used in the US, where individuals have a global score attached to them by presumptive trusted agencies. As in global-average systems, each individual has a single score that is the same no matter who is evaluating them.

Reputation systems in games have been employed at least since *Sid Meier's Pirates* (MicroProse 1987) where based on their actions players could affect their standing with different nations, and more recently in games like *Oblivion* (Bethesda Softworks 2007) and *Fable* (Microsoft, 2004). In many massively multiplayer online games (MMOGs) such as *World of Warcraft* (Vivendi 2004) players gain or lose reputation with non-player factions based on certain (typically quest-based) actions, but these games do not tend to have player-player reputation systems.

Problems with Existing Reputation Systems

While existing reputation systems in social and game settings have become widespread, they also have a number of serious issues that limit their utility.

Identity issues Most reputation systems assume that every individual has a single stable, known identity. Unfortunately in these systems it is often easy to create fake identities. These can be used to gang up on another person or add fictional support to the user's primary identity (so-called 'bombing' and 'Sybil' attacks (Douceur 2002), respectively). Both forms of "anonymity without responsibility" distort an individual's reputation by the creation of false data.

Spurious and negative ratings In any global aggregation system, bad actors can create negative results that have no basis, but which persist and are viewed by others who have no context for the rating. The reverse can also cause issues, where individuals are reluctant to supply negative ratings if doing so would put them at risk socially. For example Resnick and Zeckhauser (2002) note that "the net feedback scores that eBay displays encourages Pollyanna

assessments of reputations, and is far from the best predictor available."

This issue can be exacerbated by "reputation bombing" where individuals expressing opinions out of line with the mainstream are "bombed" into either submission or irrelevance by piling on many negative ratings by others. In some cases this approaches a form of social extortion by so-called "trustitutes" (as on sites that pay a nominal amount for opinions that are well-regarded by others) who will support you if you support them, but who will marshal others to exclude you if you do not comply (Epinions 2004).

Opinion convergence In any reputation system that depends on global or authoritative opinions, there is no room for divergence of opinion. If "everyone" thinks an individual or content item is great, those who disagree are effectively silenced; at most their opinion may change the aggregate value slightly. In other words, there is no room for valid differences of opinion or clusters of interest: if the mass of opinion says that a particular individual, movie, guild, or song is the best ever, other opinions become invisible.

This is particularly apparent in authority-based systems such as Advogato. Once an individual is assigned a high trusted status it is all but irrevocable, no matter how much others may disagree. This became starkly apparent in the celebrated 'Mentifex' example, where an individual somehow achieved Master or 'most trusted' status despite the objections of others in the community (Taylor 2007).

This convergence of opinion also leads to the devaluing of contrary opinions, in that individuals who disagree have no incentive to speak up or act because of the perception that their opinion cannot change anything.

Otello and Dynemotion

Online Alchemy has been developing an advanced form of agent-based artificial intelligence for use in games since 2002. As part of this, the company has created the Dynemotion People Engine, a system that creates agents with their own personalities, emotions, and goals, and that learn and form opinions and relationships with others around them. These agents remember and learn from what they experience first-hand, and pass on to each other their own memories and those told to them by others. This capability formed the core of Otello, a reputation engine suitable for use in games and social environments, and which enables reputation to be accrued and transmitted between players and non-player characters.

Otello is modeled on information flow and the semantics of trust in human relationships in a computationally scalable way. As noted above, Otello uses a form of subjective logic to codify opinions and reputations. All opinions are subjective and include continuous rather than discrete rating and confidence measures. Applying confidence as a form of trust transitively and accumulatively across a social graph enables the creation of a singular, subjective opinion and confidence measure. This reputation value represents a subjective view of an individual or piece of online content. Instead of "what everyone thinks" of someone or some thing, each individual sees a weighted aggregation of what their social network thinks. Thus, a single individual such as Romeo Montague or Paris Hilton will be seen as important or irrelevant and with different degrees of assurance by different people. Such individuals have entirely different reputation values based on the weighted opinions of people in different social networks.

Otello does not use any estimating algorithms to predict what people might think, nor does it depend on centralized or imposed global values. The global reputation value for any person or content can be computed, but this is typically of less value than the opinions of an individual's social network. Nor does Otello depend only on first-rank 'friends' as in many existing social networks. A reputation score is built from the recursive aggregation of the opinions of those whom you trust, and those whom those people trust, continually expanding outward until there is insufficient confidence to use additional opinions in the network.

Model and Usage

Otello is an engine that does not specify a single user interface, but the basic functions remain the same across installations. In using Otello, each individual can state their opinions via ratings along with optional tags. Each rating consists of a value and confidence, indicating an opinion and how sure the rater is about that opinion. Both are continuous values and can be represented in numerous ways, including numbers, stars, colors, etc.

Opinion ratings can be created about other individuals or things – blogs, guilds, in-game items, etc., depending on the context. Opinion ratings can also be made on tags themselves to rate an general topic: if an individual dislikes politics, he or she can rate the tag itself low, which has the effect of reducing the effective rating on all opinions carrying the tag 'politics.'

As individuals rate each other and items, Otello constructs an internal social graph of one-way opinion links. Positive values indicate increasing trust and interest, while increasingly negative values indicate interest without trust. Values near zero indicate a lack of interest and no particular trust on the part of the rater. Low confidence on a link indicates a tenuous opinion, while high confidence indicates more assurance on the part of the rater.

Using this method, "friends" are those with whom you share bi-directional positive ratings. Rating someone highly where this link is not reciprocated indicates a "subscriber" or "fan" relationship. Individuals with many more inbound than outbound links are social hubs with a certain amount of celebrity. Note that each individual can choose to make their opinion visible to others or not, so there is no social repercussion to not rating someone or rating them negatively.

Network Search Using the social graph made up of individuals' ratings, the Otello engine can perform user-centric network searches to provide subjective accumulated reputation values. The engine searches the user's outbound positive links, and then those nodes' outbound links recursively until the confidence value falls below a minimum threshold where no more useful information is available. This is a model of 'social distance' where you are unlikely to put much stock in your friend's friend's friend's opinion. The use of this kind of discounting factor also ensures that any network search is finite no matter the connectivity in it.

Note that outbound negative links are not used in personal network searches; rating someone poorly effectively reduces or removes their opinion (and that of those whom they value) from your consideration. In other words, you care what your friends' friends think, but not what those whom you don't trust think.

As the engine searches the user's network, it avoids looping back (the network is likely a graph with many cross-connections) as it aggregates all relevant opinions. These opinions are weighted by value and confidence based on the connections between the user and the node being searched.

The network search can return a value about a specific person or item ("what does my network think of person X, or about this new movie?"), or it can return a ranked list of the most important people and/or items in your network. In this way, Otello acts like a personalized content rating system, where the rankings seen are based on the trust assigned by the user, not on what "everyone" or even just "your friends" think.

Global Search In some cases it is preferable to search the entire network for an aggregated opinion or a "top ten" list. In particular, when a new user joins the system and is not yet embedded in their own social network, this provides them a way to see what others think is interesting, and to begin creating their own opinions and network.

Individual Opinions A user can always see and change their own opinions, query what another user thinks about a person or item, or what that person has rated recently. This becomes like a blog of ratings and a repository for opinions. Individuals can choose to limit the visibility of their opinions to themselves, their friends, or those in a particular group. This way Bob can keep his opinion of Debbie to himself, only let his trusted friends see his opinion of Carl, and keep his interests in international affairs and action figures separate for only individuals in those particular interest groups.

Benefits

Otello enables users to record their own opinions of online or in-game items, posts, people, etc. and to see the opinions of those whom they trust in individual or aggregated form. Visible opinions include those from their friends as well as popular authors or in-game leaders. Reputational ratings are based on each individual's outlook and not on any enforced global or authoritative ratings.

Security Each individual's opinions are secure and discreet: no one can change another's opinions, and each individual can choose how public to make their opinions. Further, no one can be "bombed" or otherwise coerced into changing their opinion, as any such social bullying can be made irrelevant by simply rating the intimidator poorly; this removes their opinion (and their coterie's opinion) from your ratings and the ratings of anyone who values your opinion. Those who are rated poorly by many others find their ability to affect them essentially eliminated, meaning that negative social actions become inherently self-limiting.

In a similar way, attempting to increase the popularity of an identity by using other fake identities (a Sybil attack) is futile, since the opinions of the fake identities first have to be valued by others to have any effect. Overall an individual cannot cede trust to anyone else, and is therefore not affected by the opinion of anyone else, unless they choose to do so.

Publish and Subscribe Opinions that an individual makes public are passed effortlessly to their friends and others interested in their opinion. For example, an individual who finds a blog post that they find important has only to rate the story, and it will appear as important to all others interested in their opinion (modified by the opinions of anyone else in their network who also rated it).

Similarly, rating and tagging an individual in a game as a valuable party member (or conversely, as a ninja-looter) makes this information available to everyone who is linked to you, and by extension all those two or more degrees separate from you. A player can thus find out what his or her friends and their friends think of someone before inviting them to join their game, party, or guild.

No Spam Otello also eliminates social spam of the "be my friend" type so often seen on social sites and in games. Having "more friends" whom you don't know is meaningless. Each person links to others because they see

value in them, but what matters is linking to people whose opinions each person finds important, not the number of people in a friends list. Linking is done without a reciprocating responsibility in the person linked to, just as with subscribing to their blog.

Operator Benefits Finally, Otello provides benefits to the game operator in terms of analyzing the social landscape inside the game. This can be used to find the leaders and early adopters and to see how ideas and ratings spread throughout the population, but also to find, localize, and remove those players who introduce cheats or other destabilizing factors into the game.

Potential Issues

As mentioned above, Otello avoids or withstands most of the issues affecting existing rating and reputation systems by making use of the same semantics of trust we use in our daily lives. Tactics like reputation bombing or Sybil attacks are fruitless.

The primary issue still open in the use of Otello is the violation of formerly earned trust. A highly influential and trusted person can turn that influence against others by rating them poorly in a malicious manner. While this is an issue in any form of powerful relationship, in an Otello-enabled environment the spurned users at least have the equal ability to limit the damage by rating the formerly trusted rater negatively, thus eliminating his or her opinion from their view of the world. The violation of anyone's trusts becomes a self-limiting action, as others will quickly stop listening to the violating individual's opinions too.

The other primary potential issue is one of scalability. Otello performs repeated spreading recursive network searches, which can be computationally intensive. Our scale tests indicate that Otello's Java implementation can handle rich networks ranging to several hundred thousand people and items per CPU with sub-second response time. Additional scaling mechanisms can be used as needed.

New Gameplay

Creating a social graph including human relationships and trust sets the stage for new forms of gameplay that will be compelling to a wide range of users. In effect this makes the social landscape – who knows whom, and who is respected by whom – part of the game. Games can be about more than who can hit fastest or hardest; now it's about who is more respected, more loved, or more feared.

Players in a multi-player setting can become known for how they play, both in terms of a reputational score and tags that others apply to their avatar. Upon meeting a person for the first time, one player can assess another based on the opinions of their trusted network, and decide if the other is worth playing with, bringing into a party, or asking for help.

In both single- and multi-player settings, non-player characters (NPCs) also become key actors on the social landscape. Now NPCs are not simply vending machines; they can not only react to what they see a character do, but pass this information on to their fellows. So if you cheat one shopkeeper, others may soon be on the lookout for you – while the head of the rogues' guild might offer you a job.

In other contexts, NPCs and PCs together can deal in rumors and noisy backchannels, with opinions passed along trusted lines. Or in a different setting, PCs and NPCs can attempt to influence public opinion and thus change styles and fashions based on affecting the opinions of those in overlapping trusted networks. Finding the early adopters and influencers in the game population becomes an aspect of gameplay.

NPCs can pass opinions to those in their social network directly, while PCs and NPCs can pass opinions tacitly through chat or other social proximity (belonging to the same guild for example). For example, if a player has a poor opinion of a particular character, there is a chance that this opinion will "rub off" on those with whom he talks or spends time. Thus, without affecting anyone's ability to set their own opinions, players who hang out together or in a particular part of town will all begin to have similar opinions. Social ideas, memes, fads, and fashions are all passed between players and between PCs and NPCs, creating a diverse, interesting social fabric within a game that can be exploited for gameplay purposes.

This has significant benefits for players beyond lifting social play out of just "chatting." Players who have the talent to cultivate the right social connections are as valuable in their own way as those who can swing a sword or shoot a gun. Such players will begin to feel socially competent and connected to their social network in the game, making them feel more immersed in the game and likely to stay around longer.

Games can also explicitly reward players who fill social roles such as being a social hub, a connector between diverse groups, a popular leader, etc. All of this is discoverable from the shape of the social graph even if it is not obvious to the players, and game operators can use the structure of the social graph to award badges, skills, or other rewards for social contribution.

Non-Game Uses

Otello has many uses alongside and apart from games as well. In general, Otello is suitable for use in any situation with a social component where individuals have long-term identities and are likely to interact multiple times. In a marketplace where identities are fluid or most people

interact only once, the 'trusted network' approach of Otello may not be appropriate. In these cases defaulting to community-wide reputation may be the only choice.

However, in many contexts such as online communities or game lobbies, Otello enables users to make better decisions based on the experiences of others. For example in a game lobby, using Otello can help people find new gaming partners from within the reaches of their social network, using opinions they can trust. Having even tenuous information about someone from several friends-of-friends is better than having to guess whether someone is going to cheat, leave early, or be obnoxious without any information at all. Having this level of information is also a powerful disincentive for people to behave badly in games, as now their actions have consequences that follow them socially into future games.

As mentioned above, Otello can also be useful for tracking opinions through a social network. Using Otello, advertisers and others are able to identify the small proportion of any population who consistently introduce new elements that quickly turn viral in the overall population, and track the spread of viral ideas through the population. Thought-leaders, hubs, and social black holes (those who consume information but rarely pass it along) can all be discovered, giving the operator a much better sense not only of the social networks in their world, but of how ideas enter and traverse it.

Otello can also be a useful way for individuals to ask questions without having to throw them out on the web in general. Whether one is asking, "does anyone know a good mechanic?" or "how can I defeat the boss monster on Level 17?" asking this of your network rather than of some general populace is more likely to create positive results.

From others' point of view, only those who are linked to you – who are interested in you, or are friends or fans of those who are – will see your questions. Each user is able to leverage the power of their social network with little effort and without spamming anyone else. Moreover, any answers provided publicly will be seen similarly, only by those in the trusted network of people linked to the person answering. This may help another person who had a similar question but had not yet asked it. This can create a positive social response, much as two overlapping conversations in a party can bring groups together.

There are other in-game and non-game uses for Otello that we are currently exploring. For example, we believe there are applications in the areas of social search and safe chat for parents and kids online. We are also interested in exploring the creation of "opinion clusters," emergent groups whose members share multiple points of agreement. This may be a powerful way to enable more satisfying social and community experiences.

Conclusion

Otello is a unique rating and reputation system modeled on the semantics of trust found in typical human relationships. Unlike existing systems it does not create global reputation values, but instead uses the opinions of those in a network of trust to create distributed, subjective reputation values. We have made every effort to have the technology follow how people act rather than the reverse without giving up computational tractability and performance.

Otello enables new ways to build and make use of one's social network and to make the social graph part of gameplay in a safe and secure fashion. It also creates entirely new forms of gameplay that go far beyond chatting as currently seen in many online social contexts. Players who are adept socially become as valuable as those who are skilled in combat, and the player character's relationships with NPCs takes on new importance.

Finally, Otello enables new forms of discovery about the social graph itself, and the ability to link these back to users of the system for rewarding positive social behavior. There remain many uses for, and new features of, Otello that are not yet understood.

References

Abdul-Rahman, A, and Hailes, S. 2000. Supporting trust in virtual communities. In *Proceedings 3rd Ann. Hawaii Int'l Conference on System Sciences*, vol 6, p. 6007

Douceur, J. 2002. The Sybil Attack. In *Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, Cambridge Massachusetts.

The Elder Scrolls IV: Oblivion. 2007, Bethesda Softworks

Epinions 2004. www.epinions.com/content_4054884484. Retrieved August 8, 2008

Fable. 2004. Microsoft Game Studios

Glass, B. 2008. Designing Your Reputation System. Presentation at IASummit, Miami FL.

Jøsang, A. 1997. Artificial Reasoning with Subjective Logic. In *Proceedings of the Second Australian Workshop on Commonsense Reasoning.*

Resnick, P. and Zeckhauser, R. 2002. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In *The Economics of the Internet and E-Commerce*, Vol. 11. Elsevier Science.

Sid Meier's Pirates!. 1987. MicroProse Software, Inc.

Taylor, B. 2007. Advogato has Failed. www.advogato.org/article/928.html. Retrieved August 8, 2008.

World of Warcraft. 2004. Vivendi Universal