Trust network-based filtering of aggregated claims

Jennifer Golbeck* and Bijan Parsia

Maryland Information and Network Dynamics Laboratory,
University of Maryland, College Park, MD 20742, USA
E-mail: golbeck@cs.umd.edu E-mail: bparsia@isr.umd.edu
*Corresponding author

Abstract: On the semantic web, assertions may be aggregated from many sources, those aggregations filtered, reasoned over, aggregated with other aggregators, displayed, scraped, extracted, recombined, and otherwise processed without significant human oversight. To preserve the connection between assertions and their source, various provenance schemes for semantic web data have been explored. However, the primary focus has been on authenticating the author of a particular statement, e.g., using digital signatures, but there is no provision for relating the authenticity of the source of the assertion and the trustworthiness of the assertion itself. This paper presents a method for using semantic web based trust networks to infer the reputation of sources for a statement and compose the reputation of several sources. By calculating a trust rating for each statement based on the ratings of its sources, the set of statements can be filtered based on the rating.

Keywords: trust; semantic web; knowledge bases; ontologies; social networks; information filtering.

Reference to this paper should be made as follows: Golbeck, J. and Parsia, B. (2006) 'Trust network-based filtering of aggregated claims', *Int. J. Metadata, Semantics and Ontologies*, Vol. 1, No. 1, pp.58–65.

Biographical notes: Jennifer Golbeck received her AB (Economics), SB (Computer Science) and SM (Computer Science) from the University of Chicago, and her PhD in Computer Science from the University of Maryland, College Park. She is currently a Research Associate at the University of Maryland, College Park. Her research interests are in trust and social networks, the Semantic Web, information visualisation, and human computer interaction.

Bijan Parsia received his BA (Philosophy) from Wesleyan University and his MA (Philosophy) from the University of North Carolina at Chapel Hill and is currently a Research Philosopher at the MIND Laboratory of the University of Maryland, College Park. His research interests are in the Semantic Web, description logics, ontology engineering, debugging and explanation of expressive ontologies, planning, visualisation, Semantic Web Services, and epistemology.

1 Introduction

Information – in particular, 'content' – on the World Wide Web is presented with an expectation that the information consumer is a human being. People are expected to make use of a variety of cues to ascertain, for example, the proponent of a claim, the author of an paper, or the photographer who took a photo and to distinguish these from the refuter of that claim, the publisher of the paper, and the aggregator of the photos. Most of these cues are traditional: bylines, attributions, quotations, citations, authorial claims, copyright notices, and the like. Some cues derive from features of web architecture, such as the use of the Domain Name System (DNS) in Universal Resource Identifiers (URIs), or HyperText Transfer Protocol (HTTP) redirects. Digital signatures can be used to verify the particular origin of a document, and that the document was unchanged in transit, but there is no provision for relating the authenticity of the source of the document and the trustworthiness of the content of that document. Human judgement is required to determine the nature of the document and its content (e.g., real purchase order, example order for debugging, or a parody for amusement). One way by which the need for continual human intervention can be eliminated is for people and organisations to set up agreements that certain documents exchanged in certain contexts will be reliable in the appropriate ways. Given that the parties of such agreements trust each other, accepting information is reduced to verifying that it came from a trusted source. Such acceptance need not be only the acceptance of that information *as true* – the modality of the acceptance depends on the agreements. On a community, website content may be acceptable for its entertainment value.

The semantic web is conceived as the 'next generation' of the World Wide Web, wherein much of the content of the web will not be solely, or even primarily, intended for human consumption. Instead, content sensitive programs will collect, process, exchange, generate, and make decisions based on web accessible information. As web

Copyright © 2006 Inderscience Enterprises Ltd.

agents make more significant decisions, it becomes more imperative that they are more sophisticated as to how they accept information from the web. While many semantic web programs will have significant domain knowledge thus, presumably, some built-in methods for evaluating the plausibility of new information, perhaps the majority of them will be less specialised. Thus, there is a need for more general, not content specific, techniques.

Many websites are *open* and anybody can submit information to be published on the site. This can range from very restricted submissions, such as comments on papers, to the entire content of the site, as with Wikis. This is relatively unproblematic when the information submitted is always presented as a cohesive chunk, say, a Wiki page, or a specific comment, or a specific blog entry. This permits the human reader to evaluate both the content of the chunk and the context of submission (i.e., the provenance).

In contrast, in an open semantic website, this is not sufficient. For example, http://owl.mindswap.org/ is an open, RDF and OWL driven website. It accepts relatively arbitrary submission of bits of RDF and OWL. It incorporates the assertions in a submission in a variety of pages, presenting the assertions in contexts divorced from their submission and using those assertions to draw inferences, which are themselves presented on different pages. The page generation software has to decide how and where to present or otherwise use each assertion in a submission.

In this paper, we present a method for integrating semantic web based trust networks with provenance information to rate and filter a set of assertions. We describe how trust networks can be created using ontology and to present an accurate algorithm for inferring trust relationships. Those inferred values are then used to compose ratings of the reputation or trustworthiness of assertions. We describe how those ratings on assertions can then be used to filter the set of statements used in an application, thereby creating a knowledge base with a known level of validity.

2 Background and previous work

Our work is based on the premise that applications will eventually access and utilise the trust data incorporated into web-based social networks. In this section, we introduce the necessary background for successfully introducing trust to networks on the web, and present some applications that have already begun to take advantage of web-based trust.

2.1 Semantic web background

The algorithms presented in this paper are designed to be used with any social network, but current implementations are semantic web based and use the Friend-of-a-Friend (FOAF, http://foaf-project.org) vocabulary. The FOAF project defines a set of terms for letting users describe people and who they know. FOAF is one of the largest projects on the semantic web, with an estimated 2–5 million users. Some of those data come from individuals creating their own FOAF files and maintaining the information in their personal web space, but increasingly, they are coming from other web-based social networks. LiveJournal (http://livejournal.com), eCademy (http://ecademy.com) and Tribe (http://tribe.net) all publish their users' social network data in FOAF format. Other websites that have gathered social network data have chosen to make those connections available in FOAF format; for example, Howard Dean's presidential campaign produced thousands of FOAF files representing the social network created when members used a feature of their website to share links with their friends. FOAF has become a recognised means of sharing social network data between social networking websites, and the ease of producing semantic web data is encouraging this evolution. The FOAF community is actively rising to this challenge by formalising their efforts in workshops and online meetings to create a stable core vocabulary that can be used by the widest range of people and applications.

Because it is a semantic web ontology, FOAF can be easily extended to capture more detailed personal and relationship information. The trust module for FOAF (Golbeck and Hendler, 2004a) extends the FOAF vocabulary by adding a property where users state how much they trust one another. It has a scale of trust ratings that range from 1 (very little trust) to 10 (very high trust) and is used in several applications (Avesani et al., 2004; Croucher, 2004; Golbeck and Hendler, 2004b).

2.2 Related work

In computer science and on the web, trust has typically referred to mechanisms of authentication, security, and privacy. Within the W3C itself, much work has been done in this direction (Eastlake et al., 2000; Eastlake and Reagle, 2002; Ford et al., 2001; Marchiori et al., 2002). Our approach is based on social trust – trust in a *person* – rather than the trust authentication offers in the *source* of a resource.

More closely related to our work is the Platform for Internet Content Selection (PICS) (Miller et al., 1996). PICS does have an RDF/XML format, which could be used for rating the trustworthiness of an individual. We opted to create our own ontology in OWL that takes advantage of more language features and allows for more complex expressions than PICS. However, the techniques described here would work equally well on a trust network built on the PICS technology.

Mechanisms for calculating the trustworthiness of an individual are important because they will form the foundation for filtering information. There are many approaches for calculating trust that span a set of applications.

The EigenTrust algorithm (Kamvar et al., 2003) is used in peer-to-peer systems and calculates trust with a variation on the PageRank algorithm (Page et al., 1998), used by Google for rating the relevance of web pages to a search. A peer creates a direct trust rating for another peer based on its historical performance. In its simple form, the algorithm

uses a matrix representation of the trust values within the system and over a series of iterations it converges to a globally accepted trust rating of each peer. Because of safeguards built into the system, EigenTrust has been shown to be highly resistant to attack. EigenTrust is designed for a peer-to-peer system while ours is designed for use in humans' social networks, and thus there are differences in the approaches to analysing trust. In the EigenTrust formulation, trust is a measure of performance and one would not expect a single peer's performance to differ much from one peer to another. Socially, though, two individuals can have dramatically different opinions about the trustworthiness of the same person. Our algorithms intentionally avoid moving towards a global trust value for each individual to preserve the personal aspects that are foundations of social trust.

Raph Levin's Advogato project (Levin and Alexander, 1998) also calculates a global reputation for individuals in the network, but from the perspective of designated *seeds* (authoritative nodes). His metric composes certifications between members to determine the trust level of a person, and thus their membership within a group. The Advogato website at http://advogato.org, for example, certifies users at three levels – apprentice, journeyer, and master. Access to post and edit website information is controlled by these certifications. Like EigenTrust, the Advogato metric is quite attack resistant. By identifying individual nodes as 'bad' and finding any nodes that certify the 'bad' nodes, the metric cuts out an unreliable portion of the network. Calculations are based primarily on the good nodes, so the network as a whole remains secure. While the perspective used for making trust calculations is still global in the Advogato algorithm, it is much closer to the methods used in this research. Instead of using a set of global seeds, we let any individual be the starting point for calculations, so each calculated trust rating is given with respect to that person's view of the network.

Richardson et al. (2003) use social networks with trust to calculate the belief a user may have in a statement. This is done by finding paths (either through enumeration or probabilistic methods) from the source to any node, which represents an opinion of the statement in question, concatenating trust values along the paths to come up with the recommended belief in the statement for that path, and aggregating those values to come up with a final trust value for the statement. Current social network systems on the web, however, primarily focus on trust values between one user to another, and thus their aggregation function is not applicable in these systems. Their paper, intentionally, does not define a specific concatenation function for calculating trust between individuals. The algorithms we define in this paper are aimed specifically at calculating trust between agents, and an exploration of how their algorithms and ours could be combined is an interesting topic for future work.

There have been some efforts on the Semantic Web dedicated to filtering content based on who stated it and the confidence in the statements. Gil and Ratnakar (2002) addressed the issue of trusting content and information sources on the semantic web without a social networking context. Their TRELLIS system derives assessments about information sources based on individual feedback about the sources. Users of the system can annotate pieces of information and the annotations can include measures of 'credibility' and 'reliability' about a statement. These are later averaged and presented to the viewer. Using the TRELLIS system, users can view information, annotations (including averages of credibility, reliability, and other ratings), and then make an analysis. Our work uses the notion of determining the credibility of a statement but approaches it from an automated perspective rather than an annotation perspective.

3 Trust

3.1 Defining trust

In human society, trust depends on a host of factors, which cannot be easily modelled in a computational system. Past experience with a person and with their friends, opinions of the actions a person has taken, psychological factors impacted by a lifetime of history and events (most completely unrelated to the person we are deciding to trust or not trust), rumour, influence by others' opinions, and motives to gain something extra by extending trust are just a few of these factors. For trust to be used as a rating between people in social networks, the definition must be focused and simplified.

Marsh (1994) addressed the issue of formalising trust as a computational concept in his PhD dissertation at the University of Stirling. His model is complex and based on social and psychological factors. Although this work is often cited, the model is highly theoretical and difficult to implement. It is particularly inappropriate for use in social networks because his focus was on interacting agents that could maintain information about history and observed behaviours. In social networks, users assign a single rating without explicit context or history to their neighbours, and thus much of the information necessary for a system like Marsh's is missing.

Deutsch (1962) contains a frequently referenced definition of trust. He states that trusting behaviour occurs when a person (say Alice) encounters a situation where she perceives an ambiguous path. The result of following the path can be good or bad, and the occurrence of the good or bad result is contingent on the action of another person (say Bob). Furthermore, the negative impact of the bad result is greater than the positive impact of the good result. This further motivates Alice to make the correct choice. If Alice chooses to go down the path, she has made a trusting choice. She trusts that Bob will take the steps necessary to ensure the good outcome. The requirement that the bad outcome must have greater negative implications than the good outcome has positive implications has been countered in other work (Golembiewski and McConkie, 1975), which does not always require disparity.

Sztompka (1999) presents and justifies a simple, general definition of trust similar to that of Deutsch: "Trust is a bet about the future contingent actions of others". There are two main components of this definition: belief and commitment. First, a person believes that the trusted person will act in a certain way. The belief alone, however, is not enough to say there is trust. Trust occurs when that belief is used as the foundation for making a commitment to a particular action. These two components are also present in the core of Deutsch's definition: we commit to take the ambiguous path if we believe that the trusted person will take the action that will produce the good outcome.

We adopt this as the definition of trust for our work: trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome. The action and commitment does not have to be significant. We could say Alice trusts Bob if she chooses to read a statement he has asserted.

3.2 Building a trust network: properties, values, and ontologies

Several properties of trust are important to our algorithms for inferring trust. The primary property is *transitivity*. Trust is not perfectly transitive in the mathematical sense; that is, if Alice highly trusts Bob, and Bob highly trusts Chuck, it does not always and exactly follow that Alice will highly trust Chuck. There is, however, a notion that trust can be passed between people. When we ask a trusted friend for an opinion about a plumber, we are taking the friend's opinion and incorporating that to help form a preliminary opinion of the plumber.

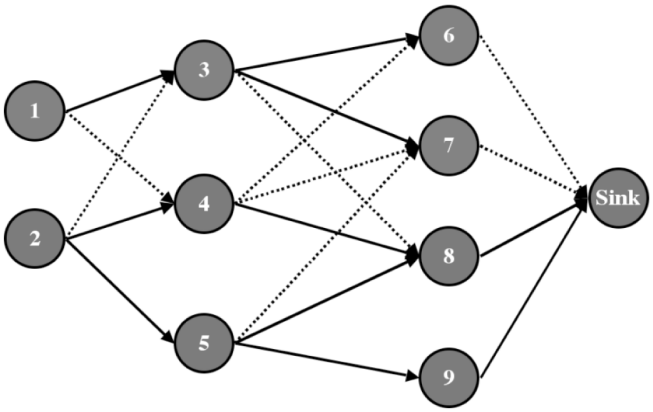
In social networks, it is also important to note the *asymmetry* of trust. For two people involved in a relationship, trust is not necessarily identical in both directions. Because individuals have different experiences, psychological backgrounds, and histories, it is understandable why two people may trust each other in different amounts. While asymmetry occurs in all types of human relationships, it is documented more in situations where the two people are not of equal status. For example, employees typically say they trust their supervisors more than the supervisors trust the employees. This is seen in a variety of hierarchies (Yaniv and Kleinberger, 2000). Even outside of hierarchies, social situations can arise with asymmetric trust. One of the more extreme instances of this is one-way trust, where circumstances force one person to trust the other, but there is no reciprocal trust (Hardin, 2002; Cook, 2001). Because trust is naturally asymmetric, trust ratings in our system are also asymmetric and represented as directed edges in a network.

One property of trust that is important in social networks, and which has been frequently overlooked in the past, is the *personalisation* of trust. Trust is inherently a personal opinion. Two people often have very different opinions about the trustworthiness of the same person. For an example, we need only to look into politics. In the USA, when asked, "do you trust the current President to effectively lead the country?" the population will be about evenly split – half will trust the person very highly, and the other half will have very little trust in the person's abilities.

Personalisation plays into calculating trust recommendations by affecting the accuracy of a recommendation. If a person wants a recommendation about how much to trust the President, an algorithm that simply composes all of the values in the system can be expected to give an answer that falls almost directly in between 'very low trust' and 'very high trust'. With most people having a strong opinion, this middle rating will not mean much. It reflects the opinion of the population, and is not a recommendation to the *individual*. Our algorithms are based on the perspective of the user. It looks at friends whom the user trusts about their opinions on a topic, the people whom those friends trust, and so on. Thus, the opinions of people whom the user does not trust much are given very little consideration, and the opinions of people whom the user trusts highly are given more consideration.

Figure 1 depicts a sample social network. The solid lines indicate relationships with trust, and the dashed lines indicate no trust. The node for which we are determining a trust rating, called the *sink*, is trusted by two nodes (8 and 9) and not trusted by two nodes (6 and 7). If a trust rating for the sink were calculated by composing all of the direct ratings of the sink, every node would get the same recommendation. However, if we take into account the information that we know about the structure of the network from the perspective of each node, a much more informative recommendation can be made. Node 1 can accept information only from its trusted neighbours. In the end, only the trust ratings given by Nodes 6 and 7 will propagate back to Node 1. Both 6 and 7 do not trust the sink, and only their opinion will be passed back to Node 3 and then to Node 1 who will calculate that the sink is not to be trusted. Similarly, Node 2 also only considers trusted paths. At the end of those paths, Nodes 8 and 9 both have directly rated the sink to be trustworthy. Their values are passed back along the network paths through Nodes 4 and 5 to Node 2. Node 2 will conclude that the sink is to be trusted Thus, if perspective is taken into account, Nodes 1 and 2 can each receive relevant and accurate information about how much to trust the sink, even though their opinions are diametrically opposed and the information in the network is mixed.

Figure 1 Nodes consider ratings from the people they trust highly (indicated by solid edges). Nodes with low trust ratings (indicated with dashed edges) are only considered when they are a direct rating of the sink, but are not used in finding paths to the sink. The ratings made by trusted nodes that directly rated the sink are used in coming up with a recommendation about how much the source should trust the sink



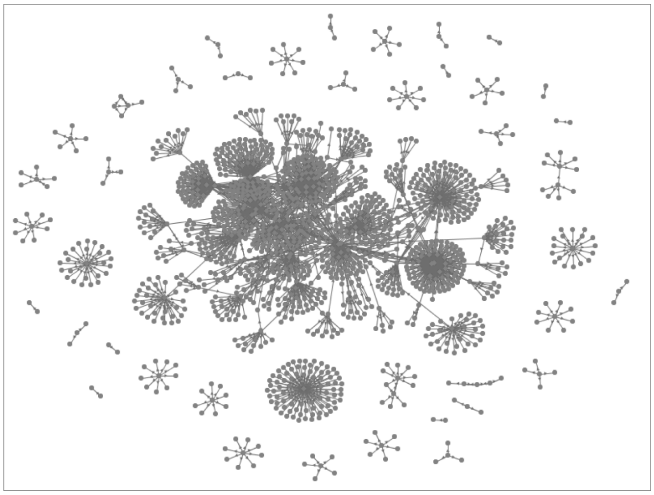
The way relationships are expressed in social networks, particularly in web-based social networks, is through an explicit statement. Users rate their connections on a scale usually made available by the service.

There are many systems for rating how much one person trusts another; in the trust literature, and the rating systems vary. The Advogato system uses a three tiered system (Apprentice, Journeyman, Master) for rating its members (Levin and Alexander, 1998). Orkut, at (http://Orkut.com), offers users the ability to rate many features of their friends, including the trustworthiness with zero to three smiley faces. Semantic Web trust projects have used a scale of 1-9, 1-10 (Golbeck and Hendler, 2004a), or a value that can fall anywhere in the [0, 1] range (Richardson et al., 2003).

This work uses the FOAF trust module, which is available at http://trust.mindswap.org/ont/trust.owl. The ontology uses a scale from 1 to 10, where 1 represents low trust and 10 represents very high trust. This is an intuitive scale for people to use and is currently the foundation of the trust network at http://trust.mindswap.org. The network currently contains nearly 2,000 people, which is a large enough foundation for running experiments.

The ontology also provides the capability to create trust ratings with respect to a specific topic. For example, Person A may rate Person B high with respect to computer science knowledge, but give a low rating to Person B with respect to movie recommendations. The subgraph of edges with the selected subject is extracted from the whole network and utilised in the same way. In the context of provenance tracking, it is possible to extract the trust ratings related to the statements of interest, and use only those ratings in the analysis. This work uses general trust ratings, not context specific ones, because there are more general ratings in the network, and the algorithms presented here are the same for general and context specific ratings.

Figure 2 The trust network at http://trust.mindswap.org



3.3 Algorithms for calculating trust

This work uses a simple recursive algorithm for inferring trust ratings in a social network. If the source has not directly rated the sink, the source queries each of its neighbours for their rating of the sink. Each neighbours' value is weighted by the trust rating the source has given that neighbour, and the weighted average is calculated. This is shown in formula 1, where t_{ij} represents the trust from node i to node j.

$$t_{is} = \frac{\sum_{j \in \text{adj}(i)} t_{ij}^{*} \; t_{js}}{\sum_{j \in \text{adj}(i)} t_{ij}}. \tag{1}$$

If the neighbour, node j, has directly rated the sink, it returns that rating. Otherwise, it repeats the process of querying neighbours and returns its own weighted average. This algorithm is similar to a breadth-first-search, and runs in polynomial time.

Analyses of this and similar algorithms for inferring trust have been presented (Golbeck and Hendler, 2004a). Experiments compared inferred trust values with values actually assigned by users in the trust project network. Our published results show that the inferred value is, on average, within 1.16 of the actual value (on a scale of 1 to 10). This was significantly more accurate than the other systems we tested.

4 Integrating trust with provenance

4.1 From trust network inferences to accepting claims

In our prototype, we focus our attention on individual claims, with the key provenance information being the set of claimants. That is, we primarily deal with knowledge provenance. A claim is simply any RDF triple submitted to our website, whether by a web form, via some aggregation

mechanism such as an RSS feed, or by a web service API. Triples are typically submitted in batches that we will call '*snippets*' with user-supplied metadata about the snippet inherited by each claim. For example, on http://www.mindswap.org/, users can submit snippets about papers they have authored, either through a free form textarea to craft their RDF directly by hand, or simple elicitation forms to help ensure data consistency, coherence, and completeness.

The key metadata for each snippet is the person submitting the claim, that is, the *claimant*. When a claimant is identifiable as a particular node in the trust network, we can attempt to determine a local trust rating for that claimant. If a user of the site has registered their trust network identifier with the site, then trust rating can be inferred with their node as the source. Given a particular trust rating for a claim, we customise the display behaviour of the site. The simplest customisation is to suppress the display of any claim from a claimant whose trust rating is below a certain, user configurable threshold.

The situation is slightly more complex if there are multiple claimants for a particular claim. There are a number of functions one could use to derive a trust rating for the claim based on the set of ratings for the claimants. However, the straightforward solution – take the maximum rating of the claimants – has a great deal of intuitive plausibility and that is what we use. Because the kind of websites we are building are community oriented portals, the general goal for the site is to be interesting, relevant, and useful to that community. Thus, trust in a person is a measure of our belief that they will create well-presented, relevant, interesting, and useful information, as determined by the portal's community standards. Because the trustworthiness of the claimant is not interpreted as evidence for or against the claim, there is no need to average or balance other divergent ratings.

4.2 Using claim ratings in semantic web systems

The first and most obvious application of the ratings for claims is to filter the content of the website based on the value of the rating. Consider applying this technology in the context of some of the many 'rumour' sites on the web. As one example, MacRumors (http://macrumors.com) allows users to submit rumours about news and technology releases related to Apple Computer. The author of each rumour is tracked, and community members already have the ability to rate rumours as positive or negative. A website with that model would significantly benefit from a semantically aware system of trust and provenance. A network of ratings that reflect one person's opinion about the quality of posts made by another user, and following the system of generating ratings for statements based on their provenance creates the groundwork for allowing users to customise the site. Users can choose a minimum trust level of statements that appear on the site, and not only is the site personalised, but optimised for the user according to their preferences and social network connections. Although 'rumour' sites provide an intuitive example because of the obvious variation in the credibility of statements, this technique clearly can be applied to any site where statements originate from a variety of sources.

We have applied such a technique in FilmTrust, a website that integrates social network and movie reviews (http://trust.mindswap.org/FilmTrust). Members of the site can rate films and write reviews. In addition, they maintain a set of friends who are rated according to how much the user trusts their opinion of movies. Using the resulting trust network, a personalised trust value is calculated for each user who has written a review of a movie. The reviews are then displayed in sorted order for the user, so those from the most trusted sources appear higher than those from less trusted sources (Figure 3).

Figure 3 A page with movie ratings and reviews from the FilmTrust website at http://trust.mindswap.org/FilmTrust/

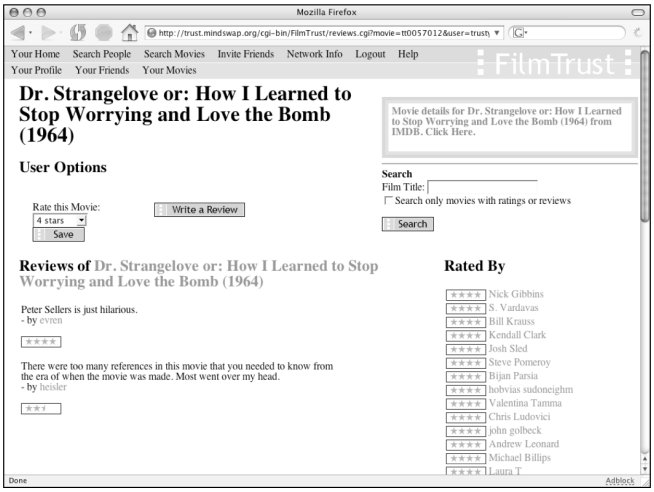


Figure 3 shows the movie reviews page for 'Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb'. There are two reviews on this page, by users 'evren' and 'heisler', ordered by the inferred trust rating from the user to the authors of the reviews. For other users, the order will be reversed because they have a higher trust value for 'evren' than for 'heisler'.

In addition to these features, users can also choose to filter reviews, displaying only those from users with trust values above a certain threshold. While movie reviews are very simple and homogenous semantic statements, this website provides a first glimpse at how trust ratings can be used to personalise the user experience and present the most useful information in a prominent way.

4.3 Filtering inferences in knowledge-bases with trust values

Filtering the base claims of the system is useful and interesting, but base claims on semantic websites form only part of the picture. semantic web portals tend to be oriented around RDFS and OWL ontologies, that is, logical theories of varying degrees of expressiveness. A semantic website, therefore, is based on a knowledge base and the character of

the website is significantly influenced by the sorts of reasoning it supports. The ordering of filtering and inferring is important to consider. If the set of base claims are filtered first, and then inferences are made over the filtered set, there are two results. First, using the filtered base claims as the fact base for inferences already filters the inferences. This allows users to conclude that any inferred statements should have at least the same trust rating as the minimum value in the filter, because all of the claims that allow the inference meet or exceed that minimum. However, this does not provide a mechanism for actually calculating a value for an inferred statement – it only sets a minimum bound. Using this filter-then infer method also means that the set of inferred statements are limited – it is possible that many other statements could have been inferred from the unfiltered base.

If the order of inferring and filtering is reversed so that all of the inferences are made over the full set, and then the set of all statements – base and inferred – are filtered, the issue becomes more complex because it requires that some trust value be established for the inferred statements. The rating for an inferred statement should clearly be some combination of the statements that lead to the inference. However, a number of different statements could lead to an inference. Consider the following set of base claims in N3. Each statement is marked with a trust rating calculated from its claimants.

```
9:    Person a owl:Class.
8:    SpouseOfStudent
      a owl:Class;
8:    rdfs:subClassOf :Person,
8:    [a owl:Restriction;
      owl:allValuesFrom :Student;
      owl:onProperty :marriedTo ],
8:    [a owl:Restriction;
      owl: cardinality "1";
      owl: onProperty :marriedTo].
7:    Student a owl: Class;
7:    rdfs:subClassOf :Person.
9:    University a owl:Class .
6:    attendsUniversity a owl:ObjectProperty;
6:    rdfs:domain :Student;
6:    rdfs:range :University.
10:   marriedTo a owl:ObjectProperty;
10:   owl:inverseOf :marriedTo.
10:   Daniel a :SpouseOfStudent;
9:    marriedTo :Jennifer.
8:    Jennifer a :Person;
6:    attendsUniversity :UMCP;
9:    marriedTo :Daniel.
```

From this example, we can infer that: *Jennifer* is a :*Student*. What should be the rating for that inferred claim? There are several ways that it can be inferred. Because *Jennifer :attendsUniversity :UMCP* (known at level 6), and the domain of :*attendsUniversity* is :*Student* (rated at level6) ,we can infer that :Jennifer is a :Student. This inference comes from two simple statements, rated at the same level, so it seems intuitive to rate the inference from these sources at a level 6, like the composite statements. There are other ways to infer that :*Jennifer* is a :*Student*, though, and they may have a higher rating than the 6 achieved with the first method. We also know that :*Daniel*, a :*SpouseOfStudent* (known at level 10), is :*marriedTo* :*Jennifer* (known at level 9). Because for instances of the :SpouseOfStudent class, the object of :marriedTo must be from the class :Student (known at level 8) and that there must be exactly one spouse (because of the cardinality restriction known at level 8) – so :*Jennifer* must be the only person that :*Daniel* is :*marriedTo* – we can infer that :*Jennifer* is a :*Student*. How to combine this series of statements into a rating for the inferred value is not clear. There is also a third way of inferring the fact, stemming from the claim that :*Jennifer is* :*marriedTo* :*Daniel* (rated at level 9). Because :*marriedTo* is the inverse of itself (rated at level 10), we know that :*Daniel is* :*marriedTo* :*Jennifer* (even if that were not explicitly stated). This leads to the second inference we made, allowing us to conclude that :*Jennifer* is a :*Student*.

This example illustrates several issues raised when considering how to rate inferred statements. First, for each set of statements that leads to an inference, there must be a way to combine the ratings of the composite statements to come up with a rating for the inferred statement. Even if we took the simple route of using just the minimum rating from the set of composite statements as the rating for the inferred statement, there are still more problems. If a statement is inferred from several sets of statements, there are now several ratings for that inferred statement. How to choose a final value for the inferred statement is not clear – it could be the maximum value from all of the possible values assigned or some composite. On top of that, the primary issue illustrated by the above example is that the number of ways a statement can be inferred can grow very quickly. To consider every possible combination of claims that lead to an inference could become computationally difficult. Because inferences are such a fundamental issue on the semantic web, the question of establishing trust values for inferred statements space will be the focus of future work in this space.

5 Conclusions and future work

In this paper, we have presented a system for inferring reputation in semantic web based trust networks, and using those values to create ratings for statements made by individuals within the network. By combining trust values with provenance information, we show how users can filter knowledge bases based on a minimum trustworthiness rating. Refining the trust metric is one point of future work. Though this analysis has shown that our simple metric is relatively accurate, considering additional structural features of the network such as path length, number of paths, and the use of intermediate nodes, may lead to more accurate metrics. Understanding which features of a trust inference algorithm should be incorporated for the most accurate metric will be an important step as this work progresses.

While we present the idea of using this system of ratings over inferences, a major future step will be to address the issue of ratings for statements derived through inferences. Section 4 presented a detailed argument as to why creating these ratings are difficult. Further analysis of this issue will be critical in extending this work into a broader application on the semantic web.

References

Avesani, P., Massa, P. and Tiella, R. (2004) 'Moleskiing: a trust-aware decentralized recommender system', *Proceedings of 1st Workshop on Friend of a Friend, Social Networking and the (Semantic) Web*, September 1–2, Galway, Ireland.

Cook, K. (Ed.) (2001) *Trust in Society*, Russell Sage Foundation, New York.

Croucher, T. (2004) 'A model of trust and anonymity in a content rating system for e-learning systems', *Proceedings of 1st Workshop on Friend of a Friend, Social Networking and the (Semantic) Web*, September 1–2, Galway, Ireland.

Deutsch, M. (1962) 'Cooperation and trust. some theoretical notes', in Jones, M.R. (Ed.): *Nebraska Symposium on Motivation*, Nebraska University Press, Lincoln, Nebraska.

Eastlake, D. and Reagle, J. (2002) *XML Encryption Syntax and Processing*, W3C Candidate Recommendation xmlenc-core, August.

Eastlake, D., Reagle, J. and Solo, D. (2000) *XML-Signature Syntax and Processing*, W3C Recommendation xmldsig-core, October.

Ford, W., Hallam-Baker, P., Fox, B., Dillaway, B., LaMacchia, B., Epstein, J. and Lapp, J. (2001) *XML Key Management Specification (XKMS)*, W3C Note xkms, March.

Gil, Y. and Ratnakar, V. (2002) 'Trusting information sources one citizen at a time', *Proceedings of the First International Semantic Web Conference (ISWC)*, Sardinia, Italy, June.

Golbeck, J. and Hendler, J. (2004a) 'Accuracy of metrics for inferring trust and reputation', *Proceedings of 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, Northamptonshire, UK, October.

Golbeck, J. and Hendler, J. (2004b) 'Reputation network analysis for email filtering', *Proceedings of the First Conference on Email and Anti-Spam*, Mountain View, CA, July.

Golembiewski, R.T. and McConkie, M. (1975) 'The centrality of interpersonal trust in group processes', in Cary Cooper (Ed.): *Theories of Group Processes*, Wiley, Hoboken, NJ.

Hardin, R. (2002) *Trust & Trustworthiness*, Russell Sage Foundation, New York.

Kamvar, S.D. Schlosser, M.T. and Garcia-Molina, H. (2003) 'The eigentrust algorithm for reputation management in P2P networks', *Proceedings of the 12th International World Wide Web Conference*, May 20–24, Budapest, Hungary.

Levin, R. and Alexander, A. (1998) 'Attack resistant trust metrics for public key certification', *7th USENIX Security Symposium*, San Antonio, Texas, January.

Marchiori, M., Cranor, L., Langheinrich, M., Presler-Marshall, M. and Reagle, J. (2002) *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*, W3C Recommendation REC-PICS-services, April.

Marsh, S. (1994) *Formalising Trust as a Computational Concept*, PhD Thesis, Department of Mathematics and Computer Science, University of Stirling, Stirling, Scotland, UK.

Miller, J., Resnick, P. and Singer, D. (1996) *Platform for Internet Content Selection (PICS) Rating Services and Rating Systems*, W3C Recommendation REC-PICS-services, October.

Orkut (2004) Available: http://orkut.com/.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1998) *The Pagerank Citation Ranking: Bringing Order to the Web*, Technical Report, Stanford University, Stanford, CA.

Richardson, M., Rakesh, A. and Pedro, D. (2003) 'Trust management for the semantic web', *Proceedings of the Second International Semantic Web Conference*, Sanibel Island, Florida.

Sztompka, P. (1999) *Trust: A Sociological Theory*, Cambridge University Press, Cambridge.

The Friend-Of-A-Friend (FOAF) Project (2004) Available: http://foaf-project.org.

Yaniv, I. and Kleinberger, E. (2000) 'Advice taking in decision making: egocentric discounting and reputation formation', *Organizational Behavior and Human Decision Processes*, November, Vol. 83, No. 2, pp.260–281.