

# Combining Trust and Reputation Management for Web-Based Services<sup>\*</sup>

Audun Jøsang<sup>1</sup>, Touhid Bhuiyan<sup>2</sup>, Yue Xu<sup>2</sup>, and Clive Cox<sup>3</sup>

<sup>1</sup> UniK Graduate School, University of Oslo, Norway  
josang @ unik.no

<sup>2</sup> Faculty of Information Technology, QUT, Brisbane, Australia.  
t.bhuiyan @ qut.edu.au                      yue.xu @ qut.edu.au

<sup>3</sup> Rumble.com, Cambridge, England  
clive.cox @ rumble.com

**Abstract.** Services offered and provided through the Web have varying quality, and it is often difficult to assess the quality of a services before accessing and using it. Trust and reputation systems can be used in order to assist users in predicting and selecting the best quality services. This paper describes how Bayesian reputation systems can be combined with trust modeling based on subjective logic to provide an integrated method for assessing the quality of online services. This will not only assist the user's decision making, but will also provide an incentive for service providers to maintain high quality, and can be used as a sanctioning mechanism to discourage deceptive and low quality services.

## 1 Introduction

Online trust and reputation systems are emerging as important decision support tools for selecting online services and for assessing the risk of accessing them. We have previously proposed and studied Bayesian reputation systems [6–8, 12] and trust models based on subjective logic [4, 5, 10]. Binomial Bayesian reputation systems normally take ratings expressed in a discrete binary form as either positive (e.g. *good*) or negative (e.g. *bad*). Multinomial Bayesian reputation systems allow the possibility of providing ratings with discrete graded levels such as e.g. *mediocre - bad - average - good - excellent* [8]. It is also possible to use continuous ratings in both binomial and multinomial reputation systems [9]. Multinomial models have the advantage that scores can distinguish between the case of polarised ratings (e.g. a combination of strictly good and bad ratings) and the case of only average ratings.

Trust models based on subjective logic are directly compatible with Bayesian reputation systems because a bijective mapping exists between their respective trust and reputation representations. This provides a powerful basis for combining trust and reputation systems for assessing the quality of online services.

---

<sup>\*</sup> Appears in the Proceedings of the 5th International Conference on Trust, Privacy & Security in Digital Business (TrustBus2008), Turin, September 2008.

A general characteristic of reputation systems is that they provide global reputation scores, meaning that all the members in a community will see the same reputation score for a particular agent. On the other hand, trust systems can in general be used to derive local and subjective measures of trust, meaning that different agents can derive different trust in the same entity. Another characteristic of trust systems is that they can analyse multiple hops of trust transitivity. Reputation systems on the other hand normally compute scores based on direct input from members in the community which is not based on transitivity. Still there are systems that have characteristics of being both a reputation system and a trust system. The matrix below shows examples of the possible combinations of local and global scores, and trust transitivity or not.

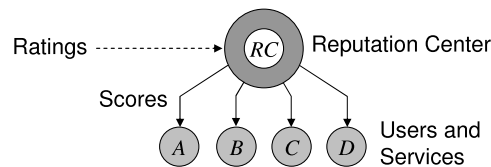
**Table 1.** Possible combinations of local/global scores and transitivity/no transitivity

	Private Scores	Public Scores
Transitivity	Trust systems, e.g. Rumble.com	Public trust systems, e.g. PageRank
No Transitivity	Private reputation systems, e.g. customer feedback analysis	Reputation systems, e.g. eBay.com

In this paper we describe a framework for combining these forms of trust and reputation systems. Because Bayesian reputation systems are directly compatible with trust systems based on subjective logic, they can be seamlessly integrated. This provides a powerful and flexible basis for online trust and reputation management.

## 2 The Dirichlet Reputation System

Reputation systems collect ratings about users or service providers from members in a community. The reputation centre is then able to compute and publish reputation scores about those users and services. Fig.1 illustrates a reputation centre where the dotted arrow indicate ratings and the solid arrows indicate reputation scores about the users.



**Fig. 1.** Simple reputation system

Multinomial Bayesian systems are based on computing reputation scores by statistical updating of Dirichlet Probability Density Functions (PDF), which therefore are called Dirichlet reputation systems [8, 9]. The *a posteriori* (i.e. the updated) reputation score is computed by combining the *a priori* (i.e. previous) reputation score with new ratings.

In Dirichlet reputation systems agents are allowed to rate others agents or services with any level from a set of predefined rating levels, and the reputation scores are not static but will gradually change with time as a function of the received ratings. Initially, each agent's reputation is defined by the base rate reputation which is the same for all agents. After ratings about a particular agent have been received, that agent's reputation will change accordingly.

Let there be  $k$  different discrete rating levels. This translates into having a state space of cardinality  $k$  for the Dirichlet distribution. Let the rating level be indexed by  $i$ . The aggregate ratings for a particular agent are stored as a cumulative vector, expressed as:

$$\vec{R} = (\vec{R}(L_i) \mid i = 1 \dots k) . \quad (1)$$

This vector can be computed recursively and can take factors such as longevity and community base rate into account [8]. The most direct form of representing a reputation score is simply the aggregate rating vector  $\vec{R}_y$  which represents all relevant previous ratings. The aggregate rating of a particular level  $i$  for agent  $y$  is denoted by  $\vec{R}_y(L_i)$ .

For visualisation of reputation scores, the most natural is to define the reputation score as a function of the probability expectation values of each rating level. Before any ratings about a particular agent  $y$  have been received, its reputation is defined by the common base rate  $\vec{a}$ . As ratings about a particular agent are collected, the aggregate ratings can be computed recursively [8, 9] and the derived reputation scores will change accordingly. Let  $\vec{R}$  represent a target agent's aggregate ratings. Then the vector  $\vec{S}$  defined by:

$$\vec{S}_y : \left( \vec{S}_y(L_i) = \frac{\vec{R}_y(L_i) + C\vec{a}(L_i)}{C + \sum_{j=1}^k \vec{R}_y(L_j)}; \mid i = 1 \dots k \right) . \quad (2)$$

is the corresponding multinomial probability reputation score. The parameter  $C$  represents the non-informative prior weight where  $C = 2$  is the value of choice, but larger value for the constant  $C$  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

$$\sum_{i=1}^k \vec{S}(L_i) = 1 . \quad (3)$$

While informative, the multinomial probability representation can require considerable space on a computer screen because multiple values must be visualised. 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  $\nu$  to each rating level  $L_i$ , and computing the normalised weighted point estimate score  $\sigma$ .

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

reputation  $\vec{R}$  is then:

$$\sigma = \sum_{i=1}^k \nu(L_i) \vec{S}(L_i) . \quad (4)$$

A point estimate in the range [0,1] can be mapped to any range, such as 1-5 stars, a percentage or a probability.

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 over all rating levels, or biased towards either negative or a positive rating levels. This must be defined when setting up the reputation system in a specific market or community.

As an example we consider five discrete rating levels, and the following sequence of ratings:

Periods 1 - 10: L1 Mediocre

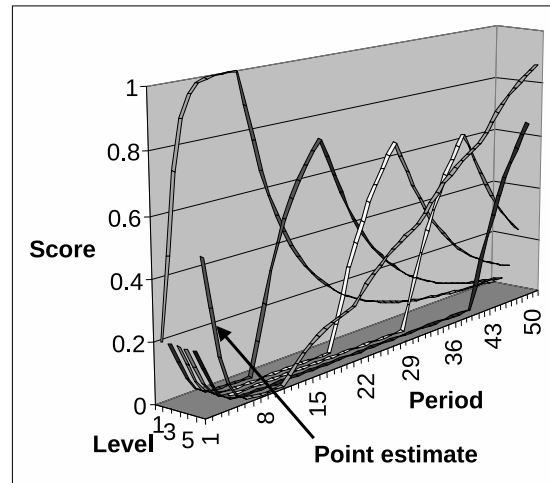
Periods 11 - 20: L2 Bad

Periods 21 - 30: L3 Average

Periods 31 - 40: L4 Good

Periods 41 - 50: L5 Excellent

The longevity factor is  $\lambda = 0.9$ , and the base rate is dynamic [8, 9]. The evolution of the scores of each level as well as the point estimate are illustrated in Fig.2.



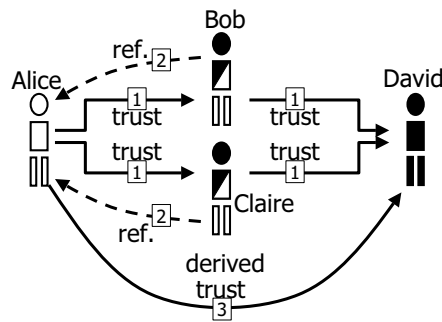
**Fig. 2.** Scores and point estimate during a sequence of varying ratings

In Fig.2 the multinomial reputation scores change abruptly between each sequence of 10 periods. The point estimate first drops as the score for L1 increases during the first 10 periods. After that the point estimate increases relatively smoothly during the subsequent 40 periods. Assuming a dynamic base rate and an indefinite series of L5 (Excellent) ratings, the point estimate will eventually converge to 1.

### 3 Trust Models Based on Subjective Logic

Subjective logic[1–3] is a type of probabilistic logic that explicitly takes uncertainty and belief ownership into account. Arguments in subjective logic are subjective opinions about states in a state space. A binomial opinion applies to a single proposition, and can be represented as a Beta distribution. A multinomial opinion applies to a collection of propositions, and can be represented as a Dirichlet distribution.

Subjective logic defines a trust metric called *opinion* denoted by  $\omega_X^A = (\vec{b}, u, \vec{a})$ , which expresses the relying party  $A$ 's belief over a state space  $X$ . Here  $\vec{b}$  represents belief masses over the states of  $X$ , and  $u$  represent uncertainty mass where  $\vec{b}, u \in [0, 1]$  and  $\sum \vec{b} + u = 1$ . The vector  $\vec{a} \in [0, 1]$  represents the base rates over  $X$ , and is used for computing the probability expectation value of a state  $x$  as  $E(x) = \vec{b}(x) + \vec{a}(x)u$ , meaning that  $\vec{a}$  determines how uncertainty contributes to  $E(x)$ . Binomial opinions are expressed as  $\omega_x^A = (b, d, u, a)$  where  $d$  denotes disbelief in  $x$ . When the statement  $x$  for example says “*David is honest and reliable*”, then the opinion can be interpreted as reliability trust in David. As an example, let us assume that Alice needs to get her car serviced, and that she asks Bob to recommend a good car mechanic. When Bob recommends David, Alice would like to get a second opinion, so she asks Claire for her opinion about David. This situation is illustrated in fig. 3 below where the indexes on arrows indicates the order in which they are formed.



**Fig. 3.** Deriving trust from parallel transitive chains

When trust and referrals are expressed as subjective opinions, each transitive trust path  $Alice \rightarrow Bob \rightarrow David$ , and  $Alice \rightarrow Claire \rightarrow David$  can be computed with the *transitivity operator*<sup>4</sup>, where the idea is that the referrals from Bob and Claire are discounted as a function Alice’s trust in Bob and Claire respectively. Finally the two paths can be combined using the cumulative or averaging fusion operator. These operators form part of *Subjective Logic* [2, 3], and semantic constraints must be satisfied in order for the transitive trust derivation to be meaningful [10]. Opinions can be uniquely mapped to beta PDFs, and in this sense the fusion operator is equivalent to Bayesian updating. This model is thus both belief-based and Bayesian.

<sup>4</sup> Also called the discounting operator

A trust relationship between  $A$  and  $B$  is denoted as  $[A:B]$ . The transitivity of two arcs is denoted as “:” and the fusion of two parallel paths is denoted as “ $\diamond$ ”. The trust network of Fig.3 can then be expressed as:

$$[A, D] = ([A, B] : [B, D]) \diamond ([A, C] : [C, D]) \quad (5)$$

The corresponding transitivity operator for opinions denoted as “ $\otimes$ ” and the corresponding fusion operator as “ $\oplus$ ”. The mathematical expression for combining the opinions about the trust relationships of Fig.3 is then:

$$\omega_D^A = (\omega_B^A \otimes \omega_D^B) \oplus (\omega_C^A \otimes \omega_D^C) \quad (6)$$

Arbitrarily complex trust networks can be analysed with TNA-SL which consists of a network exploration method combined with trust analysis based on subjective logic [4, 5]. The method is based on simplifying complex trust networks into a directed series-parallel graph (DSPG) before applying subjective logic calculus.

#### 4 Combining Trust and Reputation

A bijective mapping can be defined between multinomial reputation scores and opinions, which makes it possible to interpret these two mathematical representations as equivalent. The mapping can symbolically be expressed as:

$$\omega \longleftrightarrow \vec{R} \quad (7)$$

This equivalence which is presented with proof in [3] is expressed as:

##### Theorem 1. Equivalence Between Opinions and Reputations

Let  $\omega = (\vec{b}, u, \vec{a})$  be an opinion, and  $\vec{R}$  be a reputation, both over the same state space  $X$  so that the base rate  $\vec{a}$  also applies to the reputation. Then the following equivalence holds [3]:

For  $u \neq 0$ :

$$\left\{ \begin{array}{l} \vec{b}(x_i) = \frac{\vec{R}(x_i)}{C + \sum_{i=1}^k \vec{R}(x_i)} \\ u = \frac{C}{C + \sum_{i=1}^k \vec{R}(x_i)} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \vec{R}(x_i) = \frac{C \vec{b}(x_i)}{u} \\ u + \sum_{i=1}^k \vec{b}(x_i) = 1 \end{array} \right. \quad (8)$$

For  $u = 0$ :

$$\left\{ \begin{array}{l} \vec{b}(x_i) = \eta(x_i) \\ u = 0 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \vec{R}(x_i) = \eta(x_i) \sum_{i=1}^k \vec{R}(x_i) = \eta(x_i) \infty \\ \sum_{i=1}^k m(x_i) = 1 \end{array} \right. \quad (9)$$

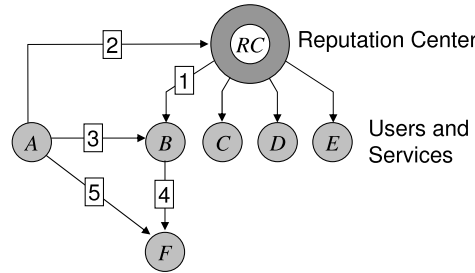
The case  $u = 0$  reflects an infinite amount of aggregate ratings, in which case the parameter  $\eta$  determines the relative proportion of infinite ratings among the rating levels. In case  $u = 0$  and  $\eta(x_i) = 1$  for a particular rating level  $x_i$ , then  $\vec{R}(x_i) = \infty$  and all the other rating parameters are finite. In case  $\eta(x_i) = 1/k$  for all  $i = 1 \dots k$ , then all the rating parameters are equally infinite. As already indicated, the non-informative prior weight is normally set to  $C = 2$ .

Multinomial aggregate ratings can be used to derive binomial trust in the form of an opinion. This is done by first converting the multinomial ratings to binomial ratings according to Eq.(10) below, and then to apply Theorem 1.

Let the multinomial reputation model have  $k$  rating levels  $x_i; i = 1, \dots, k$ , where  $\vec{R}(x_i)$  represents the ratings on each level  $x_i$ , and let  $\sigma$  represent the point estimate reputation score from Eq.(4). Let the binomial reputation model have positive and negative ratings  $r$  and  $s$  respectively. The derived converted binomial rating parameters  $(r, s)$  are given by:

$$\begin{cases} r = \sigma \sum_{i=1}^k \vec{R}_y(x_i) \\ s = \sum_{i=1}^k \vec{R}_y(x_i) - r \end{cases} \quad (10)$$

With the equivalence of Theorem 1 it is possible to analyse trust networks based on both trust relationships and reputation scores. Fig.4 illustrates a scenario where agent  $A$  needs to derive a measure of trust in agent  $F$ .



**Fig. 4.** Combining trust and reputation

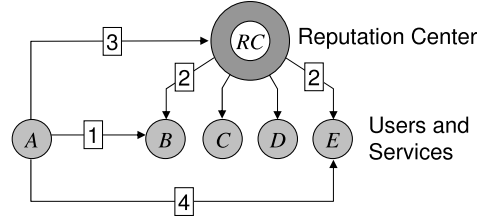
Agent  $B$  has reputation score  $\vec{R}_B^{RC}$  (arrow 1), and agent  $A$  has trust  $\omega_{RC}^A$  in the Reputation Centre (arrow 2), so that  $A$  can derive a measure of trust in  $B$  (arrow 3). Agent  $B$ 's trust in  $F$  (arrow 4) can be recommended to  $A$  so that  $A$  can derive a measure of trust in  $F$  (arrow 5). Mathematically this can be expressed as:

$$\omega_F^A = \omega_{RC}^A \otimes \vec{R}_B^{RC} \otimes \omega_F^B \quad (11)$$

The compatibility between Bayesian reputation systems and subjective logic makes this a very flexible framework for analysing trust in a network consisting of both reputation scores and private trust values.

## 5 Trust Derivation Based on Trust Comparisons

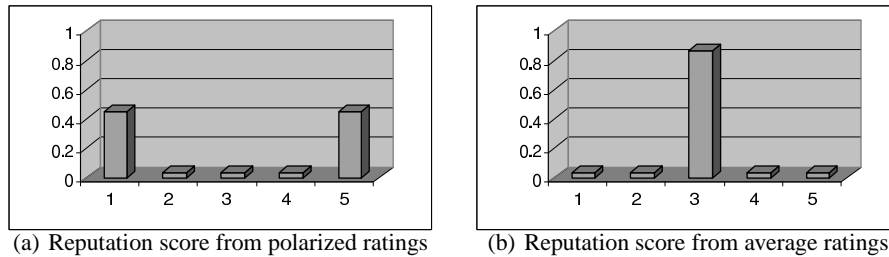
It is possible that different agents have different trust in the same entity, which intuitively could affect the mutual trust between the two agents. Fig.5 illustrates a scenario where  $A$ 's trust  $\omega_B^A$  (arrow 1) conflicts with  $B$ 's reputation score  $\vec{R}_B^{RC}$  (arrow 2).



**Fig. 5.** Deriving trust from conflicting trust

As a result  $A$  will derive a reduced trust value in the Reputation Centre (arrow 3). Assume that  $A$  needs to derive a trust value in  $E$ , then the reduced trust value must be taken into account when using  $RC$ 's reputation score for computing trust in  $E$ . The operator for deriving trust based on trust conflict produces a binomial opinion over the binary state space  $\{x, \bar{x}\}$ , where  $x$  is a proposition that can be interpreted as  $x$ : " $RC$  provides reliable reputation scores", and  $\bar{x}$  is its complement. Binomial opinions have the special notation  $\omega_x = (b, d, u, a)$  where  $d$  represents disbelief in proposition  $x$ .

What represents difference in trust values depends on the semantics of the state space. Assume that the state space consists of five rating levels, then Fig.6.a represents a case of polarised ratings, whereas Fig.6.b represents a case of average ratings. Interestingly they have the same point estimate of 0.5 when computed with Eq.(4).



**Fig. 6.** Comparison of polarized and average reputation scores

We will define an operator which derives trust based on point estimates as defined by Eq.(4). Two agents having similar point estimates about the same agent or proposition should induce mutual trust, and dissimilar point estimates should induce mutual distrust.



**Definition 1 (Trust Derivation Based on Trust Comparison).**

Let  $\omega_B^A$  and  $\omega_B^{RC}$  be two opinions on the same state space  $B$  with a set rating levels.  $A$ 's trust in  $RC$  based on the similarity between their opinions is defined as:

$$\omega_{RC}^A = \omega_B^A \downarrow \omega_B^{RC} \quad \text{where} \quad \begin{cases} d_{RC}^A = |\sigma(\vec{R}_B^A) - \sigma(\vec{R}_B^{RC})| \\ u_{RC}^A = \text{Max}[u_B^A, u_B^{RC}] \\ b_{RC}^A = 1 - b_{RC}^A - u_{RC}^A \end{cases} \quad (12)$$

The interpretation of this operator is that disbelief in  $RC$  is proportional to the greatest difference in point estimates between the two opinions. Also, the uncertainty is equal to the greatest uncertainty of the two opinions.

With the trust comparison trust derivation operator,  $A$  is able to derive trust in  $RC$  (arrow 3). With the above described trust and reputation measures,  $A$  is able to derive trust in  $E$  expressed as:

$$\omega_E^A = \omega_{RC}^A \otimes \omega_E^{RC} \quad (13)$$

This provides a method for making trust derivation more robust against unreliable or deceptive reputation scores and trust recommendations.

## 6 Numerical Example

By considering the scenario of Fig.5, assume that  $RC$  has received 5 mediocre and 5 excellent ratings about  $B$  as in Fig.6.a, and that  $A$  has had 10 average private experiences with  $B$ , as in Fig.6.b. Then  $\sigma(\vec{R}_B^{RC}) = \sigma(\vec{R}_B^A) = 0.5$ , so that  $d_{RC}^A = 0$ . According to Eq.(8) we get  $u_B^{RC} = u_B^A = 1/6$ , so that  $u_{RC}^A = 1/6$ , and according to Eq.(12) we get  $b_{RC}^A = 5/6$ . With  $a_{RC}^A = 0.9$  the derived binomial opinion is  $\omega_{RC}^A = (5/6, 0, 1/6, 0.9)$ , which indicates a relatively strong, but somewhat uncertain trust.

Assume further the aggregate ratings  $\vec{R}_E^{RC} = (0, 4, 2, 2, 0)$ , i.e. reflecting 0 mediocre, 4 bad, 2 average, 2 good and 0 excellent ratings about  $E$ . The base rate vector is set to  $\vec{a} = (0.1, 0.2, 0, 2, 0.4, 0.1)$  and the non-informative prior weight  $C = 2$ . Using Eq.(2), the multinomial reputation score is  $\vec{S}_E = (0.02, 0.44, 0.24, 0.28, 0.02)$ . The point values for each level from mediocre to excellent are: 0.00, 0.25, 0.50, 0.75 and 1.00. Using Eq.(4) the point estimate reputation is  $\sigma = 0.46$ .

Using Eq.(10) and the fact that  $\sum_{i=1}^k \vec{R}_E^{RC}(x_i) = 8$ , the reputation parameters can be converted to the binomial  $(r, s) = (3.68, 4.32)$ . Using Eq.(8)  $RC$ 's trust in  $E$  in the form of a binomial opinion can be computed as  $\omega_E^{RC} = (0.368, 0.432, 0.200, 0.500)$  where the base rate trust has been set to  $a_E^{RC} = 0.5$ .

The transitivity operator can now be used to derive  $A$ 's trust in  $E$ . The base rate sensitive operator from [11] will be used, which for this example is expressed as:

$$\begin{cases} b_E^{A:RC} = (b_{RC}^A + a_{RC}^A u_{RC}^A) b_E^{RC} \\ d_E^{A:RC} = (b_{RC}^A + a_{RC}^A u_{RC}^A) d_E^{RC} \\ u_E^{A:RC} = 1 - b_E^{A:RC} - d_E^{A:RC} \\ a_E^{A:RC} = a_E^{RC} \end{cases} \quad (14)$$

$A$ 's trust in  $E$  can then be computed as the opinion  $\omega_E^A = (0.362, 0.425, 0.213, 0.500)$ , which in terms of probability expectation value is  $E(\omega_E^A) = 0.4686$ . This rather weak trust was to be expected from the relatively negative ratings about  $E$ .

## 7 Conclusion

Trust and reputation management represents an important approach for stabilising and moderating online markets and communities. Integration of different systems would be problematic with incompatible trust and reputation systems. We have described how it is possible to elegantly integrate Bayesian reputation systems and trust analysis based on subjective logic. This provides a flexible and powerful framework for online trust and reputation management.

## References

1. A. Jøsang. Artificial reasoning with subjective logic. In Abhaya Nayak and Maurice Pagnucco, editors, *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*, Perth, December 1997. Australian Computer Society.
2. A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.
3. A. Jøsang. Probabilistic Logic Under Uncertainty. In *The Proceedings of Computing: The Australian Theory Symposium (CATS2007)*, CRPIT Volume 65, Ballarat, Australia, January 2007.
4. A. Jøsang, E. Gray, and M. Kinatader. Simplification and Analysis of Transitive Trust Networks. *Web Intelligence and Agent Systems*, 4(2):139–161, 2006.
5. A. Jøsang, R. Hayward, and S. Pope. Trust Network Analysis with Subjective Logic. In *Proceedings of the 29<sup>th</sup> Australasian Computer Science Conference (ACSC2006)*, CRPIT Volume 48, Hobart, Australia, January 2006.
6. A. Jøsang, S. Hird, and E. Faccar. Simulating the Effect of Reputation Systems on e-Markets. In P. Nixon and S. Terzis, editors, *Proceedings of the First International Conference on Trust Management (iTrust)*, Crete, May 2003.
7. A. Jøsang and R. Ismail. The Beta Reputation System. In *Proceedings of the 15th Bled Electronic Commerce Conference*, June 2002.
8. 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.
9. 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.
10. A. Jøsang and S. Pope. Semantic Constraints for Trust Transitivity. In S. Hartmann and M. Stumptner, editors, *Proceedings of the Asia-Pacific Conference of Conceptual Modelling (APCCM) (Volume 43 of Conferences in Research and Practice in Information Technology)*, Newcastle, Australia, February 2005.
11. A. Jøsang, S. Pope, and S. Marsh. Exploring Different Types of Trust Propagation. In *Proceedings of the 4th International Conference on Trust Management (iTrust)*, Pisa, May 2006.
12. 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.