

Final Year Project Interim Report

Reputation Algorithms for the Social Web

John Patrick Bracken

An thesis submitted in part fulfilment of the degree of

BA/BSc (hons) in Computer Science

Supervisor: Dr. Housseem Jerbi

Moderator: Dr. Michael O'Mahony



UCD School of Computer Science and Informatics
College of Engineering Mathematical and Physical Sciences

University College Dublin

February 6, 2014

Table of Contents

Specification	2
Abstract	3
1 Introduction (aim for 2 pages)	4
2 Background Research (aim for 6 pages)	5
2.1 Trust	5
2.2 Reputation	5
2.3 Evolution of the Web	5
2.4 Case studies	6
2.5 Reputation Algorithms	7
2.6 Development Tools	7

Specification

Subject: Social Network Analysis, Reputation Systems

Prerequisites: None

Corequisites: None

Coverage Social Network Analysis, Reputation Systems

Project Type: Design and Implementation

Software Requirements: Java, Linux, MySQL (or other)

Hardware Requirements: Laptop for development. Access to server will be provided if necessary.

Preassigned: No

Abstract

In this project I intend to compare the performance of generic reputation algorithms using the Stack Exchange Question and Answer sites' open-sourced data dumps. An attempt will also be made to improve upon the performance of the algorithms, and analyze their robustness against attack.

Chapter 1: Introduction (aim for 2 pages)

With the increasing use of the internet in our day-to-day tasks, it has become more and more important that we be able to verify the trustworthiness of the strangers we so often rely on. While the use of technologies such as public key encryption allow us to verify *who* we are talking to with reasonable confidence, we are still left with the problem of determining that person's trustworthiness as an individual—whether that be trust in their knowledge in a particular field, or that they can be relied upon to deliver a good or service to satisfaction.

To that end there has been considerable recent research into the fields of peer-to-peer trust and user reputation systems in social networks (McNally, O'Mahony and Smyth 2013; Cheng and Vassileva 2005; Mui 2002).

There are numerous contexts in which a measure of trust is desired online, from internet transactions to online tutorials. For the purposes of this project we will be focusing on the exchange of knowledge and expertise between users on the Stack Exchange Question and Answer network. We define trust as the likelihood that the user answering a question is knowledgeable on the subject matter at hand. User reputation will refer to a numerical score allocated to a user indicating their trustworthiness.

In this project, I will be implementing a number of generic approaches to reputation (Weighted Sum; Hubs and Authorities, or HITS; and PageRank) upon a graph made up of *collaboration events* [TODO: explain collaboration events] and comparing their performance on the Stack Exchange datasets to each other and to the Stack Exchange reputation model (McNally, O'Mahony and Smyth 2013).

Chapter 2: **Background Research (aim for 6 pages)**

2.1 Trust

Trust can be defined as a relationship between two relationship. A trustor is an entity who places a certain amount of faith in the actions of knowledge of another entity, (or the trustee). Trust is not a symmetric relationship, so while, for example, a student may trust a teacher's knowledge in their subject matter, the teacher will likely have less reason to trust in the student's knowledge.

Trust is an abstract relation that can occur between numerous types of entities, such as between people or web-pages (Page, L., Brin S., Motwani, R., and Winograd, T. 1999).

Trust is an inherently qualitative property, and this is what causes us to draw a distinction between trust and reputation. [TODO: expand on this and add citations]

2.2 Reputation

Reputation is a numeric score that attempts to measure a particular entity's trustworthiness. There are many proprietary and generic reputation implementations in place across the web used to measure trust, such as google's PageRank for measuring a web-page's importance, to ebay's bespoke reputation system that allows buyers to find reliable sellers. [TODO: expand on this and add citation]

2.3 Evolution of the Web

[N.B. this entire paragraph written from memory, I need to fact check and find sources, although I don't believe there are any inaccuracies]

When the web first exploded into widespread use in the nineties, it was a very static compilation of pages connected by hyperlinks. Typically, websites were only published by universities, government organisations and corporations, with content being controlled by their respective webmasters. It was much easier, then, to evaluate the trustworthiness of online resources; content related to hardware published by IBM was likely to be accurate, but IBM's advice on cake-baking may be taken with a pinch of salt (or perhaps not, as the casemay be).

Over time however, as computers became more sophisticated; so-called web 2.0 technologies such as PHP and javascript evolved; and the number of internet users exploded, [TODO: get

actual figures] the roles of producer and consumer are no longer so rigidly defined as they once were. Instead, anyone can post on a social network, publish a music review to potentially millions of people, or share off-colour remarks on a video of a cat falling from a ledge.

For the large part this is not very often an issue, as many of these forms of communication are entirely, and transparently, optional, but when a user, Alice, asks a dog lovers' web-forum how much chocolate is safe to feed to her chihuahua, she may get a range of different answers. Bob gives her the correct answer and Mallory, out of some sense of bizarre schadenfreude, intentionally gives her a malicious answer framed as genuine advice. Alice not know who to trust, but errs on the side of caution.

While the above example is contrived, it illustrates the potential dangers of the social web, and how important it is to attach an accurate reputation to users.

[TODO: add citations to the above]

2.4 Case studies

2.4.1 Stack Exchange

The Stack Exchange network is a network of Question and Answer (Q&A) communities, each focused on a specific field of expertise, with one for everything from programming to bicycling and cooking. As of the time of writing this report, the network consists of 114 Q&A sites, almost 4.5 million users and over 8 million questions with 14.6 million answers. (Stack Exchange About Page)

In the Stack Exchange system, reputation is calculated as a sum of points earned by contributing to the sites in various ways. Nearly every site activity, from asking and answering questions, to suggesting edits and flagging content for moderation, can earn the user points. This system is aimed more towards encouraging activity than accurately evaluating a user's trust. (SO: What is reputation?)

All content on the Stack Exchange network is published under a Creative Commons Attribution Share Alike license, and every three months an anonymized data dump of all the questions and answers on every site is publicly released. Each site is divided into a collection of XML files for the different objects. The ones we are primarily interested in are the Posts.xml and Users.xml files. The other files deal with edit history, comments, and in-site rewards that are unnecessary for this project. [TODO: add citation]

The files are all quite straight-forward, with each row mapping to a database row.

The relevant files I will need to use to evaluate these algorithms are the Users.xml file and the Posts.xml file.

Users.xml

The Users table has columns for user ID, display name, number of up-and-down-votes creation date etc. We are principally concerned with the user ID field so that we can match posts and votes to the correct user.

Posts.xml

The posts table has columns for all post information such as title, contents, owner ID and the post type ID. We are mainly concerned with the owner ID, the post type ID—so that we can separate answers from questions and other post types— and the score the post received.

[TODO: Case studies]

2.4.2 Klout

2.4.3 HayStaks

2.5 Reputation Algorithms

2.5.1 Collaboration Events

In the context of user interactions, we can define a so-called *collaboration event* between two entities. When entity A interacts with the collection of entities C, A then produces a reputation score for each entity in the collection C.

As we look at these events as a whole for a social network, they form a weighted directed graph of collaboration events, with the users represented by nodes, and reputation scores by the edges between nodes. There can potentially be numerous edges between the same two nodes in both directions, in which case we combine these scores so that there are at most two edges between two nodes (one for each direction).

[TODO: add brief explanations of the algorithms]

2.6 Development Tools

The project is written in Java and uses PostgreSQL to store persistent data. The Java Flyway project is used for database migrations and to ensure database consistency across development environments. Database operations are handled with the JDBC and JDBI libraries and Apache Maven is used for builds and dependency management, while the Gephi toolkit is used to build the graph and evaluate reputation scores.

Bibliography

McNally, K., O'Mahony, M.P., and Smyth, B. 2013 *A Model of Collaboration-based Reputation for the Social Web*

Cheng, R., and Vassileva, J. 2005. *Reward Mechanism for Sustainable Online Learning Community.* In *Proceedings of the 2005 conference on Artificial Intelligence in Education.* IOS Press.

Page, L., Brin S., Motwani, R., and Winograd, T. 1999 *The PageRank citation ranking: Bringing order to the Web.*

Mui, L. 2002. *A Computational Model of Trust and Reputation.* *Agents, Evolutionary Games, and Social Networks*

Resnick, P., and Zeckhauser, R. 2002. *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System.* *Advances in Applied Microeconomics* 11:127?157

Stack Exchange about page <http://stackexchange.com/about>

SO: What is reputation? <http://stackoverflow.com/help/whats-reputation>