

## POSSIBLE SOLUTIONS FOR DRAWBACKS OF WORLD WIDE WEB

Marek ŠAJNA, Karol MATIAŠKO

Department of Informatics, Faculty of Management Science and Informatics,  
University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic, e-mail: {marek.sajna, karol.matiasko}@fri.uniza.sk

### ABSTRACT

*In this paper we present some of the drawbacks of World Wide Web connected with increasing amounts of information published in it and propose solutions such as semantic metadata, personalization and trust metrics. Greater emphasis is given to trust metrics as it is subject of our current research. Several definitions and graph formalization are discussed as well as approaches to determining values of direct and indirect (transitive) trust. In the end need for more systematization is stressed along with discussion of some transitive trust issues.*

**Keywords:** trust, trust metrics, transitive trust, Semantic Web, personalization, World Wide Web

### 1. INTRODUCTION

World Wide Web as we know it today has undergone – similarly to other technological phenomena – several phases in its evolution. Changes which led to its current shape were for the first time observed at the turn of the 21<sup>st</sup> century. At the same time, term Web 2.0 began to appear denoting this different nature of the Web (1), its interaction with the users and cooperation of users themselves. Since it is sometimes criticized for being more of a marketing term, we will simply refer to current state of the Web. On the other hand we will use term Web 1.0 or talk about static web when referring to the former phase. However we would like to emphasize that there are some phases which can be recognized, even though the distinction between them is not clear and changes are being done gradually.

To bring more confusion to this segment, term Web 3.0 started to appear in recent years (2). One of its main characteristics is simplified access to information for machines. This could be achieved in several ways and so the definitions of Web 3.0 vary.

Our aim in this article is to offer several views on quality of information in World Wide Web and ways how to increase it or determine its level. Flaws in the quality are discussed in section 2, while section 3 focuses on possible solutions for them. Section 4 is focused on trust metrics, one of those solutions and paper is concluded by prospects of future research.

### 2. DRAWBACKS OF WORLD WIDE WEB

As Web originated in the academic ground, its first users were drawn from these circles and its first content was mainly written by the experts in their field. The amount of information was limited, but growing and its quality was possibly the best among all the phases. But step by step, Web was discovered by more and more companies as a new ground for presenting and publishing old media and therefore the webpages held resemblance to books, newspaper, journals, etc. However investments to infrastructure and understanding of tools needed to contribute were too high for majority of users.

Transition from Web 1.0 to the current state can be observed from two points of view, technological and

sociological. Technical progress has driven the social changes, which followed with certain delay. Important characteristic from this point of view is that Web 1.0 was static, whereas Web as we know it nowadays is highly dynamic. As users were able to comment on news stories or subscribe to discussion boards, considerable part of them was attracted to do so. Many were also able to provide full-featured webpages, establishing themselves as authors of content on the Web. These changes caused a rise of ratio of information producers to consumers. With increasing amount of information, more drawbacks began to appear. Several critics argue that lack of professionalism in some of the new information producers brought overall quality of content down.

#### 2.1. Classification

In general we can observe four categories of problems originating in vast amount of information. First of all, it is an overload, which results from the amount of information itself. It is impossible for a human being to absorb and comprehend such quantities. For instance, when searching for some facts on the web, user is presented with many possible results from many different sources. She is unable to examine all of them and eventually ends up with the information closest to the requested one from few top results. However such information can suffer flaws mentioned in this section. Since as described this problem especially applies to humans, it represents a difficulty for machines too.

Other cases of poor quality are false information, outdated information and lack of integrity. When the content is provided by somebody who is not an expert in the field, there is increased probability that credibility of such information will be disputable (3). As mentioned, the information can be simply false. This can be caused either by the fact that person providing the information does not bear enough knowledge or has provided misleading information on purpose. Such malicious behaviour can be anticipated in decentralized mass of data which Web is. Nevertheless it is vital to mention that damage from using information of poor quality is the nearly identical independently from the good or bad purpose of the publisher of the content.

Another relevant issue of World Wide Web are out-dated pieces of information. Users many times encounter information of disputable validity, just because its author has not provided any temporal context, e.g. period of validity. Out-dated information can obviously be classified as false, but we want to stress that pieces of information given into context – whether temporal or spatial – provide for less doubts about their validity. Damage done by using obsolete information is comparable to the previous case. This problem is becoming more prevalent with more content added to the Web as many of it is published without period of validity or at least publication date.

Users of the Web are sometimes confused by contradictory or partly disagreeing information. Such flaws can be filed under lack of integrity. False and out-dated information are considered to be more local properties, whereas problems with integrity are mostly global, in sense that they affect several disagreeing pieces of information or even several web locations.

### 3. PROPOSED SOLUTIONS

In this section we will sum up three concepts, which we examined as possible solutions to before-mentioned drawbacks. They can be used to enhance user experience by pointing out the quality information and include semantic information, personalization and trust metrics.

Pieces of semantic information are slowly beginning to appear in the World Wide Web. They are part of a concept referred to as Semantic Web, which is seen as one of forms of Web 3.0. It was proposed by Berners-Lee et al. (4) in 2001 and widely theoretically adopted, but not implemented in greater scope and though its feasibility was not completely proven. Berners-Lee's target with the Web was to create a decentralized platform containing universal knowledge. Semantic Web should open it to computer users and thus machines would be able to gain information and infer new relations from it. This is apparently opposed to today's web where information is mainly composed for humans. Machines are indeed also able to read it, but mainly by simulating human behaviour. However in Semantic Web semantic information attached to resources published on the web allows machines easier access to information they require for example to help people organize their lives. This process can have many forms, but these are beyond the scope of this article.

We can expect that general integrity will increase with more pieces of semantic information in the Web. Another important fact is that semantic information is giving intuitively more credit to the content and in addition putting it into some context, e.g. temporal or spatial. When person or machine is able to acquire the period of validity of some information published on the web, there is reduced chance that they will use incorrect information.

Personalization, as another concept mentioned here, can have many definitions and can be applied to several areas of human life. Our interpretation involves information technologies, which are adjusting themselves to person's own needs and expectations. Many such technologies have already appeared and are being used. They are especially effective in reducing general

information overload laid on a human user in contemporary World Wide Web.

Example of such behaviour is actually not hard to find, many major search engines are adjusting their results to provide information that their users were really looking for. Two people, for instance, from different countries can have two different meanings in mind when they search for the same phrase. This technology is sometimes referred to as context-aware search and mentioned example can be accounted for different cultural context.

However personalization can as well be understood in scope of person's social connections and can lead us to trust metrics. There is a certain intuitive idea behind the concept that users are more confident about information obtained from people that are in their social circle and people they trust (5). This is related to phenomenon of homophily (6) which declares that similar individuals – such as friends, colleagues or partners – associate with each other more often than other members of population.

Over past few years, many services providing recommendations for products, vendors, culture events or even web pages have appeared and gained popularity. Such service is referred to as recommender and relies typically on trustworthiness of those who recommend. Trust in general is a measurement of confidence one person has in another and can be observed from different angles and computed by different techniques and algorithms. Most basic of them calculate trust for a user as one static variable (7). Shortcoming of such calculations is that it is not taking into account that different people can feel different confidence in each user.

### 4. TRUST METRICS

Our research is nowadays focused on trust metrics which consider that different entities can have different attitude and confidence in particular entity. Such an intuitive thing as trust is difficult to formalize and thus we encountered several definitions. According to Gambetta (8) *trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [we] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [our] own action.* However simpler and probably better definition was provided by Li in (9), where he states that *trust can be thought of as the level of belief established between two entities in relation to a certain context.* Essential attributes of trust derived from the latter definition are that it can be represented as level of belief and is limited to some context. Context of trust is sometimes referred to as purpose or topic.

Since we need to maintain data about entities and relations between them, mathematic graphs as data structure are natural choice. Apparently nodes represent the entities or agents, subjects of trust. Edges model acquaintance of corresponding entities, which means they know each other or better say, have some defined knowledge about trustworthiness of one another. Because we defined trust as some level, it is suitable to weight the edges appropriately. As people can have mutually different trust in each other, graph has to consist of directed edges.

Form taken by the weights depends above all on the nature of trust metrics itself. It can be for instance presented in form of a scalar real number drawn from some interval. Typical widely employed interval is  $[-1, 1]$ , where apparently  $-1$  represents absolute distrust,  $0$  neutral attitude and  $1$  absolute trust. Another used interval  $[0, 1]$  can represent the same range from complete distrust to complete trust or can ignore the distrust part of spectrum and represent only neutral-to-trust scope. It depends on the applied model which interval is used. Value of trust can as well be taken from the range of integer numbers, typically  $[1, 10]$  or  $[1, 5]$  (10).

There are also several techniques which use vectors to represent trust. At first this can be caused by more precise modelling, i.e. paying attention to more attributes. The second purpose for using vectors can be effort to maintain different contexts. Context was mentioned several times regarding trust and can affect structure of the graph as well. The idea is to use one value for each context of trust between vertices connected by respective edge. Another possibility is to use multiple edges between those two vertices. The way in which selected solution affects the algorithms used to calculate trust is part of our following research.

#### 4.1. Direct trust value

When discussing trust between two vertices in the graph, we can talk about two cases depending on whether they are connected or not. Trust between two connected vertices is called direct in contrast with indirect trust between vertices without connecting edge (11). Indirect trust is sometimes referred to as inferred or merged. We will use the term transitive trust.

In this section we will discuss important decision about direct trust – how to measure it. There obviously is a possibility to let people themselves define level of trust they put in each of their acquaintances. Although such explicit judgment can bring relatively accurate results, it can be very time-consuming and requires re-evaluation every time situation between the two entities changes. Therefore we find automatized solutions more suitable.

In relation to the homophily concept mentioned earlier, we can calculate trust according to characteristics of individual entities, especially those which they have in common. When discussing human users, it can mean involvement in similar cultural or sports activities, frequency of their mutual communication (12) or similar purchasing habits. Many e-commerce sites use recommender systems based on recommendations from people that bought same items.

Since there are many parameters of similarity of human attitudes, trust value will be result of a weighted sum of them. Although only having some similar attitudes to live is not an indication of mutual trust. Trust can be influenced by series of experiences, which change its value from time to time and establish it as a dynamic parameter (13). This concept can be demonstrated on two entities, A and B. A relies on advice given by B, but is disappointed with the actual result thus gets negative experience and its trust in B will adjust accordingly. Trust is therefore a function of time and some authors suggest

automatic decrease of its value during periods without positive experiences.

#### 4.2. Trust transitivity

As mentioned in the previous section, there is direct and indirect trust and the latter is discussed in the following lines. Question at hand now is how to determine value of trust between people who are not directly acquainted in the network. Fig. 1 illustrates transitive trust with three members in a portion of a network. Thick full lines model direct and thick dashed line indirect trust.

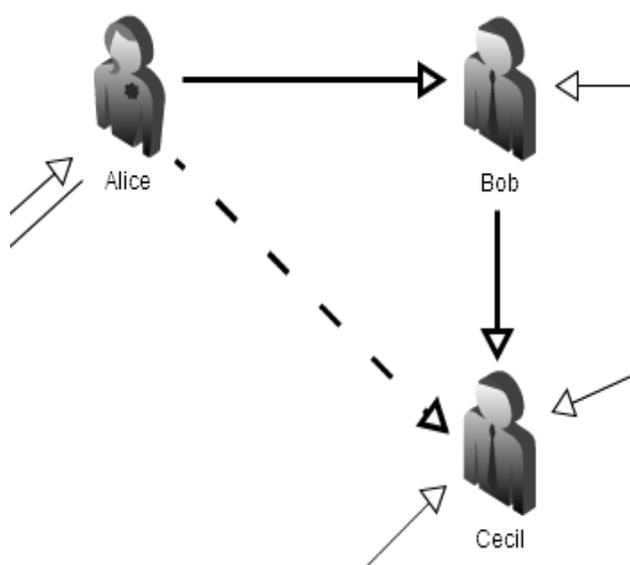


Fig. 1 Transitive trust between Alice and Cecil

Various research works have been done in the field of trust transitivity with nearly every one having its own terminology and hierarchy of sub-problems. We present two sub-problems: calculation method and realization of calculation. Several operations can be cast on examined network to get paths to examined entity and this is called calculation method. It includes operations such as shortest paths (13), maximum flow (14), breadth-first search or even enumerating all the paths between two involved vertices (15).

There are also various possibilities in scope of realization of calculation, which deals with how partial trust values along found path or paths are transformed into resulting transitive value. Beside others, this includes multiplication or means such as arithmetic or harmonic. Combination of several approaches is also a possibility. It is vital to state that purpose of trust has to stay the same along whole path.

Richters and Peixoto describe (13) an algorithm which uses weighted mean of best paths to all in-neighbours of examined vertex. The examined vertex is left out of the network in the process of calculation. Best path is the one with highest product of individual trusts. However “best path” can have different meaning in other algorithms or models.

#### 4.2.1. Related terminologies

Authors of (15) and (16) use different terminology to describe transitive trust. They identify two general tasks: concatenation and aggregation. Those can be illustrated by Fig. 2, which depicts social network with highlighted edges along paths between Alice and Eve. Concatenation function is applied to partial trusts along every path from Alice to Eve. Those paths are Alice – Cecil – Eve, Alice – Bob – Cecil – Eve and Alice – Bob – Dave – Cecil – Eve. The function is applied separately to every one of them, resulting in three distinct values.

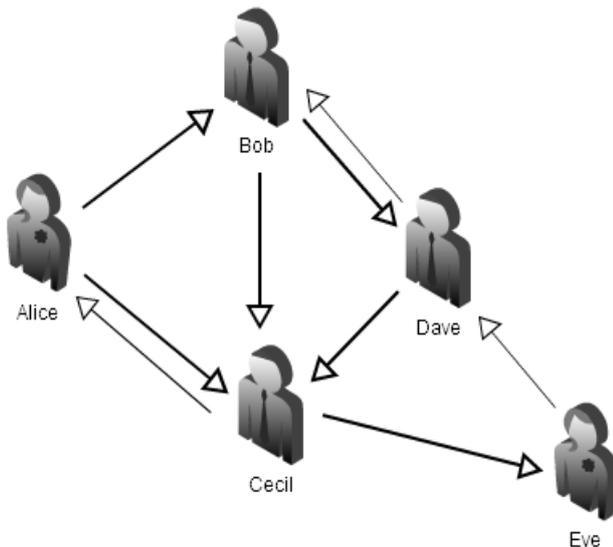


Fig. 2 Various paths between Alice and Eve

Final trust value is then calculated (or selected) by aggregation function. This can be for instance simple maximum or minimum function or more complicated weighted sums. But none of mentioned terminologies is able to present exhaustive description of hierarchy of transitive trust calculation problems. Therefore we suggest it should be subject of deeper research and systematization.

#### 4.2.2. Transitive trust issues

Transitive trust metrics is complex concept and as such it includes several issues, which have to be solved so that selected model gives relevant results. One of such issues is degradation to zero which especially applies to models using multiplication to realize calculation. It is intuitive that trust value of long paths can converge to zero (13). Therefore such models should operate on networks with specific parameters as this drawback does not affect dense networks and networks with non-zero portion of edges with absolute trust.

Another problem transitive trust suffers from is presence of conflicting paths, i.e. when two paths provide contradictory trust information about certain entity. Although it can be source of biased results, only few research works pay attention to it. Lesani and Bagheri in (17) propose a model called FuzzyTrust, which takes fuzzy approach to trust. It uses linguistic terms such as “low”, “medium” or “high” to describe direct trust,

convert it into fuzzy membership functions and then combine them. As conflicting paths between two entities each have peak at different place in the membership function, resulting function has them all. This is then reflected in final linguistic term.

Values of transitive trust can be also affected by some kind of distortion. This can happen either because of malicious behaviour of some entities or by the subjectivity of their attitudes. Not all networks suffer from such problem, but many can and the longer the paths the more uncertainty about results. This limitation has to be taken into considerations every time new transitive trust algorithm is designed.

## 5. CONCLUSION

We have discussed several drawbacks of World Wide Web, which originate in abundant amount of information. Those include poor quality, poor credibility or out-dated information. Description of problems is followed by three concepts, which are seen as solutions for them. However majority of the article is dedicated to discussion about trust metrics, its parameters and accompanying issues.

Our research proved that trust metrics is a complex concept lacking generally recognized systematization. Since there is a need to infer relations between entities presented in some sort of network, we formalize trust metrics as a graph with entities represented by vertices. Relation between two entities can be either direct or indirect. Both forms have their own calculation methods and can be considered separate research subjects. Our following research is going to concentrate on sources where to acquire data for our calculations from, such as scale-free networks (e.g. social networking services).

## REFERENCES

- [1] O'REILLY, T.: What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *O'Reilly Network*, Sep. 2005, <http://oreilly.com/web2/archive/what-is-web-20.html>
- [2] HENDLER, J.: Web 3.0 Emerging, *Computer*, Vol. 42, pp. 111-113.
- [3] KEEN, A.: Web 2.0: The second generation of the Internet has arrived. It's worse than you think, *The Weekly Standard*. [Online] February 15, 2006. [Cited: October 25, 2011] <http://www.weeklystandard.com/Content/Public/Articles/000/000/006/714fjczq.asp>
- [4] BERNERS-LEE, T. – HENDLER, J. – LASSILA, O.: The Semantic Web, *Scientific American Magazine*. [Online] May 17, 2001. [Cited: March 20, 2011] <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
- [5] ZIEGLER, C. N. – LAUSEN, G.: Analyzing Correlation between Trust and User Similarity in Online Communities, *Proc. of Trust Management, Second International Conference, iTrust 2004*, Oxford: Springer, 2004.

- [6] LAZARSFELD, P. F. – MERTON, R. K.: Friendship as a social process: A substantive and methodological analysis, *Freedom and Control in Modern Society*, Van Nostrand, Vol. 18, pp. 18–66, 1954.
- [7] RESNICK, P. – ZECKHAUSER, R.: Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system, *Advances in Applied Microeconomics A Research Annual*, Elsevier BV, Vol. 11, 2002.
- [8] GAMBETTA, D.: Can We Trust Trust ?, *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell, pp. 213–237, 1988.
- [9] LI, L.: Trust Derivation and Transitivity in a Recommendation Trust Model, *Computer Science and Software Engineering, 2008 International Conference on*, pp. 770–773, 2008.
- [10] KOH, N. S. – HU, N. – CLEMONS, E. K.: Do Online Reviews Reflect a Product's True Perceived Quality? - An Investigation of Online Movie Reviews Across Cultures, *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pp. 1 – 10, 2010.
- [11] JØSANG, A. – POPE, S.: Semantic constraints for trust transitivity. *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling - Volume 43*, Newcastle, New South Wales, Australia: Australian Computer Society, Inc., pp. 59–68, 2005.
- [12] ADAMIC, L. A. – ADAR, E.: How to Search a Social Network. *Social Networks*, Vol. 3, pp. 187–203, July 2005.
- [13] RICHTERS, O. – PEIXOTO, T. P.: Trust Transitivity in Social Networks. *PLoS ONE*, Vol. 6, 2011.
- [14] LEVIEN, R. – AIKEN, A.: Attack resistant trust metrics for public key certification, *Proc. of the 7th USENIX Security Symposium*, pp. 265–298, 1998.
- [15] RICHARDSON, M. – AGRAWAL, R. – DOMINGOS, P.: Trust Management for the Semantic Web, *Proc. of the Second International Semantic Web Conference*, pp. 351–368, 2003.
- [16] BINTZIOS, V. G. – PAPAIOANNOU, T. G. – STAMOULIS, G. D.: An effective approach for accurate estimation of trust of distant information sources in the semantic Web, *Security, Privacy and Trust in Pervasive and Ubiquitous Computing. SecPerU 2006. Second International Workshop on. 2006*.
- [17] LESANI, M. – BAGHERI, S.: Fuzzy Trust Inference in Trust Graphs and its Application in Semantic Web Social Networks, *World Automation Congress, 2006. WAC '06*, pp. 1–6, 2006.

Received October 12, 2011, accepted December 20, 2011

## BIOGRAPHIES

**Marek Šajna** was born on 11.1.1987. In 2010 he graduated (MSc) with honours at the Faculty of Management Science and Informatics at University of Žilina. He is currently a doctoral student at Department of Informatics at the Faculty of Management Science, University of Žilina. His scientific research is focusing on solutions for problems related with information overload in World Wide Web, especially semantic metadata and trust metrics.

**Karol Matiaško** was born in 1952. In 1975 he graduated in the study field of Cybernetics at the Faculty of Electrical and Mechanical engineering of the University of Transport and Communications in Žilina. He received PhD in Technical Cybernetic in 1988 at University of Transport and Communication in Žilina like employee of the Research Institute of Transport. Since 1990 he was joined with the Faculty of Management Science and Informatics of University of Žilina and his research and educational activities have been oriented in the area of Database system, Distributed processing and Data processing.