COMPREHENSIVE TRUST METRICS FOR INFORMATION NETWORKS

Jeff Pasternack* & Dan Roth
University of Illinois at Urbana-Champaign
Urbana, Illinois, 61801

ABSTRACT

Existing computational trust systems analyze information networks to determine the "trustworthiness" of the nodes, but the scalar values they produce are both opaque and semantically variable, and knowing only that the trustworthiness of a website is "27" is not helpful to the user. Moreover, the simplistic means by which they are typically calculated can yield misleading results, sometimes dramatically so. We present a new, standardized set of trust metrics that instead compute the trustworthiness of an information source as a triple of truthfulness, completeness, and bias scores, and argue that these must be calculated *relative to the user* to be meaningful. We then explore these new metrics with a user study.

1 INTRODUCTION

Information-based trust systems (Artz and Gil 2007) are algorithms that, given an information network containing information sources (e.g. authors, publishers, websites, etc.) making a variety of claims ("the atomic mass of gold is 42"), determine how much a user should trust the former and believe the latter (a highly trusted source reliably provides trustworthy documents and a highly believed claim is likely to be true). As data has become more abundantly available to military agents, the importance of such systems has grown tremendously. The potential of online, open source intelligence has exploded with the mass adoption of blogs, wikis, message boards and other collaborative media, but gone is the high entry barrier (and enforced journalistic standards) of older, established media such as newspapers and television. Similarly, with growing scale, reduced budgets, and instant communication, this "information overload" also vexes communications and human intelligence, with an ever-increasing need to sift out the true from the spurious.

Unfortunately, the assessments produced by these algorithms are simplistic, assigning a probability or arbitrary weight as the scalar trustworthiness of a source, based upon the accuracy of their claims. For example, TruthFinder (Yin, Yu, and Han 2008) calculates a source's trustworthiness as the arithmetic mean of the probabilities of its claims. Consider a short document authored by Sarah: "John is running against me. Last year, John spent $100,000 of taxpayer money on travel. John recently voted to confiscate, without judicial process, the private wealth of citizens." If all of these statements are true, Sarah and her document would thus be considered highly trustworthy.

However, if we know more about the background, we might find that Sarah is misleading us and is, in fact, quite *un*trustworthy despite her factuality (i.e. we should not consider her to be a reliable source of information). If "John is running against Sarah" is a well-known, "easy" fact (Galland et al. 2010), Sarah's correct assertion thereof is unimportant (taken to the extreme, one could otherwise pad a document with banalities like "1+1=2, 1+2=3..." to produce a seemingly-trustworthy document, regardless of its other content). Further, if $100,000 in travel expenses is par for John's office because it necessitates a great deal of travel, Sarah has conveniently neglected to mention this, instead inviting the reader to compare his costs to their own prior expectation of what "typical" travel expenses should be and conclude, incorrectly, that John has enjoyed gratuitously luxurious accommodation. Similarly, Sarah's "wealth confiscation" typically goes by the slightly more innocuous term "taxation", but her biased language suggests to the reader that John has approved of something unusually nefarious.

To counter these problems, we propose assessing trustworthiness not as a scalar, but rather three separate values: truthfulness, completeness, and bias. Decomposing trust into these three components allows the trust system's user to *meaningfully* assess the extent to which a document or information source can be relied upon, and calculating them consistently across algorithms permits the output of competing trust systems to be directly compared and evaluated. We also find that the user himself is essential in calculating the trustworthiness of a source, in addition to the prior knowledge he may bring to the trust system itself (e.g. as in (Pasternack and Roth 2010a)); for example, if an investor and a politician are each reading about House debates on a new corporate tax bill, the investor may not care who introduced the bill, while the politician may not care about the fine-grain details of the tax. An author that occasionally flubs those details may thus nonetheless still be trusted by the politician, but not the investor. By considering the relative importance of information and preexisting beliefs of the user, we can better approximate the user's own judgment and provide a more accurate trust analysis.

2 BACKGROUND

Broadly speaking, information networks can be categorized as homogeneous and heterogeneous networks; our work applies to the latter type, but we briefly discuss the former below.

2.1 Homogenous (Reputation) Networks

Homogeneous networks have only a single type of entity, with edges forming recommendations, votes, or other relationship between two entities in the graph; these are more commonly known as reputation networks. Reputation systems are frequently employed in P2P applications and social groups, such as Eigentrust (Kamvar, Schlosser, and Garcia-molina 2003) and Advogato (Levien 2008). Alternatively, PageRank (Brin and Page 1998) and (Kleinberg 1999)'s Hubs and Authorities can be seen as reputation systems where links imply recommendation. In these systems there is a core principle of transitivity: if you trust Bob, and he trusts Jane, then you should also trust Jane. The mechanics of transitivity vary; (Josang, Marsh, and Pope 2006) uses subjective logic, while other systems use an iterative algorithm for "trust propagation". However, the amount of information that can be encoded in homogeneous networks is limited; for news websites we might add edges corresponding to links between them (again on the basis that these are implicit recommendations), but we would not look at the actual articles on those websites, or the claims they contain. Thus, while the semantics of trustworthiness within a reputation network are often relatively straightforward (based on "flows" of trust among entities) they are a poor choice for information sources, where recommendations among the sources (if present) are much less important than what the source actually says.

2.2 Heterogeneous Networks

The heterogenous networks seen in trust problems comprise of a number of (possibly hierarchical, e.g. document $\rightarrow$ author $\rightarrow$ publisher) information sources, each asserting a number of claims. The trust system seeks to both find the trustworthiness of these sources *and* the believability of their claims, although in the user's interest may be specific, e.g. determining the atomic weight of gold or finding trustworthy articles about Bill Clinton.

Fact-finders are the predominant class of algorithms for finding trust in these networks; these take as input a bipartite graph of sources and claims, with edges connecting each source to the claims it asserts, and output a trustworthiness score for each source and a belief score for each claim (with the semantics of these scores varying with the particular algorithm). The algorithms proceed iteratively, calculating the truth of facts based on the trust in their sources, and the trust in sources based on the truth of their facts. TruthFinder (Yin, Han, and Yu 2007; Yin, Yu, and Han 2008) and Investment (Pasternack and Roth 2010a) are straightforward implementations of this idea, and some reputation systems (like Hubs and Authorities (Kleinberg 1999)) can be adapted to heterogeneous networks with relatively little effort (Pasternack and Roth 2010a). (Dong, Berti-equille, and Srivastava 2009) additionally calculates source dependence (where one source copies another) to give higher credibility to independent sources. (Galland et al. 2010) incorporates an estimate of the "hardness" of a fact in its calculations, such that knowing the answer to an easy question earns less trust than to a hard one. Finally, (Pasternack and Roth 2010b) "lifts" fact-finders from unweighted bipartite graphs to weighted k-partite graphs, allowing for attributes and group membership of the sources (like the quality of a website's design or a newspaper's Associated Press membership) to be modeled, among other extensions.

One important observation is that, regardless of the algorithm, we can readily standardize the believability of each claim as the probability that the claim is true. However, source trustworthiness scores are computed differently from algorithm to algorithm, and the most immediate choice, used by TruthFinder and others, of calculating this as the probability of the source producing a true claim (the arithmetic mean of the probabilities of the claims made by the source) is, as already discussed, readily misled. While the trustworthiness score remains an internal parameter within the trust system, we can nonetheless report a more meaningful trustworthiness evaluation using our own metrics, which (among other advantages) provide consistency by virtue of being derived directly from the (standardized) belief in the claims rather than the arbitrary trustworthiness score used by each particular algorithm.

3 THE METRICS

3.1 The Components of Trustworthiness

Besides inconsistent and problematic semantics, existing scalar trust scores for information sources suffer from being overly broad–trustworthiness cannot always be summarized with a single number. Consider, for example, two news articles about a topic, both of which consist of entirely true claims. One article may omit a number of important details, while the other may be strongly biased towards one position. A human reader interested in only the gist of the topic would be satisfied by the incomplete article, while an information extraction system building a knowledge base would find bias irrelevant.

We therefore view the trustworthiness of an information source as three interrelated, but separate, $[0, 1]$ components: truthfulness, completeness, and bias. This has a number of advantages for information consumers:

- It allows them to moderate their reading by factoring in the source's inaccuracy, incompleteness or bias (and, for example, questioning claims from a somewhat inaccurate source, or carefully maintaining objectivity when confronted with a biased source).
- They can select information sources appropriate to their needs: completeness and bias may not be important to every user.
- Similarly, when a single score is preferred as a "summary" of a source's trustworthiness, this can be computed from the truthfulness, completeness and bias with respect the user's needs.
- Each component may be explained separately to the user: a low truthfulness score is explained by inaccurate claims, low completeness by listing some of the important claims the source *does not* mention, and bias by listing claims supporting the source's favored position together with a list of counterpoint claims they omitted.

3.2 Truthfulness

Given a single, atomic assertion, truthfulness is simply our belief in the claim; that is, $\mathcal{T}(c) = P(c)$. For simplicity we restrict ourselves to Bayesian belief, but our definitions may readily be extended to Dempster-Shafer or subjective logic, allowing us to qualify our belief with the ignorance or uncertainty arising from insufficient evidence. For a collection of assertions $\mathbf{C}$, such as documents, we define

$$\mathcal{T}(\mathbf{C}) = \frac{\sum_{c \in \mathbf{C}} P(c) \cdot \mathcal{I}(c, P(c))}{\sum_{c \in \mathbf{C}} \mathcal{I}(c, P(c))}$$

where $\mathcal{I}(c, P(c))$ is the subjective importance of a claim given its truth, as determined by the user. Declaring "Dewey Defeats Truman" is more significant than an error reporting the price of corn futures–unless the user happens to be a futures trader. The truth of a claim also affects its importance; e.g. given the claim "tech stocks will be highly volatile over the next five years", an investor would be indifferent to this claim if true (since it is in line with market expectations) but would definitely want to know if it was false (since it would mean the risk penalty incorporated into the price of these stocks is undeserved).

3.3 Completeness

We are also concerned with how thorough collections of claims (and their providers) are; a reporter who reports the military casualties of a battle but ignores civilian losses cannot be trusted as a source of information about the war. While incompleteness is often symptomatic of bias, this is not always the case–it is possible to provide an incomplete view on a topic without attempting to sway the reader to a particular position. If a collection $\mathbf{C}$ purports to cover a topic t (e.g. "the war"), and $\mathbf{A}$ is the collection of all claims in the corpus, we can calculate completeness with respect to t as

$$\mathcal{C}(\mathbf{C}) = \frac{\sum_{c \in \mathbf{C}} P(c) \cdot \mathcal{I}(c, P(c)) \cdot \mathcal{R}(c, t)}{\sum_{c \in \mathbf{A}} P(c) \cdot \mathcal{I}(c, P(c)) \cdot \mathcal{R}(c, t)}$$

where $\mathcal{R}(c, t)$ is the $[0, 1]$ relevance of a given claim c to the topic t. Thus, completeness is the proportion of the topic's true, importance- and relevance-weighted claims in a given collection. A collection that omits true, important and highly relevant claims will have a low completeness, but omitting untrue, unimportant or irrelevant claims will have no effect.

3.4 Bias

Bias results from supporting a favored position with either untruthful statements or a targeted incompleteness ("lies of omission"). A single claim may also have bias depending on its representation; e.g. "freedom fighter" and "terrorist" can refer to the same person, but with very different connotations. Like truthfulness and completeness, the degree of bias depends on the user–a political conservative, for example, may find Fox News less biased than MSNBC, with the converse being true for a liberal. Here we will restrict ourselves to considering a finite, discrete set $\mathbf{S}$ of possible positions for the topic (e.g. pro- and anti-gun control), where $\mathcal{S}(c,s)$ is the $[0,1]$ degree to which a claim c supports a position $s \in \mathbf{S}$, and $\sum_{s\in\mathbf{S}} \mathcal{S}(c,s) \leq 1$. Let us also denote the *user's* support for each position as $\mathcal{S}(s)$, where $\sum_{s\in\mathbf{S}} \mathcal{S}(s) = 1$. Now we can calculate the bias of a single claim c and collection of claims $\mathbf{C}$ as

$$\mathcal{B}(c) = \sum_{s\in\mathbf{S}} |\mathcal{S}(s) - \mathcal{S}(c,s)|$$

$$\mathcal{B}(\mathbf{C}) = \frac{\sum_{s\in\mathbf{S}} |\sum_{c\in\mathbf{C}} P(c)\cdot\mathcal{I}(c,P(c))\cdot(\mathcal{S}(s)-\mathcal{S}(c,s))|}{\sum_{c\in\mathbf{C}} P(c)\cdot\mathcal{I}(c,P(c))\cdot sum_{s\in\mathbf{S}}\mathcal{S}(c,s)}$$

A collection of claims that, on the whole, matches the user's belief among the possible positions then has no bias, whereas a collection whose claims, on the whole, contradict the user's stance is held to have high bias. Notice that a collection need not contain claims that each match the user's belief in the positions to to be considered unbiased, but rather must be balanced so that, taken together (and weighted by truth and importance) these claims *collectively* support each position to the same degree the user does.

3.5 From Collections to Sources

From our metrics over single-topic collections of claims such as documents, we can calculate of trustworthiness of the information source $\mathbf{P}$ (such as an author or publisher) providing them: $\mathcal{X}(\mathbf{P}) = \frac{\sum_{\mathbf{C}\in\mathbf{P}} \mathcal{X}(\mathbf{C})\cdot\mathcal{W}(\mathbf{C})}{\sum_{\mathbf{C}\in\mathbf{P}} \mathcal{W}(\mathbf{C})}$, where $\mathcal{W}$ is the relative weight of each constituent collection (e.g. the importance of the collection's topic t to the user), and $\mathcal{X}$ is the measure of interest ($\mathcal{T}$, $\mathcal{C}$, or $\mathcal{B}$).

3.6 Relativity

Notice that we rely upon the subjective importance of claims to the *user*, as well as his stance on the positions of each topic. Trustworthiness cannot be assessed "globally", but–as we have seen–must be calculated with respect to each user's viewpoint to be meaningful. When evaluating the performance of an algorithm relative to the user's judgment, we have two choices: solicit the user's estimates of $\mathcal{S}(s)$ and $\mathcal{I}(c,\mathcal{T}(c))$, or obtain his opinion of the truthfulness, completeness and bias of each collection directly; the former method is generally more tedious, but better suited to large datasets where many collections share relatively few claims and topics.

4 USER STUDY

We wished to explore our new metrics, and specifically to contrast our new trust metrics with the simplistic trustworthiness scores calculated by existing trust systems, such as the arithmetic mean ($\mathcal{T}_m(\mathbf{C}) = |\mathbf{C}|^{-1}\sum_{c\in\mathbf{C}} P(c)$). To do this, we needed to evaluate these alternatives relative to human trustworthiness judgments over a given collection of claims.

4.1 The Article

We selected a news article from the English version of The People's Daily on the topic of the effectiveness of China's family planning policy (commonly referred to as the "one-child policy", although this is an inaccurate oversimplification). The People's Daily is of particular interest because it is operated by the Chinese Communist Party (CPC) and thus has a definite bias, and yet tends to be factually accurate. Consequently, we could expect $\mathcal{T}_m$ to perform poorly as a trustworthiness measure, as it is oblivious to bias and completeness and would award high trustworthiness scores to both the articles and The People's Daily on the whole, compared to the lower trustworthiness assessment we expected humans to assign. The title of the article is itself highly suggestive of this: "China's population policy draws wide praise", and its contents unsurprisingly highlight the purported benefits of 30 years of China's family planning policy to the Chinese people. Predictably, it completely ignores the large body of criticism that has been leveled at the policy, including accusations of forced abortions and infanticide, as well as unintended consequences such as sex-selective abortion and the resultant male to female gender imbalance.

4.2 Setup

Each user in our study was given the text of the news article, but *not* told its title, author or publisher to prevent this from prejudicing their trust assessments. We asked that they read the article while keeping the two possible positions ("China's family planning policy has been good for China" and "China's family planning policy has been bad for China") in mind. We then gave them a set of questions, one regarding their own position on the topic (expressed as a real value between the extremes of "China's family planning policy has been entirely bad for China" and "China's family planning policy has been entirely good for China"), six on their assessment of the article's overall trustworthiness, and 57 questions about 19 specific claims made by the article (3 for each claim). All answers were in the form of real numbers between 0 and 10 (inclusive). We had nine participants, all computer scientists.

Table 1: Survey results on the overall trustworthiness of the article

Question	Min	Max	Mean	Std Dev
What is your position on China's family planning policy? 0 = entirely bad for China 10 = entirely good for China	4	8	6.1	3.9
How trustworthy is this article as a source of information about the topic? 0 = completely untrustworthy 10 = completely trustworthy	5	10	7.4	5.2
How reliable is this article as a source of information about the topic? 0 = completely unreliable 10 = completely reliable	5	9	7.6	4.3
How would you rate this article as a source of information about the topic? 0 = worst possible article 10 = best possible article	3	8	6.1	4.3
How accurate was the information in the article? 0 = completely incorrect 10 = completely correct	5	9	7.2	4.6
How informative was the article with respect to the topic? 0 = wholly uninformative 10 = everything I needed to know	3	8	6.0	5.1
How biased was the article? 0 = completely objective 10 = completely biased	7	10	8.7	3.8

4.3 Overall Trustworthiness Assessments

The overall trustworthiness assessments given by respondents are summarized in Table 1. A number of interesting observations can be made based on these results:

- Participants on the whole felt that China's family planning policy was only mildly positive for China, though nobody thought very negatively about it.
- Participants gave similar trustworthiness and reliability scores, suggesting that these two concepts are roughly synonymous for users, although there was greater variance in the assessed trustworthiness.
- By contrast, when asked to assign a rating to the article, respondents were significantly less positive in their appraisal. This suggests that the criteria for overall rating differs from that for trustworthiness, possibly including additional factors such as writing style or placing greater emphasis on bias, or, more generally, that the perception of trustworthiness can exceed that of quality.
- Interestingly, the mean score for trustworthiness exceeded that for accuracy (truthfulness), informativeness (completeness), and unbiasedness (1 - bias), suggesting that users may be more generous in their assessment of overall trustworthiness than they are in its specific components.
- The perceived high bias did not seem to preclude a reasonable overall trustworthiness; this may indicate that the participants (highly educated computer scientists) consider themselves savvy enough to correct for bias themselves, or it may be an artifact of the survey's phrasing–asking how trustworthy the article is "as a source of information about the topic" may have implied that only the quantity and accuracy of the information, and not its bias, was to be considered.

4.4 The Claims

From the news article we extracted 19 claims, ranging from unimportant details ("Carl Haub is a senior demographer at the Population Reference Bureau") to broad claims that directly address the topic ("China's family planning policy has alleviated problems from overpopulation in China"). For each claim, we asked three questions:

1. To what degree do you believe this claim [0 = definitely not true, 10 = definitely true]?

2. How important would it be to you if this claim were true [0 = I wouldn't care at all, 10 = I would care a lot]?
3. How important would it be to you if this claim were false [0 = I wouldn't care at all, 10 = I would care a lot]?

Question 1 essentially asks for the user's estimate of the probability of the claim, P(c). The answers to questions 2 and 3 (A_2 and A_3) allow us to estimate $\mathcal{I}(c, P(c))$ as $P(c) \cdot A_2/10 + (1-P(c)) \cdot A_3/10$. We found respondents gave somewhat similar answers for these two questions (the mean difference was 1.2), although for the claim "China is the most populous country in the world" the difference was relatively large (means of 4.4 and 7.6, respectively) reflecting the value of surprise: if this claim were false, it would mean that a large number of trusted sources of information (such as major mass media outlets and textbook publishers) had been incorrect.

When we researched these 19 claims ourselves, we found sixteen to be true with high certainty. The remaining three claims were speculative (such as "India's population will surpass China's in 2040") and found to be reasonable, but obviously not provable. This suggests a mean accuracy of at least 84%, which would then also be the minimum $\mathcal{T}_m$ simple trust score reported by a trust system (assuming it determined the truth of the claims correctly). Our survey respondents, however, were limited to their background knowledge and the article itself, and did not conduct any additional research before making their accuracy assessment. Consequently, their average estimate of the claims' accuracy is a significantly lower 75% (7.5/10), reflecting their greater level of uncertainty.

4.5 Calculated Metrics

We estimated $\mathcal{R}(c, t)$ and $\mathcal{S}(c, s)$ for each claim, and, taken together with the surveyed values for $P(c)$ and $\mathcal{I}(c, P(c))$, we are able to calculate the truthfulness and bias of the document for each participant. Computing completeness, on the other hand, is not possible since we are looking at a single document rather than a corpus, and thus do not have $\mathbf{A}$ and do not know which claims we are missing. The question on informativeness instead captured this from the users directly, giving the article a mean of 6.0, which at first seems rather generous given the brevity and one-sidedness of the article. However, as respondents did not, on the whole, have a strong interest in China's family planning policies, only the broad details were important to them, and thus an absence of detail need not necessitate a low completeness.

We calculate a truthfulness of 0.77 and a bias of 0.58. The slightly higher truthfulness versus the simple mean accuracy of the respondents' claim beliefs (0.75) indicates that users were more confident in the truth value of those claims which were more important to them. Unfortunately, as the article is consistently factually accurate, there is little room for differentiation between these two measures here; a more interesting example would be an article that gets the important claims correct but fine details wrong. Our calculated bias, however, is much lower than than the bias assigned by the respondents, 0.87 (8.7/10). In this particular article, the coverage is clearly and uniformly one-sided, but our bias measure assumes that participants will perceive less bias when a collection of claims (on the whole) agrees with their position on the topic. This is reasonable when the bias level is moderate and somewhat subtle (e.g. in MSNBC and Fox News) but apparently does not hold when bias is extreme and blatant, as the user can no longer "ignore" it. Interestingly, when we calculate an "absolute" bias, setting $\mathcal{S}(s) = 1/|\mathbf{S}|$, we find a bias of 0.82, which is much more in line with our users judgement. This suggests, perhaps, that when absolute bias is high, it should be preferred to (or interpolated with) relative bias.

4.6 User Metric Preference

After conducting the survey and calculating our metrics, we wanted to know which trustworthiness scores users would actually prefer in practice: our new metric triple, or a scalar rating. The exact question we asked our survey participants was "which of these you think best capture the trustworthiness of the article you read in the survey (i.e. which set of metrics would be most helpful to you if you were researching China's family planning policy and were considering reading the article)?". We gave four choices. "The trustworthiness of the article is 7.4 (out of 10)", based on the mean overall trustworthiness given by the respondents, was preferred by 28%. "The trustworthiness of the article is 8.7 (out of 10)", selected as a moderately higher value, was preferred by 11%. "The trustworthiness of the article is 10 (out of 10)", based on the mean accuracy $\mathcal{T}_m$ (taking the three speculative claims as true), was (predictably) preferred by 0%. Finally, the composite "the truthfulness of the article is 7.7 (out of 10), the completeness of the article was 6 (out of 10), and the bias of the article was 8.2 (out of 10)", based on the calculated truthfulness, the mean "informativeness" rating assigned by respondents, and the calculated absolute bias, respectively, was preferred by the remaining 61% of respondents. Note that one respondent was undecided and so split his vote between the first and last choices. Overall, our respondents preferred the new metrics by a wide margin.

5 CONCLUSION

We have introduced three new metrics for measuring the trustworthiness of information sources: truthfulness, completeness, and bias, and shown that these are able to convey a more useful and more robust idea of how much (and in what way) an information source should be trusted than the current practice of presenting the user with a single, trust algorithm-dependent scalar value. By computing trustworthiness consistently across algorithms, we also enable direct cross-system performance comparisons, and the evaluation of computed trustworthiness against human judgment.

However, as our survey shows, human perceptions of trustworthiness are not simple to capture. Strikingly, survey participants assigned the news article a trustworthiness score that exceeded their opinion of its accuracy, completeness, and unbiasedness! We also found that perception of bias, in particular, was difficult to predict; the very high bias did not seem to be reflected to a significant degree in the overall trustworthiness assessment, and could not be estimated relative to the user as we had expected but instead required an objective estimate.

As our survey was limited in size and scope, further analysis of our metrics may be useful. Future work should approach an entire corpus to allow calculation (or at least estimation) of the set of all claims ($\mathbf{A}$) to allow completeness to be calculated directly, should use domain experts to ensure accurate estimates of $P(c)$, and should include articles with varying levels of accuracy and bias to allow for more thorough comparison.

Still, it is nonetheless already clear that predicting consistent, semantically well-defined components of trust is a dramatic improvement; an existing trust system might have assigned our newspaper article a perfect trustworthiness score based upon its factual accuracy, completely ignoring the incompleteness and bias that proved obvious to human readers. Our new metrics avoid this trap and allow us to express a more complete picture of an information source's trustworthiness.

ACKNOWLEDGEMENTS

The authors would like to thank our survey respondents for their participation. This research was partly sponsored by the Army Research Laboratory (ARL) (accomplished under Cooperative Agreement Number W911NF-09-2-0053). Any opinions, findings, and conclusion or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the ARL.

REFERENCES

Artz, D., and Gil, Y. 2007. A survey of trust in computer science and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web* 5(2):58–71.

Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30(1-7):107–117.

Dong, X.; Berti-equille, L.; and Srivastava, D. 2009. Integrating conflicting data: the role of source dependence. *Technical report, AT&T Labs-Research, Florham Park, NJ.*

Galland, A.; Abiteboul, S.; Marian, A.; and Senellart, P. 2010. Corroborating information from disagreeing views. In *Proceedings of the third ACM international conference on Web search and data mining*, 131–140. ACM.

Josang, A.; Marsh, S.; and Pope, S. 2006. Exploring different types of trust propagation. *Lecture Notes in Computer Science* 3986:179.

Kamvar, S.; Schlosser, M.; and Garcia-molina, H. 2003. The Eigentrust algorithm for reputation management in P2P networks. *WWW '03.*

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46(5):604–632.

Levien, R. 2008. Attack-resistant trust metrics. *Computing with Social Trust* 121–132.

Pasternack, J., and Roth, D. 2010a. Knowing What to Believe (when you already know something). In *Proc. the International Conference on Computational Linguistics (COLING).*

Pasternack, J., and Roth, D. 2010b. Lifted Fact-Finders. *(In submission).*

Yin, X.; Han, J.; and Yu, P. S. 2007. Truth discovery with multiple conflicting information providers on the web. *In Proc. of SIGKDD.*

Yin, X.; Yu, P. S.; and Han, J. 2008. Truth Discovery with Multiple Conflicting Information Providers on the Web. *IEEE Transactions on Knowledge and Data Engineering* 20(6):796–808.