Advanced Features in Bayesian Reputation Systems[⋆]

Audun Jøsang[1] and Walter Quattrociocchi[2]

[1] University of Oslo - UNIK Graduate Center
josang @ matnat.uio.no

[2] Institute of Cognitive Sciences and Technologies (ISTC), Italian National Research Council
walter.quattrociocchi @ istc.cnr.it

Abstract. Bayesian reputation systems are quite flexible and can relatively easily be adapted to different types of applications and environments. The purpose of this paper is to provide a concise overview of the rich set of features that characterizes Bayesian reputation systems. In particular we demonstrate the importance of base rates during bootstrapping, for handling rating scarcity and for expressing long term trends.

1 Introduction

Reputation systems are used to collect and analyse information about the performance of service entities with the purpose of computing reputation scores for service objects and service entities. A fundamental assumption of reputation systems is that reputation scores can help predict the future performance of the respective entities and thereby reduce uncertainty of relying parties during the decision making processes [5]. The idea is that transactions with reputable parties are likely to result in more favourable outcomes than transactions with disreputable parties.

In the case of centralised reputation systems, ratings are collected centrally and the computed reputation scores are published online. In the case of distributed reputation systems, relying parties must discover and collect ratings from other members in the community who have had prior experience with the service entity. The relying party must then analyse the collected information to derive a private reputation score. Two fundamental elements of reputation systems are:

1. *Communication protocols* that allow participants to provide ratings about service entities to the reputation centre, as well as to obtain reputation scores of potential entities from the reputation system.
2. *A reputation computation engine* used by the reputation system to derive reputation scores for each participant, based on received ratings, and possibly also on other information.

This paper focuses on the reputation computation engines, and in particular on Bayesian computational engines. Binomial and multinomial Bayesian reputation systems have been proposed and studied e.g. in [1, 3, 4, 6].

[⋆] In the proceedings of the 6th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS'09), Linz, Austria, August-September, 2009.

Many different reputation systems, including computation engines, have been proposed in the literature, and we do not intend to provide a survey or comparison in this paper, but simply refer to [2]. The purpose of this paper is concisely describe the advanced features of Bayesian reputation systems, some of which have been described previously and some of which are presented here.

2 Mathematics of Bayesian Reputation Systems

2.1 Computing Reputation Scores

Binomial reputation systems allow ratings to be expressed with two values, as either positive (e.g. *Good*) or negative (e.g. *Bad*). Multinomial reputation systems allow the possibility of providing ratings in different discrete levels such as e.g. *mediocre* - *bad* - *average* - *good* - *excellent*.

Binomial Reputation Scores. Binomial Bayesian reputation systems apply to the binary state space $\{\text{Bad, Good}\}$ which reflect a corresponding performance of a service entity. The Beta distributions is a continuous distribution functions over a binary state space indexed by the two parameters α and β. The beta PDF denoted by $\mathrm{Beta}(p \,|\, \alpha, \beta)$ can be expressed using the gamma function Γ as:

$$\mathrm{Beta}(p \,|\, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

where $0 \leq p \leq 1$ and $\alpha, \beta > 0$, with the restriction that the probability variable $p \neq 0$ if $\alpha < 1$, and $p \neq 1$ if $\beta < 1$. The probability expectation value of the Beta distribution is given by:

$$\mathrm{E}(p) = \alpha/(\alpha+\beta). \tag{2}$$

Binomial reputation is computed by statistical updating of the Beta PDF. More specifically, the *a posteriori* (i.e. the updated) reputation is continuously computed by combining the *a priori* (i.e. previous) reputation with every new rating. It is the expectation value of Eq.(2) that is used to represent the reputation score. The Beta distribution itself only provides the underlying statistical foundation, and is otherwise not used in the reputation system.

Before receiving any ratings, the *a priori* distribution is the Beta PDF with $\alpha = Wa$ and $\beta = W(1-a)$ where W denotes the non-informative prior weight and a denotes the base rate of the outcome ”Good”. The default base rate is $a = 0.5$. Normally $W = 2$, which with the default base rate produces a uniform *a priori* distribution. Then after observing r ”Good” and s ”Bad” outcomes, the *a posteriori* distribution is the Beta PDF with $\alpha = r + Wa$ and $\beta = s + W(1-a)$. By using Eq.(2) the reputation score S is:

$$S = (r + Wa)/(r + s + W). \tag{3}$$

This score should be interpreted as the probability that the next experience with the service entity will be ”Good”.

Multinomial Reputation Scores. Multinomial Bayesian reputation systems allow ratings to be provided over k different levels which can be considered as a set of k disjoint elements. Let this set be denoted as $\Lambda = \{L_1, \ldots L_k\}$, and assume that ratings are provided as votes on the elements of Λ. This leads to a Dirichlet probability density function over the k-component random probability variable $\vec{p}(L_{\,i}),\ \ i = 1 \ldots k$ with sample space $[0,1]^k$, subject to the simple additivity requirement $\sum_{i=1}^{k} \vec{p}(L_i) = 1$.

The Dirichlet distribution with prior captures a sequence of observations of the k possible outcomes with k positive real rating parameters $\vec{r}(L_{\,i}),\ i = 1 \ldots k$, each corresponding to one of the possible levels. In order to have a compact notation we define a vector $\vec{p} = \{\vec{p}(L_i) \mid 1 \leq i \leq k\}$ to denote the k-component probability variable, and a vector $\vec{r} = \{r_i \mid 1 \leq i \leq k\}$ to denote the k-component rating variable.

In order to distinguish between the *a priori* default base rate, and the *a posteriori* ratings, the Dirichlet distribution must be expressed with prior information represented as a base rate vector $\vec{a}$ over the state space.

$$\text{Dirichlet}(\vec{p} \mid \vec{r}, \vec{a}) = \frac{\Gamma\left(\sum_{i=1}^{k}(\vec{r}(L_i)+W\vec{a}(L_i))\right)}{\prod_{i=1}^{k}\Gamma(\vec{r}(L_i)+W\vec{a}(L_i))} \prod_{i=1}^{k} \vec{p}(L_i)^{(\vec{r}(L_i)+W\vec{a}(L_i)-1)} \ ,$$

$$\text{where} \begin{cases} \sum\limits_{i=1}^{k} \vec{p}(L_i) = 1 \\ \\ \vec{p}(L_i) \geq 0, \forall i \end{cases} \text{and} \begin{cases} \sum\limits_{i=1}^{k} \vec{a}(L_i) = 1 \\ \\ \vec{a}(L_i) > 0, \forall i\ . \end{cases} \tag{4}$$

It can be seen that Eq.(4) simply is a generalisation of Eq.(1). Similarly to the binomial case, the multinomial reputation score $\vec{S}$ is the vector of probability expectation values of the k random probability variables expressed as:

$$\vec{S}(L_i) = \mathrm{E}(\vec{p}(L_i) \mid \vec{r}, \vec{a}) = \frac{\vec{r}(L_i) + W\vec{a}(L_i)}{W + \sum_{i=1}^{k} \vec{r}(L_i)}\ . \tag{5}$$

The non-informative prior weight W will normally be set to $W = 2$ when a uniform distribution over binary state spaces is assumed. Selecting a larger value for W would result in new observations having less influence. over the Dirichlet distribution, and can in fact represent specific *a priori* information provided by a domain expert or by another reputation system.

2.2 Collecting Ratings

Assume k different discrete rating levels. This translates into having a state space of cardinality k. For binomial reputation systems $k = 2$ and the rating levels are ”Bad” and ”Good”. For multinomial reputation system $k > 2$ and any corresponding set of suitable rating levels can be used. Let the rating level be indexed by i. The aggregate ratings for a particular agent y are stored as a cumulative vector, expressed as:

$$\vec{R}_y = (\vec{R}_y(L_i) \mid i = 1 \ldots k)\ . \tag{6}$$

The simplest way of updating a rating vector as a result of a new rating is by adding the newly received rating vector $\vec{r}$ to the previously stored vector $\vec{R}$. The case when old ratings are aged is described in Sec.2.3.

Each new discrete rating of agent y by an agent x takes the form of a trivial vector $\vec{r}_y^{\,x}$ where only one element has value 1, and all other vector elements have value 0. The index i of the vector element with value 1 refers to the specific rating level.

2.3 Aggregating Ratings with Aging

Ratings may be aggregated by simple addition of the components (vector addition).

Agents (and in particular human agents) may change their behaviour over time, so it is desirable to give relatively greater weight to more recent ratings. This can be achieved by introducing a longevity factor $\lambda \in [0, 1]$, which controls the rapidity with which old ratings are aged and discounted as a function of time. With $\lambda = 0$, ratings are completely forgotten after a single time period. With $\lambda = 1$, ratings are never forgotten.

Let new ratings be collected in discrete time periods. Let the sum of the ratings of a particular agent y in period t be denoted by the vector $\vec{r}_{y,t}$. More specifically, it is the sum of all ratings $\vec{r}_y^{x}$ of agent y by other agents x during that period, expressed by:

$$\vec{r}_{y,t} = \sum_{x \in M_{y,t}} \vec{r}_y^{x} \tag{7}$$

where $M_{y,t}$ is the set of all agents who rated agent y during period t.

Let the total accumulated ratings (with aging) of agent y after the time period t be denoted by $\vec{R}_{y,t}$. Then the new accumulated rating after time period $t + 1$ can be expressed as:

$$\vec{R}_{y,(t+1)} = \lambda \cdot \vec{R}_{y,t} + \vec{r}_{y,(t+1)}, \text{ where } 0 \leq \lambda \leq 1 \ . \tag{8}$$

Eq.(8) represents a recursive updating algorithm that can be executed once every period for all agents, or alternatively in a discrete fashion for each agent for example after each rating. Assuming that new ratings are received between time t and time $t+n$ periods, then the new rating can be computed as:

$$\vec{R}_{y,(t+n)} = \lambda^n \cdot \vec{R}_{y,t} + \vec{r}_{y,(t+n)} \ , \quad 0 \leq \lambda \leq 1. \tag{9}$$

2.4 Convergence Values for Reputation Scores

The recursive algorithm of Eq.(8) makes it possible to compute convergence values for the rating vectors, as well as for reputation scores. Assuming that a particular agent receives the same ratings every period, then Eq.(8) defines a geometric series. We use the well known result of geometric series:

$$\sum_{j=0}^{\infty} \lambda^j = \frac{1}{1-\lambda} \quad \text{for } -1 < \lambda < 1 \ . \tag{10}$$

Let $\vec{e}_y$ represent a constant rating vector of agent y for each period. The Total accumulated rating vector after an infinite number of periods is then expressed as:

$$\vec{R}_{y,\infty} = \frac{\vec{e}_y}{1-\lambda}, \text{ where } 0 \leq \lambda < 1 \; . \tag{11}$$

Eq.(11) shows that the longevity factor λ determines the convergence values for the accumulated rating vector according to Eq.(8). In general it will be impossible for components of the accumulated rating vector to reach infinity, which makes it impossible for the score vector components to cover the whole range $[0, 1]$. However, entities that provide maximum quality services over a long time period wold naturally expect to get the highest possible reputation score. An intuitive interpretation of this expectation is that each long standing entity should have its own personal base rate which is determined as a function of the entity's total history, or at least a large part of it. This approach is used in the next section to include individual base rates.

2.5 Individual Base Rates

A base rate normally expresses the average in a population or domain. Here we will compute individual base rates from a ”population” consisting of individual performances over a series of time periods. The individual base rate for entity y at time t will be denoted as $\vec{a}_{y,t}$. It will be based on individual evidence vectors denoted as $\vec{Q}_{y,t}$.

Let $\vec{a}$ denote the community base rate as usual. Then the individual base rate for entity y at time t can be computed similarly to Eq.(5) as:

$$\vec{a}_{y,t}(L_i) = \frac{\vec{Q}_{y,t}(L_i) + W\vec{a}(L_i)}{W + \sum_{i=1}^{k} \vec{Q}_{y,t}(L_i)} \; . \tag{12}$$

Reputation scores can be computed as normal with Eq.(5), except that the community base rate $\vec{a}$ is replaced with the individual base rate $\vec{a}_{y,t}$ of Eq.(12). It can be noted that the individual base rate $\vec{a}_{y,t}$ is partly a function of the community base rate $\vec{a}$, which thereby constitutes a two-level base rate model.

The components of the reputation score vector computed with Eq.(5) based on the individual base rate of Eq.(12) can theoretically be arbitrarily close to 0 or 1 with any longevity factor and any community base rate.

The simplest alternative to consider is to let the individual base rate for each entity be a function of the entity's total history. A second similar alternative is to let the individual base rate be computed as a function of an entity's performance over a very long sliding time window. A third alternative is to define an additional high longevity factor for base rates that is much closer to 1 than the common longevity factor λ. The formalisms for these three alternatives are briefly described below.

Total History Base Rate. The total evidence vector $\vec{Q}_{y,t}$ for entity y used to compute the individual base rate at time period t is expressed as:

$$\vec{Q}_{y,t} = \sum_{j=1}^{t} \vec{R}_{y,j} \tag{13}$$

Sliding Time Window Base Rate. The evidence vector $\vec{Q}_{y,t}$ for computing an entity y's individual base rate at time period t is expressed as:

$$\vec{Q}_{y,t} = \sum_{j=u}^{t} \vec{R}_{y,j} \quad \text{where Window Size } = (t - u) \ . \tag{14}$$

The Window Size would normally be a constant, but could also be dynamic. In case e.g. $u = 1$ the Window Size would be increasing and be equal to t, which also would make this alternative equivalent to the total history alternative described above.

High Longevity Factor Base Rate. Let λ denote the normal longevity factor. A high longevity factor λ_{H} can be defined where $\lambda_{\text{H}} > \lambda$. The evidence vector $\vec{Q}_{y,t}$ for computing an entity y's individual base rate at time period t is computed as:

$$\vec{Q}_{y,t} = \lambda_{\text{H}} \cdot \vec{Q}_{y,(t-1)} + \vec{r}_{y,t}, \text{ where } \lambda < \lambda_{\text{H}} \leq 1 \ . \tag{15}$$

In case $\lambda_{\text{H}} = 1$ this alternative would be equivalent to the total history alternative described above. The high longevity factor makes ratings age much slower than the regular longevity factor.

2.6 Reputation Representation

Reputation can be represented in different forms. We will here illustrate reputation as *multinomial probability scores*, and as *point estimates*. Each form will be described in turn below.

Multinomial Probability Representation. The most natural is to define the reputation score as a function of the probability expectation values of each element in the state space. The expectation value for each rating level can be computed with Eq.(5).

Let $\vec{R}$ represent a target agent's aggregate ratings. Then the vector $\vec{S}$ defined by:

$$\vec{S}_y : \left(\vec{S}_y(L_i) = \frac{\vec{R}^*_y(L_i) + W\vec{a}(L_i)}{W + \sum_{j=1}^{k} \vec{R}^*_y(L_j)}; \mid i = 1 \ldots k \right) \ . \tag{16}$$

is the corresponding multinomial probability reputation score. As already stated, $W = 2$ is the value of choice, but a larger value for the constant W can be chosen if a reduced influence of new evidence over the base rate is required.

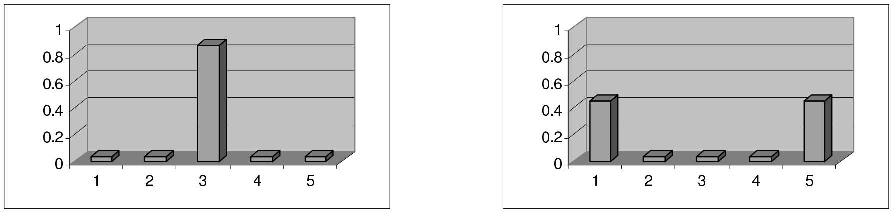
The reputation score $\vec{S}$ can be interpreted like a multinomial probability measure as an indication of how a particular agent is expected to behave in future transactions. It can easily be verified that

$$\sum_{i=1}^{k} S(L_i) = 1 \ . \tag{17}$$

The multinomial reputation score can for example be visualised as columns, which would clearly indicate if ratings are polarised. Assume for example 5 levels:

$$\text{Discrete rating levels:} \begin{cases} L_1 : \text{Mediocre,} \\ L_2 : \text{Bad,} \\ L_3 : \text{Average,} \\ L_4 : \text{Good,} \\ L_5 : \text{Excellent.} \end{cases} \tag{18}$$

We assume a default base rate distribution. Before any ratings have been received, the multinomial probability reputation score will be equal to $1/5$ for all levels. Let us assume that 10 ratings are received. In the first case, 10 *average* ratings are received, which translates into the multinomial probability reputation score of Fig.1.a. In the second case, 5 mediocre and 5 excellent ratings are received, which translates into the multinomial probability reputation score of Fig.1.b.



(a) 10 average ratings

(b) 5 mediocre and 5 excellent ratings

Fig. 1. Illustrating score difference resulting from average and polarised ratings

With a binomial reputation system, the difference between these two rating scenarios would not have been visible.

Point Estimate Representation. While informative, the multinomial probability representation can require considerable space to be displayed on a computer screen. A more compact form can be to express the reputation score as a single value in some predefined interval. This can be done by assigning a point value ν to each rating level i, and computing the normalised weighted point estimate score σ.

Assume e.g. k different rating levels with point values evenly distributed in the range [0,1], so that $\nu(L_i) = \frac{i-1}{k-1}$. The point estimate reputation score is then:

$$\sigma = \sum_{i=1}^{k} \nu(L_i) S(L_i) \ . \tag{19}$$

However, this point estimate removes information, so that for example the difference between the average ratings and the polarised ratings of Fig.1.a and Fig.1.b is no longer visible. The point estimates of the reputation scores of Fig.1.a and Fig.1.b are both 0.5, although the ratings in fact are quite different. A point estimate in the range [0,1] can be mapped to any range, such as 1-5 stars, a percentage or a probability.

2.7 Dynamic Community Base Rates

Bootstrapping a reputation system to a stable and conservative state is important. In the framework described above, the base rate distribution $\vec{a}$ will define initial default reputation for all agents. The base rate can for example be evenly distributed, or biased towards either a negative or a positive reputation. This must be defined by those who set up the reputation system in a specific market or community.

Agents will come and go during the lifetime of a market, and it is important to be able to assign new members a reasonable base rate reputation. In the simplest case, this can be the same as the initial default reputation used during during bootstrap.

However, it is possible to track the average reputation score of the whole community, and this can be used to set the base rate for new agents, either directly or with a certain additional bias.

Not only new agents, but also existing agents with a standing track record can get the dynamic base rate. After all, a dynamic community base rate reflects the whole community, and should therefore be applied to all the members of that community. The aggregate reputation vector for the whole community at time t can be computed as:

$$\vec{R}_{M,t} = \sum_{y_j \in M} \vec{R}_{y,t} \tag{20}$$

This vector then needs to be normalised to a base rate vector as follows:

Definition 1 (Community Base Rate). *Let $\vec{R}_{M,t}$ be an aggregate reputation vector for a whole community, and let $\vec{S}_{M,t}$ be the corresponding multinomial probability reputation vector which can be computed with Eq.(16). The community base rate as a function of existing reputations at time $t+1$ is then simply expressed as the community score at time t:*

$$\vec{a}_{M,(t+1)} = \vec{S}_{M,t}. \tag{21}$$

The base rate vector of Eq.(21) can be given to every new agent that joins the community. In addition, the community base rate vector can be used for every agent every time their reputation score is computed. In this way, the base rate will dynamically reflect the quality of the market at any one time.

If desirable, the base rate for new agents can be biased in either negative or positive direction in order to make it harder or easier to enter the market.

When base rates are a function of the community reputation, the expressions for convergence values with constant ratings can no longer be defined with Eq.(11), and will instead converge towards the average score from all the ratings.

2.8 Continuous Ratings

It is common that the subject matter to be rated is measured on a continuous scale, such as time, throughput or relative ranking, to name a few examples. Even when it is natural to provide discrete ratings, it may be difficult to express that something is strictly good or average, so that combinations of discrete ratings, such as *"average-to-good"* would better reflect the rater's opinion. Such ratings can then be considered continuous. To handle this, it is possible to use a fuzzy membership function to convert a continuous rating into a binomial or multinomial rating. For example with five rating levels the sliding window function can be illustrated as in Fig.2. The continuous q-value determines the r-values for that level.

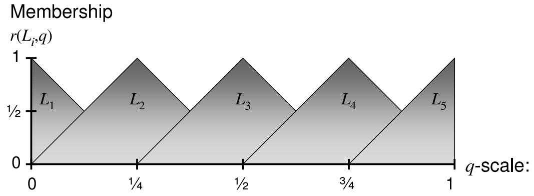


Fig. 2. Fuzzy triangular membership functions

3 Simple Scenario Simulation

A simple scenario can be used to illustrate the performance of a multinomial reputation system that uses some of the features described above. Let us assume that agent A and agent B receive the following ratings over 70 rounds or time periods.

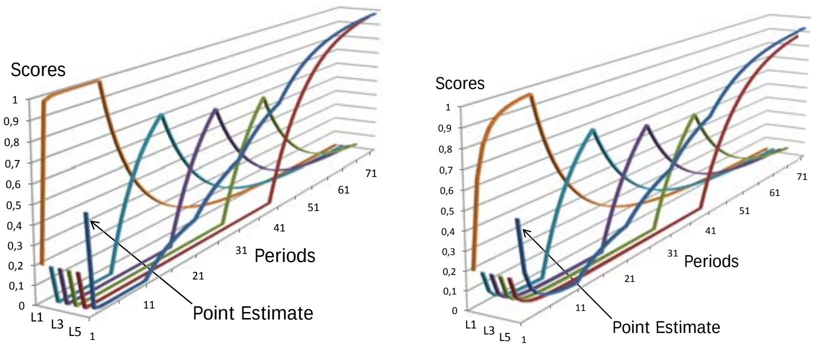
Sequence	Agent A	Agent B
Periods 1 - 10	10 × L1 ratings in each period	1 × L1 rating in each period
Periods 11-20	10 × L2 ratings in each period	1 × L2 rating in each period
Periods 21-30	10 × L3 ratings in each period	1 × L3 rating in each period
Periods 31-40	10 × L4 ratings in each period	1 × L4 rating in each period
Periods 41-70	30 × L5 ratings in each period	3 × L5 ratings in each period

Table 1. Sequence of ratings

The longevity of ratings is set to $\lambda = 0.9$ and the individual base rate is computed with the high longevity approach described in Sec.2.5 with high longevity factor for the base rate set to $\lambda_{\mathrm{H}} = 0.999$. For simplicity in this example the community base rate is assumed to be fixed during the 70 rounds, expressed by $a(\mathrm{L1}) = a(\mathrm{L2}) = a(\mathrm{L3}) = a(\mathrm{L4}) = a(\mathrm{L5}) = 0.2$. Fig.3 illustrates the evolution of the scores of Agent A and Agent B during the period.

The scores for both agents start with the community base rate, and then vary as a function of the received ratings. Both agents have an initial point estimate of 0.5.

The scores for Agent B in Fig.3.b are similar in trend but less articulated than that of agent A in Fig.3.a, because agent B receives equal but less frequent ratings. The final score of agent B is visibly lower than 1 because the relatively low number of ratings is insufficient for driving the individual base rate very close to 1. Thanks to the community



(a) Scores for agent A

(b) Scores for agent B

Fig. 3. Score evolution for agents A and B

base rate, all new agents in a community will have a meaningful initial score. In case of rating scarcity, an agents score will initially be determined by the community base rate, with the individual base rate dominating as soon as some ratings have been received,

4 Conclusion

Bayesian reputation systems have a solid basis in classical statistics which make them both sound and simple to adapt to various contexts. Bayesian reputation is compatible with opinions of subjective logic, thereby allowing computational trust to be combined with reputation models. This paper has shown the great flexibility of binomial and multinomial Bayesian reputation systems.

References

1. A. Jøsang and R. Ismail. The Beta Reputation System. In *Proceedings of the 15th Bled Electronic Commerce Conference*, June 2002.
2. A. Jøsang, R. Ismail, and C. Boyd. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems*, 43(2):618–644, 2007.
3. A. Jøsang and Haller J. Dirichlet Reputation Systems. In *The Proceedings of the International Conference on Availability, Reliability and Security (ARES 2007)*, Vienna, Austria, April 2007.
4. A. Jøsang, X. Luo, and X. Chen. Continuous Ratings in Discrete Bayesian Reputation Systems. In *The Proceedings of the Joint iTrust and PST Conferences on Privacy, Trust Management and Security (IFIPTM 2008)*, Trondheim, June 2008.
5. W. Quattrociocchi, M. Paolucci, and R. Conte. Dealing with Uncertainty: Simulating Reputation in an Ideal Marketplace. In *Proceedings of the 2008 Trust Workshop, at the 7th Int. Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS)*, 2008.
6. A. Withby, A. Jøsang, and J. Indulska. Filtering Out Unfair Ratings in Bayesian Reputation Systems. *The Icfain Journal of Management Research*, 4(2):48–64, 2005.