Journal of Trust Management
a SpringerOpen Journal

RESEARCH Open Access

Two sides of the coin: measuring and communicating the trustworthiness of online information

Jason RC Nurse[1*], Ioannis Agrafiotis[1], Michael Goldsmith[1], Sadie Creese[1] and Koen Lamberts[2]

* Correspondence:
jason.nurse@cs.ox.ac.uk
[1]Cyber Security Centre, Department of Computer Science, University of Oxford, Oxford, UK
Full list of author information is available at the end of the article

Abstract

Information is the currency of the digital age – it is constantly communicated, exchanged and bartered, most commonly to support human understanding and decision-making. While the Internet and Web 2.0 have been pivotal in streamlining many of the information creation and dissemination processes, they have significantly complicated matters for users as well. Most notably, the substantial increase in the amount of content available online has introduced an information overload problem, while also exposing content with largely unknown levels of quality, leaving many users with the difficult question of, what information to trust? In this article we approach this problem from two perspectives, both aimed at supporting human decision-making using online information. First, we focus on the task of measuring the extent to which individuals should trust a piece of openly-sourced information (e.g., from Twitter, Facebook or a blog); this considers a range of factors and metrics in information provenance, quality and infrastructure integrity, and the person's own preferences and opinion. Having calculated a measure of trustworthiness for an information item, we then consider how this rating and the related content could be communicated to users in a cognitively-enhanced manner, so as to build confidence in the information only where and when appropriate. This work concentrates on a range of potential visualisation techniques for trust, with special focus on radar graphs, and draws inspiration from the fields of Human-Computer Interaction (HCI), System Usability and Risk Communication. The novelty of our contribution stems from the comprehensive approach taken to address this very topical problem, ensuring that the trustworthiness of openly-sourced information is adequately measured and effectively communicated to users, thus enabling them to make informed decisions.

Keywords: Information trustworthiness; Information quality; Trust metrics; Trust visuals; Decision-making; Social-media content; Risk communication

Introduction

We live in a world where the ability to access and publish information is practically considered a human right. The adoption of the Internet into our daily lives has facilitated this capability, leading to the creation of a global information marketplace that has become central to decision-making online and offline. There remain, however, significant concerns regarding the quality of online information, and thus its trustworthiness, that cannot be overlooked. Numerous real-life cases have highlighted this misinformation problem within social-media sites, and more recently we have even witnessed the severe impact which inaccurate information can have; that is, in the Boston bombings case, where an individual, after being wrongly identified as a suspect by a popular social site, was found dead [1]. Here, poor quality information resulted not only in an ill-judged decision, but a tragic loss of an innocent life.

To address the issues surrounding the quality and trustworthiness of online content, there have been a number of proposals focusing on various aspects of the problem. Agichtein *et al.* [2] for instance, propose a system that can automatically identify high-quality content items in question-and-answer networks using several contextual and intrinsic features. This drive towards automated assessment of social content can also be seen in Castillo *et al.* [3] as applied to analysing the credibility of Twitter data, and in Suzuki and Yoshikawa [4], who focus on evaluating editor and text features to determine the quality of Wikipedia articles. These proposals all draw on well-defined sets of sub-factors – e.g., provenance, reputation, competence, corroboration and recency – which tend to be indicative of quality and trust [5], and from these deduce useful metrics and approaches to arrive at a trustworthiness score that may be associated with the online content.

One aspect not covered by these and similar works, however, is the fact that trust, credibility and quality (and the sub-factors of which they are comprised) are intrinsically subjective, i.e., they can be perceived and interpreted in different ways, and arguably may have varying levels of importance depending on the user of the information or the context. One user may rely heavily on the reputation of its source in determining how much to trust content, while another user may be more concerned with how up-to-date it is. With this in mind, we believe that there is substantial value in allowing users of information-trustworthiness measures to influence the scores which they will receive from automated tools, so as better to represent their individual preferences and situation. User influence could be realised at several levels – as simplistic as disregarding some trustworthiness/quality sub-factors completely, or as complicated as defining ranges of weights (i.e., importance) for trust sub-factors depending on scenario and decision context.

Measuring information trustworthiness is only one half of the problem. Once trustworthiness scores have been calculated, it is crucial that they are appropriately communicated to users in such a way as to enable them to make well-informed decisions. Although not extensive, there has been some work on this task. In Idris *et al.* [6], for example, authors present a simple traffic-light system (with red, amber and green, thus drawing on the real-world metaphor) to convey quality. Adler *et al.* [7] take a finer-grained approach to presenting trust as they colour the text background of Wikipedia content – from white (high trustworthiness) to dark orange (low trustworthiness) dependent on how trustworthy their system deems that segment of the information to be. Although useful approaches, these proposals at times lack the strong foundation in Cognitive Science and HCI that is imperative to designing interfaces and visuals that will ultimately be effective in communicating trust. This risks the possibility of confusing users even more, thus resulting in bad decisions and ill-informed actions.

The aim of this paper is therefore to present our comprehensive, interdisciplinary approach to address the shortcomings highlighted above, and to support decision-making that takes advantage of online content. This article brings together and consolidates a number of our previous contributions, while also reflecting on their utility with the overall problem in mind. We begin in Section 2 by briefly recapping our policy-based approach to measuring the degree to which users should trust openly-sourced information (e.g., tweets, Facebook updates, and blog posts) – this takes into account information provenance, quality and infrastructure-integrity factors and metrics, along with the individual's preferences and situation. Section 3 then considers visual approaches for communicating trustworthiness, i.e., those that predate ours and our previous work on this topic, and subsequent evaluations and reflections. In Section 4, we enter the core of the paper, which focuses on radar graphs and their use as a visual mechanism for communicating detailed trustworthiness information effectively. The discussion here highlights their advantages and shortcomings, but also covers general points applicable to any technique to communicate trust via visuals. Finally, Section 5 concludes the article and presents directions for extending this work.

An approach to calculate information's trustworthiness

Assessing the quality and trustworthiness of information is not a new problem. Early articles such as Wang and Strong [8] (on data quality) and Chopra and Wallace [9] (on trust) have researched this at length and identified several sub-factors that could be used to assess and measure these aspects. Examples of sub-factors within these areas include the level of competence of a source, their reputation and authority, the recency of information, how well corroborated the information is, and even how information is presented. In more recent articles, these and other sub-factors have been applied to automatically analyse and rate the trustworthiness of content, with commendable degrees of accuracy; cf. Castillo *et al.* [3]. The shortcoming with these specific automated types of proposals, however, is the lack of explicit appreciation of the subjective nature of trust (and related concepts) thereby overlooking the importance of allowing a user to influence the final information-trustworthiness score that is presented to them.

To address this problem, we have devised a policy-based approach and framework through which the trustworthiness of information can be measured [10,11]. As it is a framework, we allow for the 'plug-in' of any of the variety of techniques currently used to measure sub-factors of trust, such as those by Agichtein *et al.* [2] and Castillo *et al.* [3]. Most importantly, this approach allows users to specify their preferences via policies, and then have these applied to trustworthiness calculations to guide the calculation of a final content trustworthiness score. This may mean attenuation or accentuation of the scores given to specific sub-factors (e.g., weighting the information's calculated recency score higher than its corroboration score), or indeed, changing the way that they are combined to arrive at the single trust score. This is somewhat similar to the idea of personalisation systems for news and other such information (e.g., [36]).

Broadly, we assess information trustworthiness in three domains: the provenance of information, its quality, and the integrity of the information infrastructure used to communicate the content from author to final consumer (hereafter, referred to as III). While there are several publications focused on the first two of these domains, the third has thus far largely been neglected. As such, we have outlined a preliminary technique for accounting for attacks detected against the information infrastructure (e.g., information poisoning) – practically limited to local technology deployed within the user's network – which takes into account known and unknown threats, vulnerabilities in local infrastructure, and probabilities of attacks [11]. This will allow users to attain a better understanding of the environment they are operating in as well as assisting them to decide whether or not to believe associated provenance or quality scores. At this point, our research in this area is still on-going as we aim to better comprehend infrastructure scopes, and automated linkage to vulnerability and attack databases (e.g., CVE [12], NVD [13] and CAPEC [14]), en route to proposing a viable III metric.

As we envisage our general approach and any developed tools to be used within organisations as well (e.g., an Emergency Operations Centre (EOC) trying to understand an on-going crisis situation using live Twitter and Facebook data), we also allow organisations to define policies that reflect their beliefs. For instance, an EOC may always want tweets from the BBC to be rated as highly competent or, equally, may prefer that its employees never listen to some blacklisted information sources. Of course, the danger with allowing this personalisation of trust scores – by user or organisation – is that calculated information trustworthiness values are highly biased and ultimately unreliable. It remains our opinion however, that users of this system will be more focused on the ground truth, i.e., what is actually happening regarding a topic, situation or scene, and therefore doctoring content's trust scores will only put them at a disadvantage; the adage, 'garbage in garbage out' applies here.

Figure 1 presents the approach, with a simple example.

As illustrated in Figure 1, the approach starts with openly-sourced information being fed into the system. This could be from any source, including content on Twitter,

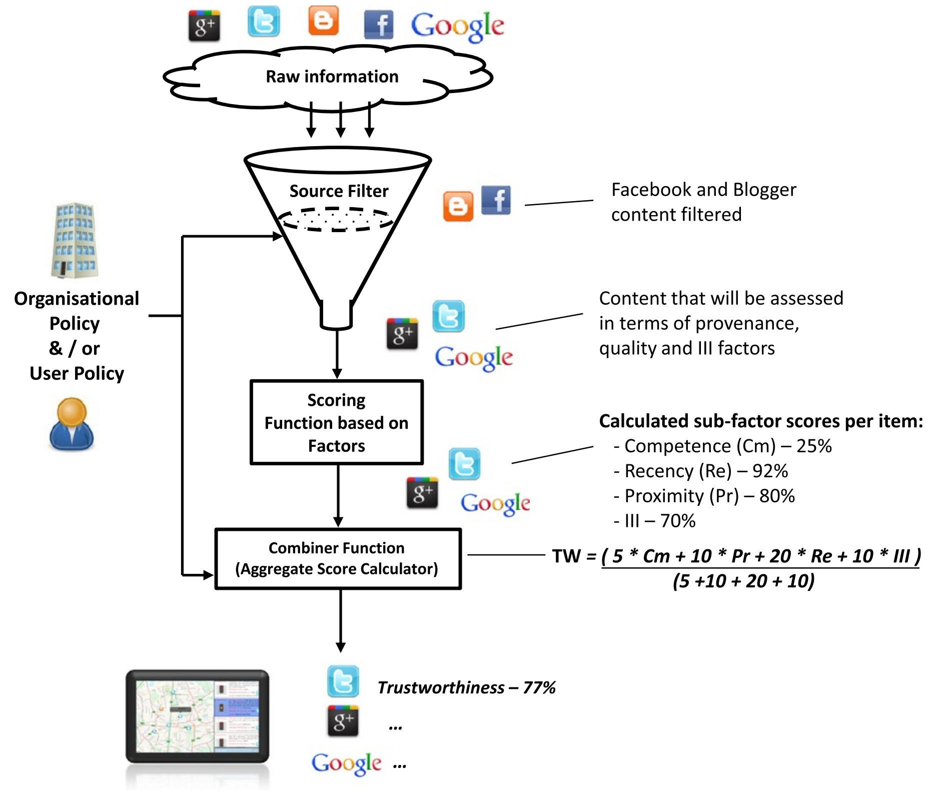


Figure 1 Policy-based approach to determining trustworthiness of openly-sourced information. This approach utilises an information filter, scoring function and combiner function, while also considering the organisational and user policies of its users.

Facebook and Google+, and could either be directly entered or gathered by a Web crawler or from an RSS feed. Information is then filtered according to the specific source, e.g., the tweeter on Twitter, or by source type, so as to remove content from unwanted sources. In the figure, the organisation, an EOC, has set a company-wide policy to block content from Facebook and Blogger, but allow content from Twitter, Google Plus and Google. Next, the filtered information (and any associated metadata) is assessed using several trust metrics to gain sub-factor ratings for each piece of information. These metrics can be drawn from existing work (e.g., [2,3,7]), or contributions by the efforts of our research consortium [11]. In the figure, we see examples of scores that can result for each information item: 25% for the Competence of a source (i.e., they have been found not to be very competent), 92% for information Recency (meaning the information is quite up to date), 80% for Proximity (i.e., the source is physically fairly close to the event of interest), and III of 70% (therefore although generally good, there may be some concerns about the integrity of the information infrastructure).

The approach then combines the sub-factor scores into a single trustworthiness score. Crucially, this combination can be influenced by the organisation's policy and any specific user policies that have been set beforehand [15]. We allow two types of configuration input at this stage. The first enables organisations and users to assign importance levels to the different sub-factors, to define their weight and ultimate impact on the final trustworthiness score. In the example in Figure 1, the user has assigned Competence to have a weight of 5; Proximity, 10; Recency, 20; and III, 10. Other research that we have been involved in [16] actually looks in detail at this notion of importance of factors and understanding people's perceptions; we expect that this could be used in the future to help specify importance policies. The second configuration available is the ability to select different types of combiner functions. In the illustration above we use the weighted arithmetic mean, however, there are numerous others including geometric mean, harmonic mean, and variations on root-mean-square (quadratic mean) that may be of use [17]. The advantage in having this range of functions available is that they all offer nuances that may be preferred by different users and contexts. In the case of crisis response, one could imagine an EOC preferring that trustworthiness scores associated with information items are slightly underestimated rather than overestimated, given that human lives are at stake. As a result, the harmonic mean would be a better choice than the quadratic mean because of its tendency to consistently underrate combined values [17].

After the policies have been applied and the combiner function executed, the output is a trustworthiness score for each piece of information. This score would have made use of the range of novel techniques to automatically measure quality and trust, but also would accommodate organisational and user preferences and context. In Section 3, we take the trustworthiness score, as it is here – a percentage – and consider whether and how visuals may be applied to allow for effective communication of the associated information risk.

Visuals for communicating trustworthiness

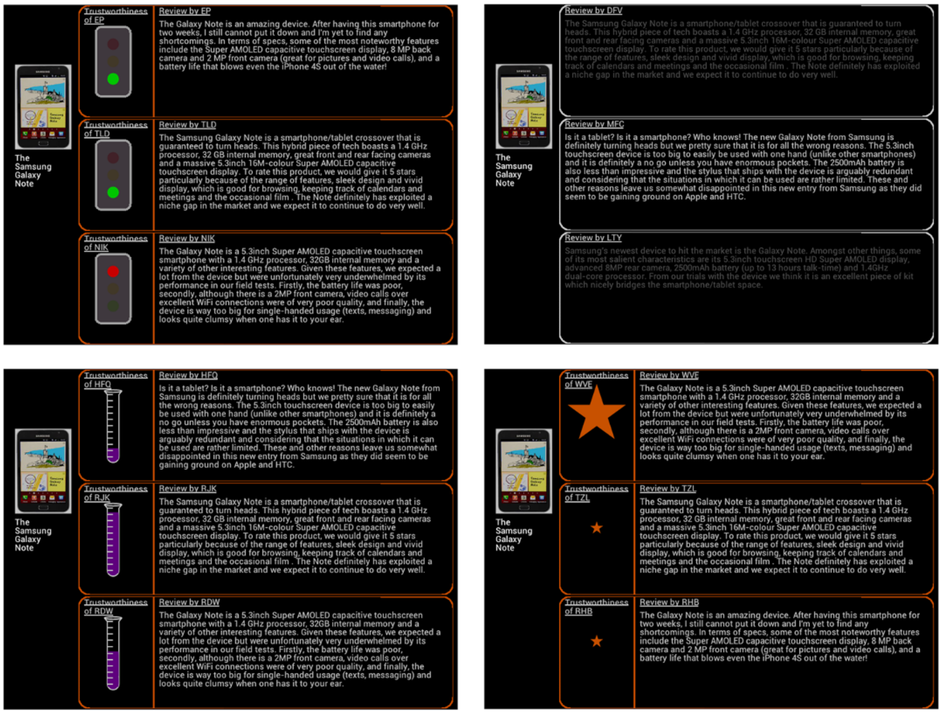
Over the last few years, there have been an increasing number of articles using visuals to communicate the quality and trustworthiness of information. In our review in

Section 1, we highlighted the work of Idris *et al.* [6] and the use of traffic-lights, and that of Adler *et al.* [7] that advocates changing the background colour of related text to indicate trust. Another notable approach is that of Chevalier *et al.* [18]. In their article, they concentrate heavily on visuals, particularly charts and graphs, as a support tool to assist information users to assess the quality of openly-sourced content. While these articles are indeed valuable, assumptions are often made regarding users and their cognitive abilities and preferences for interfaces, inclusive of visuals for trust. Unfortunately, ill-conceived assumptions may introduce other problems for users, such as information overload and confusion, which can negatively affect decision-making.

As a result of this, we have pursued a research agenda focused specifically on understanding ways in which information could be effectively communicated to users, in light of their cognitive abilities, perceptions and biases. Two well-established fields were central to this work: (1) Risk Communication, given that our aim of presenting users with trustworthiness data is in effect, to mitigate the risk of believing and acting on that information; (2) HCI and System Usability, for general guidance and principles of interface design. Having critically reflected on these and related domains, we engaged in several exercises to evaluate existing and new visual techniques for conveying information trustworthiness. Broadly, our investigative approach involved gathering a large number of visualisation techniques (e.g., iconography, glyphs, charts, and imagery), assessing them in terms of accepted Risk Communication and Usability guidelines and in the context of trust, and lastly, conducting user experiments to test a subset of the more useful visual approaches.

In detail, our user experiments tested four techniques for visualising the trustworthiness of information. These were: traffic lights (with red, amber and green indicating low, medium and high trustworthiness respectively), transparency (where highly trustworthy information would be shown normally, medium trustworthy content would be made 30% transparent, and low trustworthy content 70% transparent), stars (using size of star to indicate trustworthiness, the bigger, the better), and test-tubes (filled to portray trustworthiness – the fuller it was the higher the trust). As indicated, these all operated on three levels of trustworthiness, namely, high, medium and low; the assumption being that the percentage score calculated would have to be mapped to these levels. Figure 2 displays the visuals in the context of the experiment which was conducted on a tablet PC device; full details are available in [20].

From our analysis of the experimental results, traffic lights were found to be the most preferred technique for presenting trustworthiness, while transparency was the least preferred. Participants' choice of traffic lights was reportedly linked to modern man's instilled behaviour towards them and the colours they presented; in society today, a red traffic light is the universal sign to stop. The use of real-world metaphors is heavily advocated in Usability and Risk Communication principles but it was good to actually verify these recommendations. The poor performance by the transparency method was surprising, given analogous approaches have been used before [21] with some success. Feedback from individuals pointed to the difficulty in perceiving and understanding the degree of trustworthiness actually being conveyed, i.e., perceiving the difference between being 30% or 70% transparent. Moreover, if we were to allow five levels of trust as opposed to three – the positive there being further granularity to trust scores – it would be even more challenging for individuals to perceive. Of course, there would also be a challenge with regards to using traffic lights.



The Samsung Galaxy Note

Trustworthiness of EP

Review by EP

The Galaxy Note is an amazing device. After having this smartphone for two weeks, I still cannot put it down and I'm yet to find any shortcomings. In terms of specs, some of the most noteworthy features include the Super AMOLED capacitive touchscreen display, 8 MP back camera and 2 MP front camera (great for pictures and video calls), and a battery life that blows even the iPhone 4S out of the water!

Trustworthiness of TLD

Review by TLD

The Samsung Galaxy Note is a smartphone/tablet crossover that is guaranteed to turn heads. This hybrid piece of tech boasts a 1.4 GHz processor, 32 GB internal memory, great front and rear facing cameras and a massive 5.3inch 16M-colour Super AMOLED capacitive touchscreen display. To rate this product, we would give it 5 stars particularly because of the range of features, sleek design and vivid display, which is good for browsing, keeping track of calendars and meetings and the occasional film . The Note definitely has exploited a niche gap in the market and we expect it to continue to do very well.

Trustworthiness of NIK

Review by NIK

The Galaxy Note is a 5.3inch Super AMOLED capacitive touchscreen smartphone with a 1.4 GHz processor, 32GB internal memory and a variety of other interesting features. Given these features, we expected a lot from the device but were unfortunately very underwhelmed by its performance in our field tests. Firstly, the battery life was poor, secondly, although there is a 2MP front camera, video calls over excellent WiFi connections were of very poor quality, and finally, the device is way too big for single-handed usage (texts, messaging) and looks quite clumsy when one has it to your ear.

The Samsung Galaxy Note

Review by DFV

The Samsung Galaxy Note is a smartphone/tablet crossover that is guaranteed to turn heads. This hybrid piece of tech boasts a 1.4 GHz processor, 32 GB internal memory, great front and rear facing cameras and a massive 5.3inch 16M-colour Super AMOLED capacitive touchscreen display. To rate this product, we would give it 5 stars particularly because of the range of features, sleek design and vivid display, which is good for browsing, keeping track of calendars and meetings and the occasional film . The Note definitely has exploited a niche gap in the market and we expect it to continue to do very well.

Review by MFC

Is it a tablet? Is it a smartphone? Who knows! The new Galaxy Note from Samsung is definitely turning heads but we pretty sure that it is for all the wrong reasons. The 5.3inch touchscreen device is too big to easily be used with one hand (unlike other smartphones) and it is definitely a no go unless you have enormous pockets. The 2500mAh battery is also less than impressive and the stylus that ships with the device is arguably redundant and considering that the situations in which it can be used are rather limited. These and other reasons leave us somewhat disappointed in this new entry from Samsung as they did seem to be gaining ground on Apple and HTC.

Review by LTY

Samsung's newest device to hit the market is the Galaxy Note. Amongst other things, some of its most salient characteristics are its 5.3inch touchscreen HD Super AMOLED display, advanced 8MP rear camera, 2500mAh battery (up to 13 hours talk-time) and 1.4GHz dual-core processor. From our trials with the device we think it is an excellent piece of kit which nicely bridges the smartphone/tablet space.

The Samsung Galaxy Note

Trustworthiness of HFQ

Review by HFQ

Is it a tablet? Is it a smartphone? Who knows! The new Galaxy Note from Samsung is definitely turning heads but we pretty sure that it is for all the wrong reasons. The 5.3inch touchscreen device is too big to easily be used with one hand (unlike other smartphones) and it is definitely a no go unless you have enormous pockets. The 2500mAh battery is also less than impressive and the stylus that ships with the device is arguably redundant and considering that the situations in which it can be used are rather limited. These and other reasons leave us somewhat disappointed in this new entry from Samsung as they did seem to be gaining ground on Apple and HTC.

Trustworthiness of RJK

Review by RJK

The Samsung Galaxy Note is a smartphone/tablet crossover that is guaranteed to turn heads. This hybrid piece of tech boasts a 1.4 GHz processor, 32 GB internal memory, great front and rear facing cameras and a massive 5.3inch 16M-colour Super AMOLED capacitive touchscreen display. To rate this product, we would give it 5 stars particularly because of the range of features, sleek design and vivid display, which is good for browsing, keeping track of calendars and meetings and the occasional film . The Note definitely has exploited a niche gap in the market and we expect it to continue to do very well.

Trustworthiness of RDW

Review by RDW

The Galaxy Note is a 5.3inch Super AMOLED capacitive touchscreen smartphone with a 1.4 GHz processor, 32GB internal memory and a variety of other interesting features. Given these features, we expected a lot from the device but were unfortunately very underwhelmed by its performance in our field tests. Firstly, the battery life was poor, secondly, although there is a 2MP front camera, video calls over excellent WiFi connections were of very poor quality, and finally, the device is way too big for single-handed usage (texts, messaging) and looks quite clumsy when one has it to your ear.

The Samsung Galaxy Note

Trustworthiness of WVE

Review by WVE

The Galaxy Note is a 5.3inch Super AMOLED capacitive touchscreen smartphone with a 1.4 GHz processor, 32GB internal memory and a variety of other interesting features. Given these features, we expected a lot from the device but were unfortunately very underwhelmed by its performance in our field tests. Firstly, the battery life was poor, secondly, although there is a 2MP front camera, video calls over excellent WiFi connections were of very poor quality, and finally, the device is way too big for single-handed usage (texts, messaging) and looks quite clumsy when one has it to your ear.

Trustworthiness of TZL

Review by TZL

The Samsung Galaxy Note is a smartphone/tablet crossover that is guaranteed to turn heads. This hybrid piece of tech boasts a 1.4 GHz processor, 32 GB internal memory, great front and rear facing cameras and a massive 5.3inch 16M-colour Super AMOLED capacitive touchscreen display. To rate this product, we would give it 5 stars particularly because of the range of features, sleek design and vivid display, which is good for browsing, keeping track of calendars and meetings and the occasional film . The Note definitely has exploited a niche gap in the market and we expect it to continue to do very well.

Trustworthiness of RHB

Review by RHB

The Galaxy Note is an amazing device. After having this smartphone for two weeks, I still cannot put it down and I'm yet to find any shortcomings. In terms of specs, some of the most noteworthy features include the Super AMOLED capacitive touchscreen display, 8 MP back camera and 2 MP front camera (great for pictures and video calls), and a battery life that blows even the iPhone 4S out of the water!

Figure 2 Screenshots of the tool's prototype interface for visually communicating information trustworthiness using traffic lights, transparency, test-tubes and stars.

Other general points worth noting include the fact that transparency appears to subconsciously direct participants away from low trustworthy content (a potentially good feature depending on where the tool is deployed and the level of subconscious persuasion desired); different visuals may elicit varying interpretations from users (i.e., even though participants were told beforehand about the three levels of trust, some perceived a low test-tube as worse than a red traffic light); and often, individuals desire more information on trustworthiness scores (i.e. how it was determined, what factors led to it). It is the last of these points that has given rise to the work presented in the next section.

Communicating trustworthiness with radar graphs

As mentioned above, a key finding in our experiments was that users would like more information about how the trustworthiness score was calculated by the system, to help them determine how confident they could be in the trust score presented. From a human trust perspective, this is perfectly reasonable, as understanding is a common antecedent to trust [5]. To assist users with building confidence in trust scores, we engaged in two experiments. The first experiment explored people's perception of a set of trust factors using radar graphs as visuals. Figure 3 shows the interface that was presented to study participants. This allowed us to assess: (i) whether individuals could understand the detailed factors that they were requesting – we found that they could; (ii) their ability to perceive and comprehend radar graphs to deduce trust – generally, also positive conclusions were drawn; and (iii) the existence of variations in the levels of importance

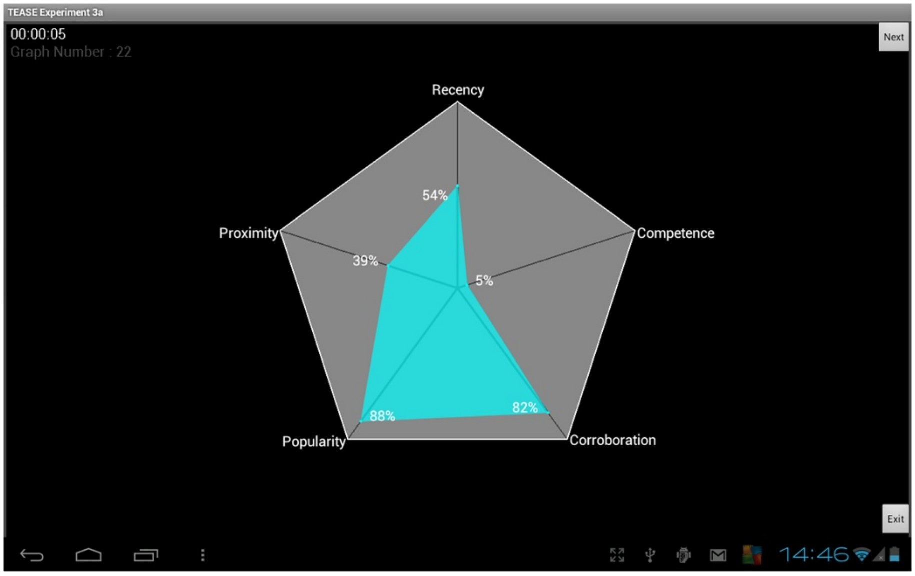


Figure 3 Radar graphs application interface displaying five detailed information-trustworthiness sub-factors: Recency, Proximity, Popularity, Corroboration and Competence.

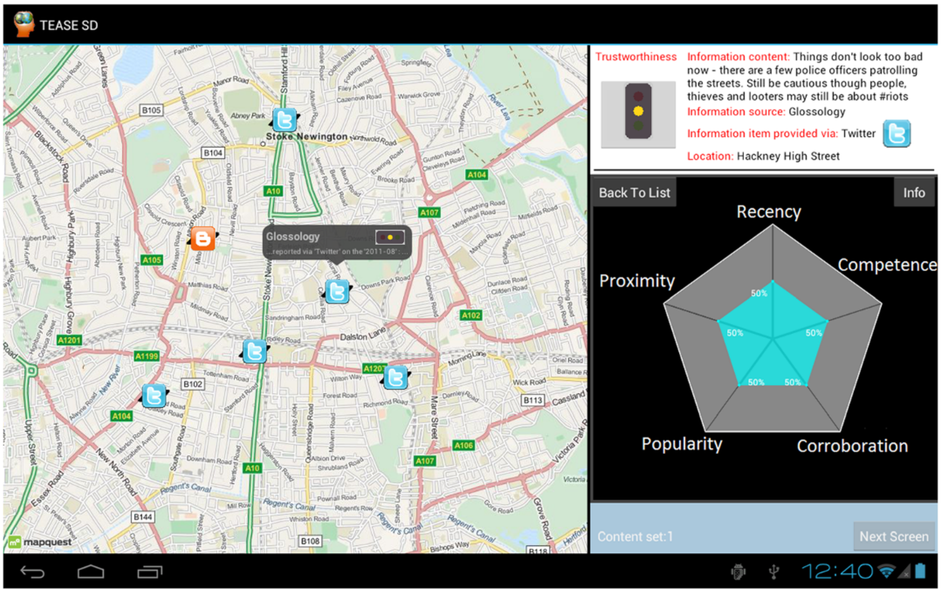
attached to trust sub-factors – findings did point to a clear difference, with Competence being the most important of the five and Popularity being the least important [22].

The second experiment, which built on the first, sought to use radar-graph visuals to allow users to 'drill-down' into trustworthiness scores shown initially as traffic lights. The context of the experiment was crisis-response and users had the task of making decisions on the level of risk in a scenario containing openly-sourced information with associated trust scores; a screenshot of the interface used is shown in Figure 4. Trustworthiness was visualised at three levels: at the highest level, there was a traffic light symbol; the second level showed radar graphs with associated measures for provenance, quality and III (users could tap the traffic light to drill down to this level of detail); and the third level displayed a radar graph with the actual factors used to calculate each of these measures. The purpose of allowing drill-down to radar graphs was to build confidence in the higher-level trust scores, where and when necessary; a goal that was achieved according to the results from the study [16].

One crucial finding from these experiments, especially the first, was that people's perceptions of particular graph schemas could actually be quite varied and also subject to visual biases. To understand why this was the case and when it was likely to occur, we focused on these graphs and explored the nuances in schemas that might have led to these differences in perceptions. Below, we describe that work in detail and reflect on key findings that will be useful for any future work on visualisations for trustworthiness.

Experiment background and context

The aim of our radar graphs experiment was to explore people's perception of five trust factors, which our previous findings suggest are key to trust online [5,22], and also to assess their importance to individuals in so far as they pertain to judgements on trustworthiness. The factors were: Competence (*Cm*), the level of knowledge of a person or



Trustworthiness

Information content: Things don't look too bad now - there are a few police officers patrolling the streets. Still be cautious though people, thieves and looters may still be about #riots

Information source: Glossology

Information item provided via: Twitter

Location: Hackney High Street

Figure 4 Crisis management application interface that allows the tool's users to drill-down into trustworthiness scores (initially represented by traffic lights) to get more detail on how scores were conceived. This detail is presented in the context of the trustworthiness sub-factors of Recency, Proximity, Popularity, Corroboration and Competence.

information source; Proximity (*Pr*), the geographical closeness of a source to an event of interest; Popularity (*Po*), how well-known is a source; Recency (*Re*), how recent or up-to-date is information to the event of interest; and Corroboration (*Cr*), how well supported the information is by a variety of different sources. The experiment design consisted of 200 radar graphs, each presenting ranges of values between 0-100% for the five factors; Excel's *RAND()* function was used to produce a set of 200 random values which was then implemented in our application for experimentation. In Figure 3, we show an example visual, which displays Graph #22; with Recency at 54%, Competence at 5%, Corroboration at 82%, Popularity at 88%, and Proximity at 39%. Further below we present graphs exactly as they were presented to study participants, to give readers the best insight into the conditions and tasks of the experiment.

A total of 40 individuals (29 females, 11 males, mean age of 23.7, age range: 18–58 years) participated in the study. Recruitment was conducted through the use of flyers posted within the University of Warwick and the University of Oxford. Participants were from a variety of disciplines (sciences, humanities, and social sciences) and there was also a diversity of levels, i.e., students were both postgraduates and undergraduates, and working professionals spanned from hospitality clerks to personal assistants, researchers and administrators. Participants were compensated for assisting with the experiment.

The experiment consisted of participants being presented with each of the 200 graphs and then given a maximum of 10 seconds before they were asked to give a rating of 0–100 (100 being the maximum trustworthiness) to represent the level of trustworthiness the graph conveyed to them. A timer was displayed on screen and therefore participants were always aware of the time remaining. The restricted time allowed served two purposes. Firstly, we were aiming to get participants' first and instinctive impression, and secondly, we would decrease the chances of study participants recalling how they assessed similarly shaped graphs, thereby avoiding them simply using their memory.

To present the graphs to participants, we used a Motorola Xoom tablet PC. At the beginning of the experiment, participants were briefed on the goals of the study and requested to sign a consent form. To ensure that they had a clear understanding of the five trust factors, they were also shown short definitions and examples of how the terms could be used. As there were a large number of graphs, participants were advised that in case of discomfort (e.g., tired eyes), they were free to take a break at any time. Regarding the responses to the graphs, it was emphasised that there were no correct or incorrect answers. Once participants were comfortable, the experiment commenced, and they were asked to evaluate the trustworthiness degree represented by each of the 200 radar graphs by assessing measures of the five factors included and any personal preferences they held.

The findings from the analysis, as reported initially in [16], were very encouraging and highlighted distinct significance levels of factors across participants. Specifically, we used linear regression analysis to identify importance (via coefficients) for each factor per participant, and then averaged across the sample to define a regression formula for the group of participants. The formula for trustworthiness (as a function of the five factors) that resulted is shown below:

Equation (1) presented in [16]:

$$\begin{aligned} \mathit{Trustworthiness} \; = \; & -5.425 \; + \; 0.176\,Re \; + \; 0.405\,Cm \; + \; 0.235\,Cr \\ & + \; 0.127\,Po \; + \; 0.141\,Pr \end{aligned}$$

Coefficients preceding each factor were taken to define factor importance. Thus, Competence was the most influential factor, followed by Corroboration, Recency, Proximity and Popularity. Finally, it is worth mentioning that participants did not report any significant difficulties in understanding the factors (or their relation to trustworthiness), graphs or combining them to deduce an overall trust score. This built our confidence in the findings above.

Detailed radar graph analysis

In addition to the more general evaluation in [16], we have subsequently engaged in several smaller and more focused statistical analyses pertaining to graphs and respective participants' scores. In the first investigation, we conducted a basic relationship analysis to verify that values produced by the formula derived from the sample above (hereafter, 'expected values') correlated with participants' trustworthiness scores for each of the 200 graphs. This was to establish some link between expected values and participants' actual scores rather than an influence or causal association; although both types of values would have the underlying trust-factor scores as a basis. Another key benefit was also the potential to identify any outlying participants (i.e., those with different opinions than the general populous) that might have been marginalised after averaging across coefficients to define the trustworthiness formula for the sample.

The analysis consisted of first computing expected values (using the formula in (1) and respective factor scores) for all of the graphs, and running a Pearson product–moment correlation [23] to determine the relationship between these values and each of the 40 participants' graph scores. As an example, in Table 1 we present the details of

Table 1 Data table which shows the graph data (i.e., scores for Recency, Competence, Corroboration, Popularity, and Proximity), expected trustworthiness values (based on the trustworthiness formulae proposed in equation (1)) and actual trustworthiness scores provided by two participants in the study

Graph #	Re	Cm	Cr	Po	Pr	Expected	P1	P2
22	54	5	82	88	39	42	40	40
33	1	14	12	30	9	8	9	5

two graphs, their respective trust-factor scores, the expected value for the graphs and the actual scores given by Participants #1 (P1) and #2 (P2).

From the analysis conducted, we found a positive correlation between the expected and actual scores, with all Pearson correlation coefficients statistically significant at $p < 0.001$. This confirmed a link between these two values and suggested that there were no extreme outliers (i.e., participants with very contrary opinions and perceptions of factor importance). To reinforce our findings and focus more on an evaluation of the similarity of expected and given scores, we also calculated the *deviation* from the expected value for every graph, per participant. The equation we used is presented below, where n is the number of the graphs, x is the score that the participant gave to the graph and E is the expected value for the specific graph. The findings from this calculation did highlight some deviation across participants; however, the results were supportive of the similarity of values in general.

Equation (2):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - E_i)^2}$$

From the correlation analysis, we did identify a participant with a Pearson coefficient of *0.353*, which was much smaller than the average across the sample of *0.745*. Not surprisingly, the same participant had the highest *deviation* value, a score of *24.49*, as compared to the sample average of *15.93*. Upon further investigation, we found that this participant viewed the importance of trust factors very differently, with Recency being most important (*0.385*), followed by Proximity (*0.184*), Corroboration (*0.162*), Competence (*0.135*) and then Popularity (*0.081*). This variation could therefore explain the lowest Pearson coefficient and the highest deviation value. Generally however, the strong correlation in other scores did emphasise the utility of the main formula presented in (1).

From our in-depth analysis of graphs and scores assigned to them (in addition to deviation analyses mentioned above), we were able to identify a set of graphs that were persistent outliers where some participants allocated much higher or lower scores than expected. Our further assessment therefore focused on understanding the cause of these outliers (i.e. graphs where the expected and actual score differ more than 50%) in an attempt to comprehend their nature and, if possible, mitigate their future effects.

Elaborating on the outliers, we noted that in 4 graphs, at least 40% of participants agreed in their reporting of scores greater than 50% different to expectations. Specifically: for Graph #3, 53% of participants provided scores 50% or more greater than the expected value; for Graph #45, 45% of individuals gave scores of at least 50% smaller; for Graph #52, 40% of study participants supplied scores of at least 50% greater than expected; and for Graph #66, 45% of participants reported scores of 50% or more smaller than the expected values. We noted two potential reasons for these significant variations. Either, there was a marked deviation in factors' importance (i.e., coefficients) for those groups of participants, or possibly the overall size or area of the radar graph (i.e., the filled part of the pentagon) or indeed, its schema, influenced individuals. An example of the latter case is that smaller-sized graphs, independent of specific factor importance levels, subconsciously swayed participants to give lower scores than normal for the sample.

To ascertain which of these (if any) may have been true, we first calculated the average coefficient value for each trust factor, for the four groups of participants that gave the differing scores to the graphs. One should recall that coefficients resulted from linear regression analysis on each participant's 200 graph scores. Next, we compared these coefficients (i.e., levels of importance) to those of the main sample. This assessment highlighted several variations in factors' importance, to the extent that it could have been the reason for differing final graph scores.

In Graph #66 (in Figure 5) for instance, where 45% of participants gave scores 50% or more lower than expected, we found that these participants felt that Recency and Popularity were respectively 11% and 17% less important than in the overall sample formula and Corroboration was 9% more important. Reflecting on the graph itself, one can begin to understand why a lower score from those participants might therefore have resulted. That is, the factors with higher graph values were less important and those with lower values (here, Corroboration only) were more important. Similar observations were apparent with the other three graphs. This was a notable finding that pointed to deviation in factors' importance as the actual reason for different (or outlier) scores, and potentially not subconscious influences of graph size. A possible solution to mitigate the effect of this specific outlier could be to allow users to customise the importance of the five factors according to their trust perceptions.

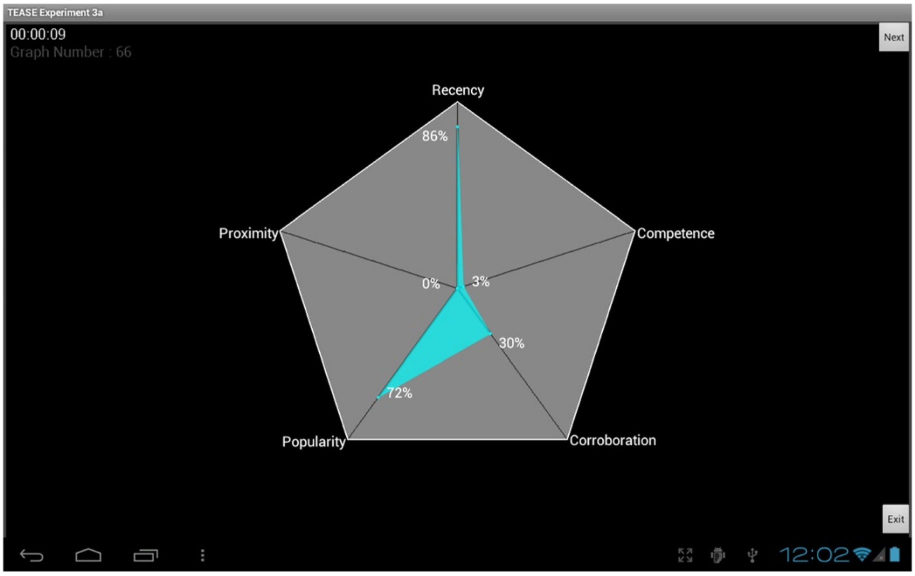


Figure 5 **Screenshot of Graph #66, where the high percentage of Recency among extremely low percentages of Competence and Proximity results in a very small volume.**

Apart from comparing participants' scores to the expected values for the sample, another more focused approach to determine whether size of graphs subconsciously influenced participants was to compare their individual scores to their expected values (i.e., values calculated based on their own trustworthiness formulae). This was therefore moving away from assessing outliers as it pertains to the expected score across the sample, to identifying and evaluating outliers relating to the respective participant's expected graph score. For this investigation, therefore, we used the formula generated for each participant based on the linear regression analysis, and compared the new expected values to the actual scores that were given to the respective graphs. To gather an idea of whether graph size or area might have influenced participants, we checked for cases where actual scores were 50% or more different to the expected scores, *and* size of the graph (or more accurately, the calculated area of the filled graph) was at least 25% different to the expected value. To calculate the filled area, we used the equation presented below in (3). This splits the pentagon into five triangles, calculates the blue-shaded area in each, sums these, then divides by the total pentagon area to determine the size percentage filled.

Equation (3):

$$\frac{{}^{1}/_{2} \times \sin 72 \times (Re \times Cm + Cm \times Cr + Cr \times Po + Po \times Pr + Pr \times Re)}{{}^{5}/_{2} \times 100 \times 100 \times \sin 72}$$

Therefore, if a participant gave an actual graph score 65% greater than their expected value for that graph, and the graph size was 40% bigger than that expected score, we hypothesised this to mean that size may have had some impact on their decision to award that higher graph score.

The results from this analysis indicated that there were a significant number of cases – i.e., 49% of the cases where there was larger than a 50% difference – where this situation occurred, which led us to believe that graph size may have had some noteworthy impact on participants' scores. Considering the situations where this transpired, we became interested in the respective graphs and whether a particular type of schema (i.e., graph arrangement) or area may have resulted in under or over-estimations in scores. This could give us valuable further insight into how graphs were perceived by participants in addition to allowing us to look in more detail at the outliers present.

In testing for this, we found that for a number of the cases particularly small sizes did feature. The graphs most underestimated by participants, and therefore those where size may have had a real influence, were Graph #45 shown in Figure 6 and Graph #66 previously presented in Figure 5; 17 participants underestimated these graphs. Assessing Graph #45's schema, the very low values for Recency, Popularity and Competence led to a particularly small graph size with considerably thin filled areas for even the higher rated factors of Proximity and Corroboration. Graph #66 exhibited a similar size phenomenon just with different trust factors. The other graphs which may have been particularly influenced by graph area are documented in Table 2; incidentally, all of these graphs result in underestimations of trustworthiness scores. Readers can easily recreate the graphs to view their schemas as necessary.

Further to the test above, we also considered the possibility that extremely low levels of Competence (generally the most important factor) may have been the cause of low

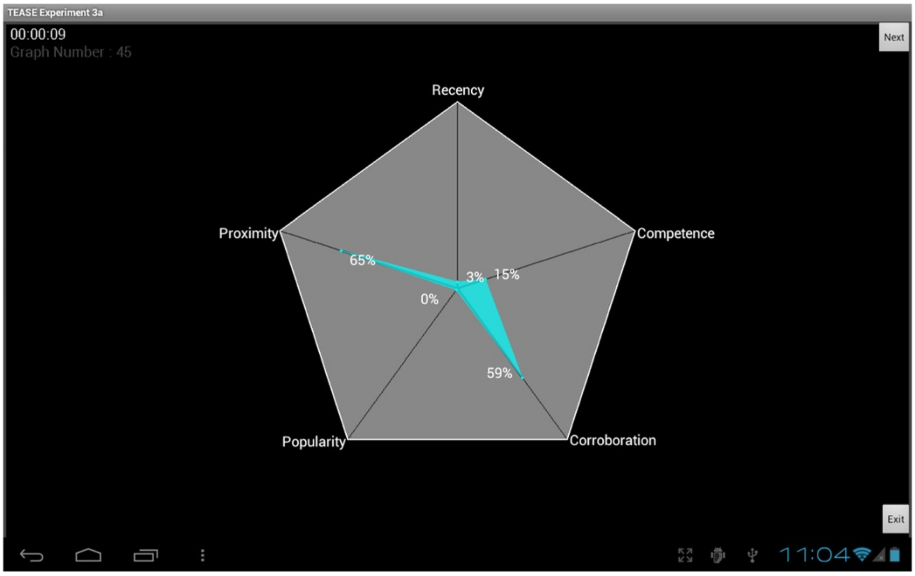


Figure 6 **Screenshot of Graph #45, where factors with high percentages are adjacent to factors with low percentages, resulting in a very low volume.**

scores across the participants. This would be easily conceivable, as a low score in the most important factor may have led to individuals completely discounting the graph. As can be seen in Graph #2 in Table 2 and Figure 7 however, even in cases where Competence was extremely high, there was some level of understatement in participants' scores. The outliers identified in this case will require further research to fully characterise and mitigate their effects.

The final cause that we considered as a potential source of outliers was the time constraint of 10 seconds within which participants had to provide their answers for each graph. We observed, however, that all the participants would respond within the first five to seven seconds, thus rendering the effect of time probably insignificant.

Next steps in using radar graphs and other visuals for trust

There are several avenues for further research which we intend to pursue, in order to follow up on the analyses and exploratory findings in terms of communicating

Table 2 Graphs (and their associated values for Recency, Competence, Corroboration, Popularity, and Proximity) where there is a possibility that graph size may have impacted trustworthiness scores given by study participants

Graph #	Re	Cm	Cr	Po	Pr	Number affected by size	Expected value	Average deviation of scores from the expected value
2	13	94	20	33	46	9	51	21
15	21	0	31	70	15	9	17	16
33	1	14	12	30	9	11	8	5
106	19	14	30	0	32	10	15	8
143	63	0	2	57	37	10	19	18
148	63	2	25	25	1	11	16	16
150	35	1	73	35	7	9	24	20

We also show the expected trustworthiness value and the average deviation of scores given by participants to that value.

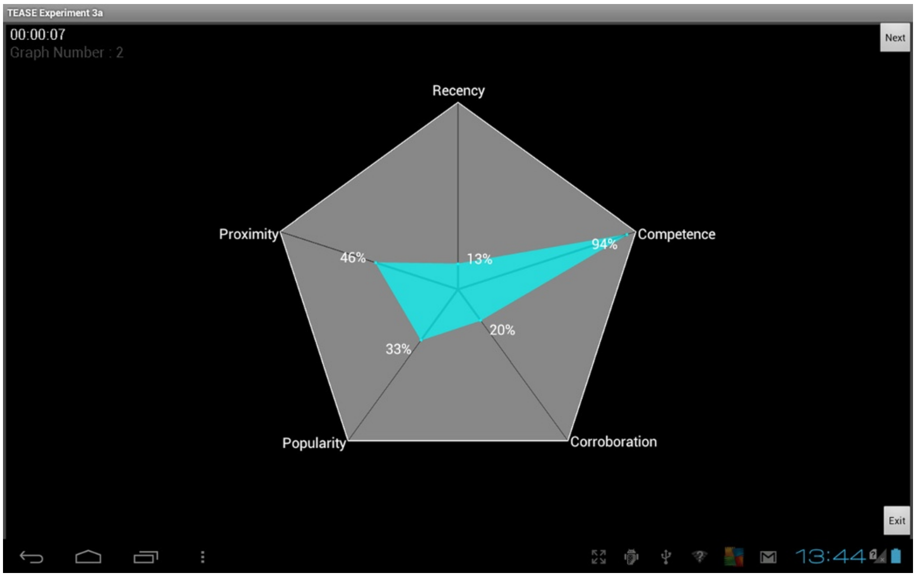


Figure 7 Screenshot of Graph #2, where an extremely high value on Competence is among very low values of the adjacent factors Recency and Corroboration, resulting in underestimations due to low volume of the graph.

trustworthiness especially using radar graphs. The first pertains to the layout of the graph itself – i.e., the position and ordering of each trust factor and its axis – and what happens when the positioning or ordering is changed. To take Graph #66 as an example, if Recency (86%), Popularity (72%) and Corroboration (30%) were next to each other, the graph's schema would be noticeably different and the area would jump from 5% to 17%. Figure 8 shows a mocked-up example of this altered format.

The research question therefore remains, might these variations in ordering have an impact on participants' perceptions and the scores that they award? Or, do people see

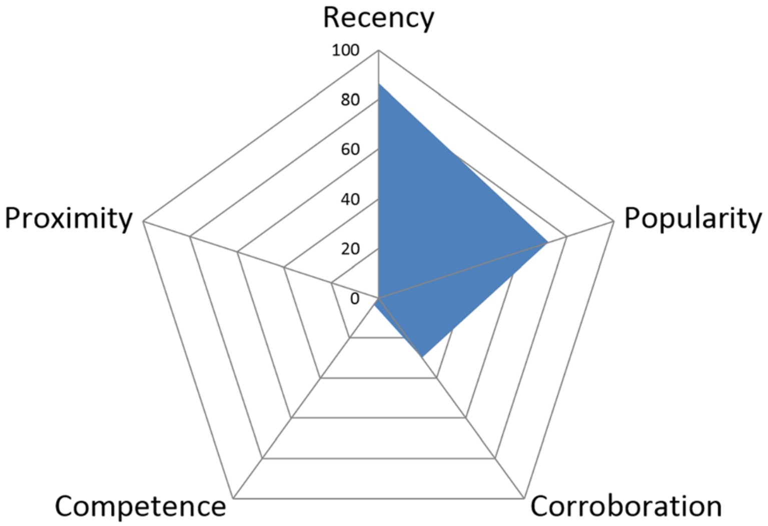


Figure 8 Graph #66 with factors ordered in such a manner to maximise the volume of the graph. Notice the difference in volume with Figure 5, where the factors with the same values are ordered differently.

past the area and schema differences and award similar scores regardless of ordering? If ordering does have an impact then we would need to consider whether there are any 'better' orderings or indeed, whether the choice of ordering should be left to system users to decide. Finding the optimal ordering solution may mitigate the outlier effect caused by the graph size as well, thereby potentially rendering future attempts to communicate trustworthiness as effective instead of problematic, as they seem to be now.

We have already started to consider orderings that would maximise graph size, where the graph might be unduly underestimated by some individuals because of its schema. Graph factors ordered in the following way may result in a consistently maximised graph area: largest factor, second largest factor, fourth largest factor, fifth largest factor and then third largest factor (i.e., 1-2-4-5-3). We need to be careful however, because this approach could lead to consistent overestimations in the trustworthiness of graphs and related information, since the participants will always view a maximum area for the graph. Moreover, users might become confused with the constant change of the position of the trustworthiness factors in the edges of the pentagon; usability will also therefore be an issue.

One possibility is to assess whether placing the trust factors with the same ordering in the radar graph (i.e., 1-2-4-5-3) but focusing on the importance of the trustworthiness factors to the participants. This approach could allow the system to provide higher volume in graphs that should be trusted and lower volumes in graphs that should not be trusted. Unfortunately, a side effect of this is that it leads to system biases, not to mention the initial task of ascertaining how important individual factors are to users (although, this could arguably be expressed by them at system setup). Nonetheless, as mentioned above, continued work on this and other approaches will be necessary.

Another avenue for follow-up research picks up on a feedback point from interviews with participants who have undertaken the graphs experiment. There were suggestions that we should investigate the potential of additional axes in radar graphs to represent more factors. Our current emphasis on 5 factors was based on simplicity and reducing cognitive effort required by system users (a guideline from the Risk Communication and Usability fields), but as several research articles have suggested (e.g., Miller [24] and Saaty and Ozdemir [25]), the human brain may be capable of coping with possibly 7 or 9 items. Future experiments might therefore seek to evaluate people's ability and desire to assess additional trust factors – of which there are many [5] – and assess whether there is as preferred number of factors that should be displayed. This could also allow us to further validate our existing work and investigate the importance of other factors as it pertains to trustworthiness. Dependent on the research available, we may need to conduct a broader study on the importance of factors outside of graphs and then use a subset (e.g., the most important) for display within graphs during decision-making. In terms of graphs, the field of Risk Communication (especially seminal research work such as Lipkus and Hollands [19]) will undoubtedly continue to be a key area of reference during our study design and subsequent user experimentation.

An area where we may also conduct further research is to consider the impact of *context* on individuals, and assess the influence it has on people's perception of trust factors and their importance both within and outside of visuals and graphs. For the previous experiments, we specifically avoided context as a variable because we were more focused on how people perceived the importance of those factors generally. In the real-world however, as has been highlighted in existing research [26,27], context is crucial and is likely to have a notable influence on the importance of factors and perceived trust. This was also hinted at in interviews with participants as some said that they may have allocated different scores to the same graphs if they were presented with different scenarios. It will also be intriguing to investigate whether individuals' perceptions concur regarding factors' importance levels within specific contexts.

Finally, in terms of radar graphs, we aim to assess the impact that cognitive biases may have on the perception of graphs within our experiments. Cognitive biases refer to systematic weaknesses in human's cognitive processing and have been discussed at length in several articles [28-30]. The bias of most immediate interest is the *anchoring effect*, which defines the tendency of decision-makers to systematically base judgements on initial (and potentially even irrelevant) information [30]; future decisions are 'anchored' or biased to that starting information. We would therefore be specifically investigating (using quantitative – i.e., scores – and interview-based approaches) whether such an effect is prevalent when individuals score graphs, and if it is, to what extent. Key questions include, assessing whether we could find clear links between this and incorrect perception of certain graphs. A good reference point for this information is Furnham and Boo [31] and their recent, comprehensive review on the anchoring effect, its causes and attempts to tackle it in the past.

Conclusions and future work

The proliferation of information in online environments has rendered the designing of tools able to support decision-making using this information more crucial than ever. Towards this end, exploring the notion of trustworthiness of information is a vital step. Many researchers investigating this multidimensional concept have thus far focused either on proposing novel information-trustworthiness metrics, or on visuals through which quality and trustworthiness scores can be presented to users.

In this article, we have taken a lateral approach, synthesising a number of our previous contributions and presenting our comprehensive approach to measuring *and* visualising trust, in an attempt to address 'both sides of the coin'. We first recapped our policy-based approach which evaluates the degree of trustworthiness which users should place in online information. This is comprised of a number of quality and trust factors and respective metrics, and of the preferences of decision-makers, i.e., the users of the system. In addition to the application of policies, our approach also benefits from allowing any metric to be plugged into our framework to measure quality and trust. As these techniques mature, therefore, our approach will become more efficient and accurate. This will result in better support for system users when drawing on online information to assist decision-making online and offline.

Presenting a method to calculate trustworthiness addresses only one dimension of the problem at hand. To address the other main issue, i.e., effective communication of information-trustworthiness scores, we have taken steps towards understanding how best to communicate trustworthiness to individuals, by exploring and experimenting with various visualisation techniques including traffic lights, transparency, stars, and more recently, radar graphs. Apart from presenting and discussing our general two-pronged approach to the problem, this paper focused in detail on these radar graphs and evaluated their use as techniques for effectively and efficiently communicating trustworthiness information. We were also able to use graphs to assess the importance of trust factors to ranges of test participants, while also identifying challenges to communication realised via inconsistencies in human perceptions (in schemas, etc.). Most importantly, these irregularities could also apply to other visual techniques, and so wider research should note and consider these issues. Section 4.3 outlined several steps that we intend to take to further these discussions and address how misperceptions could be handled across ranges of similar visualisation approaches.

Briefly commenting on the use cases for our system, we believe that there are several of interest. The most compelling one, however, is that of crisis management and response, as alluded to in the various experiments conducted. Here, responders would be able to use a tool implemented on a tablet device, for instance, en route to or at a crisis scene. This could inform them of related content, who is posting content in or about the area, and how trustworthy that content is (likely strongly influenced by an Emergency Operations Centre's overarching organisational policy). Other cases include supporting decision-making when doing online research for the purchase of a new device or searching for quality information about an emerging topic.

Reflecting on our work in general, there are underlying issues which raise further challenges to be addressed, some practical and others requiring further novel research. One practical challenge is that we assume that the information upon which our approaches are applied is readily consumable. While this may be the case in some situations, there may be practical difficulties in gathering, parsing and reading information in an automated fashion, e.g.: How are topics searched? Is the system real-time? Is information persisted? Typically, we envisage that our system assessing trustworthiness will reside in the user domain and therefore, initial queries could dictate what information to gather and when to access it. There are solutions suggested that constantly scan sources (mainly focusing on social media) that are of specific interest [32,33], searching for information containing particular words. These will be considered in future work, as well as the notions of real-time monitoring, parsing and rating of information. There is also the possibility of drawing on advances in Natural Language Processing (NLP) and Semantic analysis to give further insight into information content, what is being said, and what exactly is meant.

Regarding the processing of the information and acquiring the necessary data to determine the score for the trust and quality factors, proposed methods for measurement rely heavily on the metadata (e.g. timestamps to determine information's recency, or geo-tags to determine the proximity of a source to a reported event). There can be cases, however, where such metadata is absent (the dearth of geo-tagged tweets is a good example) or has been maliciously tampered with, in the case of direct attacks on our system, for instance. For malicious tampering, we hope that the III score assigned to information will be able to reflect this, but where metadata is completely absent, this will be more difficult to handle. There are some approaches we have been considering to address this gap, such as that proposed by Sultanik and Fink [34], and this and other techniques are directly within the scope of future work.

Catering for users' needs when faced with information of unknown quality, albeit of crucial importance, is not the only research area where our approaches could be applied. We envisage adapting the knowledge from our research in understanding and visually communicating trust and quality to explore whether it could be beneficial to other disciplines, such as Cybersecurity; an initial exploration into that problem is presented in [35]. Opportunities have arisen in that field in particular because of the increase in attacks which target human weaknesses, both perceptual and cognitive. If systems and interfaces could be designed that better take these weaknesses into account and also understand what makes users trust or ignore security warnings and messages, this should lead to better security decisions and behaviour.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors listed contributed significantly in the research and experimentation leading to this article, and the preparation of the manuscript itself. All authors read and approved the final manuscript.

Acknowledgements

This work was conducted as a part of the TEASE project, a collaboration between the University of Oxford, University of Warwick, HW Communications Ltd and Thales UK Research and Technology. The project was supported by the UK Technology Strategy Board's Trusted Services Competition (www.innovateuk.org) and the Research Councils UK Digital Economy Programme (www.rcuk.ac.uk/digitaleconomy).

Author details

[1]Cyber Security Centre, Department of Computer Science, University of Oxford, Oxford, UK. [2]Department of Psychology, University of York, York, UK.

Received: 7 October 2013 Accepted: 5 March 2014
Published: 20 May 2014

References

1. BBC (2013) Falsely accused student of Boston attacks confirmed dead. In: Falsely accused student of Boston attacks confirmed dead. http://www.bbc.co.uk/news/world-us-canada-22297568. Accessed 30 Sept 2013
2. Agichtein E, Castillo C, Donato D, Gionis A, Mishne G (2008) Finding high-quality content in social media. In: Proceedings of the International conference on Web search and web data mining. pp 183–194, ACM
3. Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World Wide Web. pp 675–684, ACM
4. Suzuki Y, Yoshikawa M (2012) Mutual evaluation of editors and texts for assessing quality of Wikipedia articles. In: Proceedings of the 8th Annual International Symposium on Wikis and Open Collaboration. ACM
5. Nurse JRC, Rahman SS, Creese S, Goldsmith M, Lamberts K (2011) Information Quality and Trustworthiness: A Topical State-of-the-Art Review. In: Proceedings of the International Conference on Computer Applications and Network Security (ICCANS). IEEE
6. Idris NH, Jackson MJ, Abrahart RJ (2011) Colour coded traffic light labeling: A visual quality indicator to communicate credibility in map mash-up applications. In: Proceedings of International Conference on Humanities Social Sciences, Science & Technology
7. Adler BT, Chatterjee K, De Alfaro L, Faella M, Pye I, Raman V (2008) Assigning trust to Wikipedia content. In: Proceedings of the 4th International Symposium on Wikis. ACM
8. Wang RY, Strong DM (1996) Beyond accuracy: What data quality means to data consumers. J Manag Inform Syst 12(4):5–33
9. Chopra K, Wallace WA (2003) Trust in electronic environments. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences. p 10–19, IEEE
10. Rahman SS, Creese S, Goldsmith M (2012) Accepting information with a pinch of salt: handling untrusted information sources. In: Security and Trust Management (pp. 223–238). Springer, Berlin Heidelberg
11. Nurse JRC, Creese S, Goldsmith M, Rahman SS (2013) Supporting Human Decision-Making Online Using Information-Trustworthiness Metrics. In: Human Aspects of Information Security, Privacy, and Trust (pp. 316–325). Springer, Berlin Heidelberg
12. MITRE (n.d.) Common Vulnerabilities and Exposures (CVE). http://cve.mitre.org/. Accessed 30 Sept 2013
13. NIST (n.d.) National Vulnerability Database (NVD). http://nvd.nist.gov/. Accessed 30 Sept 2013
14. MITRE Common Attack Pattern Enumeration and Classification. http://capec.mitre.org/. Accessed 4 Sept 2013
15. Helfert M, Foley O, Ge M, Cappiello C (2009) Limitations of weighted sum measures for information quality. In: Proceedings of the 15th Americas Conference on Information Systems
16. Nurse JRC, Agrafiotis I, Creese S, Goldsmith M, Lamberts K (2013) Building Confidence in Information - Trustworthiness Metrics for Decision Support. In: Proceedings of 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-13). IEEE
17. Agarwal B (2007) Programmed Statistics. New Age International Ltd, New Delhi.
18. Chevalier F, Huot S, Fekete JD (2010) Wikipediaviz: Conveying article quality for casual Wikipedia readers. In: IEEE Pacific Visualization Symposium (PacificVis). pp 49–56, IEEE
19. Lipkus IM, Hollands JG (1999) The visual communication of risk. JNCI Monographs 1999(25):149–163

20. Nurse JRC, Creese S, Goldsmith M, Lamberts K (2012) Using Information Trustworthiness Advice in Decision-Making. In: Proceedings of the International Workshop on Socio-Technical Aspects in Security and Trust (STAST) at the 25th IEEE Computer Security Foundations Symposium (CSF-2012), (pp. 35–42). IEEE
21. Bisantz AM, Stone RT, Pfautz J, Fouse A, Farry M, Roth E, Nagy AL, Thomas G (2009) Visual representations of meta-information. J Cognit Eng Decis Making 3(1):67–91
22. Nurse JRC, Agrafiotis I, Creese S, Goldsmith M, Lamberts K (2013) Communicating Trustworthiness using Radar Graphs: A Detailed Look. In: Proceedings of 11th International Conference on Privacy, Security and Trust (PST-2013), (pp. 333–339). IEEE
23. Howell DC (2011) Fundamental statistics for the behavioral sciences, 7th edn. Wadsworth Publishing Company, Belmont, CA.
24. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):81
25. Saaty TL, Ozdemir MS (2003) Why the magic number seven plus or minus two. Math Comput Model 38(3):233–244
26. Kelton K, Fleischmann KR, Wallace WA (2008) Trust in digital information. J Am Soc Inf Sci Technol 59(3):363–374
27. Marsh S, Basu A, Dwyer N (2012) Rendering unto Caesar the Things That Are Caesar's: Complex Trust Models and Human Understanding. In: Proceedings of 6th IFIP International Conference on Trust management (IFIP'TM 2012). Springer, Berlin
28. Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. Science 185(4157):1124–1131
29. Croskerry P (2002) Achieving quality in clinical decision making: cognitive strategies and detection of bias. Acad Emerg Med 9(11):1184–1204
30. Peters E, McCaul KD, Stefanek M, Nelson W (2006) A heuristics approach to understanding cancer risk perception: contributions from judgment and decision-making research. Ann Behav Med 31(1):45–52
31. Furnham A, Boo HC (2011) A literature review of the anchoring effect. J Socio Econ 40(1):35–42
32. Indiana University Bloomington (2012) Truthy - Information diffusion research, http://truthy.indiana.edu. Accessed 30 Sept 2013
33. Streams K (2012) LazyTruth Chrome Extension Fact Checks Chain Emails. http://www.theverge.com/2012/11/14/3646294/lazytruth-fact-check-chain-email. Accessed 30 Sept 2013
34. Sultanik EA, Fink C (2012) Rapid Geotagging and Disambiguation of Social Media Text via an Indexed Gazetteer. In: Rothkrantz L, Ristvej J, Franco Z (eds) Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM-2012, vol 190. pp 1–10
35. Nurse JRC, Creese S, Goldsmith M, Lamberts K (2011) Trustworthy and Effective Communication of Cybersecurity Risks: A Review. In: Proceedings of the International Workshop on Socio-Technical Aspects in Security and Trust (STAST) at the 5th International Conference on Network and System Security (NSS-2011), (pp. 60–68). IEEE
36. Streibel O, Alnemr R (2011) Trend-based and reputation-versed personalized news network. In: Proceedings of the 3rd International Workshop on Search and Mining User-generated contents at the 20th ACM Conference on Information and Knowledge Management. pp 3–10, ACM

doi:10.1186/2196-064X-1-5
Cite this article as: Nurse *et al.*: Two sides of the coin: measuring and communicating the trustworthiness of online information. *Journal of Trust Management* 2014 **1**:5.