

Trust Networks on the Semantic Web

Jennifer Golbeck¹, Bijan Parsia¹, James Hendler¹

¹University of Maryland, College Park
A. V. Williams Building
College Park, Maryland 20742
golbeck@cs.umd.edu, hendler@cs.umd.edu, bparsia@isr.umd.edu

Abstract. The so-called "Web of Trust" is one of the ultimate goals of the Semantic Web. Research on the topic of trust in this domain has focused largely on digital signatures, certificates, and authentication. At the same time, there is a wealth of research into trust and social networks in the physical world. In this paper, we describe an approach for integrating the two to build a web of trust in a more social respect. This paper describes the applicability of social network analysis to the semantic web, particularly discussing the multi-dimensional networks that evolve from ontological trust specifications. As a demonstration of algorithms used to infer trust relationships, we present several tools that allow users to take advantage of trust metrics that use the network.

1 Introduction

"Trust" is a word that has come to have several very specific definitions on the Semantic Web. Much research has focused on authentication of resources, including work on digital signatures and public keys. Confidence in the source or author of a document is important, but trust, in this sense, ignores many important points.

Just because a person can confirm the source of documents does not have any explicit implication about trusting the content of those documents. This project addresses "trust" as credibility or reliability in a much more human sense. It opens up the door for questions like "how much credence should I give to what this person says about a given topic," and "based on what my friends say, how much should I trust this new person?"

In this paper, we will discuss the application of a social network to the semantic web. Section 2 discusses how to build a meaningful social network from the architecture of the semantic web, and how it conveys meaning about the structure of the world. Section 3 will describe the implementation of such a network. We describe a sample ontology, an algorithm for computing trust in a network, and present tools that use this network to provide users with information about the reputation of others.

1.2 Related Work

This paper uses techniques developed in the field of social network analysis, and applies that to the issue of trust on the semantic web. This section describes the most relevant works from each area.

1.2.1 Social Networks

Social networks have a long history of study in a wide range of disciplines. The more mathematical of the studies have appeared in the "Small World" literature. The work on Small Worlds, also commonly known as "Six Degrees of Separation" originated out of Stanley Milgram's work in the 1960s. His original studies indicated that any two people in the world were separated by only a small number acquaintances (theorized to be six in the original work) [18]. Since then, studies have shown that many complex networks share the common features of the small-world phenomenon: small average distance between nodes, and a high connectance, or clustering coefficient [21].

Small world networks have been studied in relation to random graphs [29]. For social systems, both models have been used to describe phenomena such as scientific collaboration networks [20], and models of game theory [29]. The propagation of effects through these types of networks has been studied, particularly with respect to the spread of disease [19, 7]. The web itself has shown the patterns of a small world network, in clustering and diameter [4, 1].

Viewing the current web as a graph, where each page represents a node, and the hyperlinks translate to directed edges between nodes, has produced some interesting results. The main focus of this research has been to improve the quality of search [7,6,5,8,14,26]. Other work has used this structure for classification [9] and community discovery [15].

1.2.2 Trust on the Semantic Web

Yolanda Gil and Varun Ratnakar addressed the issue of trusting content and information sources [12]. They describe an approach to derive assessments about information sources based on individual feedback about the sources. As users add annotations, they can include measures of Credibility and Reliability about a statement, which are later averaged and presented to the viewer. Using the TRELIS system, users can view information, annotations (including averages of credibility, reliability, and other ratings), and then make an analysis.

Calculating trust automatically for an individual in a network on the web is partially addressed in Raph Levin's Advogato project [17]. His trust metric uses group assertions for determining membership within a group. The Advogato [3] website, for example, certifies users at three levels. Access to post and edit website information is controlled by these certifications. On any network, the Advogato trust metric is extremely attack resistant. By identifying individual nodes as "bad" and finding any nodes that certify the "bad" nodes, the metric cuts out an unreliable portion of the network. Calculations are based primarily on the good nodes, so the network as a whole remains secure.

2 Networks on the Semantic Web

Studying the structure of the hypertext web can be used to find community structure in a limited way. A set of pages clustered by hyperlinks may indicate a common topic among the pages, but it does not show more than a generic relationship among the pages. Furthermore, pages with fewer outgoing links are less likely to show up in a cluster at all because their connectance is obviously lower. These two facts make it difficult for a person to actually see any relationship among specific concepts on the web as it currently stands – classification is not specific enough, and it relies on heavy hyperlinking that may not be present.

The Semantic Web changes this. Since the semantic data is machine-understandable, there is no need to use heuristics to relate pages. Concepts in semantically marked up pages are automatically linked, relating both pages and concepts across a distributed web.

By its nature, the semantic web is one large graph. Resources (and literals) are connected by predicates. Throughout the rest of this paper, we will refer to resources as objects, and predicates as properties. Mapping objects to nodes and properties to labeled edges in a graph yields the power to use the algorithms and methods of analysis that have been developed for other manifestations of graphs.

While the graph of the entire Semantic Web itself is interesting, a subgraph generated by restricting the properties, and thus edges, to a subset of interest allows us to see the relationships among distributed data. Applications in this space are vast. Semantic markup means that retrieving instances of specific classes and the set of properties required for a particular project, becomes easy. Furthermore, merging data collected from many different places on the web is trivial.

Generating social networks on the semantic web is a similar task with useful results. Information about individuals in a network is maintained in distributed sources. Individuals can manage data about themselves and their friends. Security measures, like digital signatures of files, go some way toward preventing false information from propagating through the network. This security measure builds trust about the authenticity of data contained within the network, but does not describe trust between people in the network.

There are many measures of "trust" within a social network. It is common in a network that trust is based simply on knowing someone. By treating a "Person" as a node, and the "knows" relationship as an edge, an undirected graph emerges (Figure 1 shows the graph of acquaintances used in this study). If A does not know B, but some of A's friends know B, A is "close" to knowing B in some sense. Many existing networks take this measure of closeness into account. We may, for example, reasonably trust a person with a small Erdos number to have a stronger knowledge of graph theory than someone with a large or infinite number.