

Mining the Social Web

by Morten Goodwin Olsen

Note that this document has been revised from the original version. The original version included an evaluation of the attended course which does not have any public interest.

In this short paper we present the correlation between the Ph.D. course *S2_4 Mining the Social Web Mining for web2.0* at Aalborg University and eGovernment Monitoring of online public services. The intention is to show how the course is applicable to the Ph.D. of Morten Goodwin Olsen.

Morten Goodwin Olsen's Ph.D. is regarding automatic measurements of eGovernment services including accessibility, transparency, efficiency and impact. Clearly, for automatic web monitoring, web mining techniques are essential, including techniques for mining the social web.

Navigability and small world

An example of an approach with correlation is measuring navigability within a web site. According to WCAG¹ and UWEM², navigation within a web site should be clear and consistent. Based on this, we can assume that there are at least two requirements for a web site to be accessible with regards to navigability:

- (1) Any web page (or any part of the web site graph) should be easily available from any other page – with only a few clicks.**

An extreme case of the opposite would be to have a web site where each web page contains only one link. This is typically done in online quiz pages where each page has a question and includes only a link to the next question. To arrive at for example page 15, a user needs to click through 14 before he/she arrives at the page he/she wants.

- (2) There should not be too many links on each web page as this would make it hard for users to find the correct link.**

As an extreme case of the opposite we can assume that for each web page there are links to all other web pages. Even though there is only one click between each page, this example is clearly inaccessible as each page would consist of a huge number of links making it hard for any user to navigate.

It is clear that the above two requirements are very similar to graph analysis such as the analysing small world phenomena³. Furthermore, based on the above, an algorithm for measuring navigability within a web site could clearly benefit from the research on social

1 WCAG – Web Content Accessibility Guidelines <http://www.w3.org/TR/WCAG10/>

2 UWEM – Unified Web Evaluation Methodology http://www.wabcluster.org/uwem1_2/

3 Note that this is not a one-to-one mapping as a site map within a web site is regarded as accessible, even though this contradicts requirement (2).

In graph theory this would be similar to one important node. However, in a small world networks do not consist of one special node more important than all others.

graphs.

Efficiency and findings shortest paths

Another example where social network analysis would be beneficial for eGovernment measuring is efficiency within a municipality. Several municipalities publish selected e-mail logs online, e.g. regarding building permission. The intention of this is for inhabitants to be able to check online whether e.g. a building application is accepted or not. A general assumption is:

- (1) The shorter path an e-mail needs to take within the administration of the municipalities, the more efficient the municipality is.**

Again, it is clear that there is much correlation between this approach and social web mining since there exists a lot of research on measuring path lengths in graphs. Note that for this approach to be applicable, assumption (1) needs to be verified.

Impact and predicting trust

Impact is another important measure of eGovernment services. In this setting, impact can be defined as how much an online eGovernment application is used by its citizens. Measuring this may have similarities with social network analysis.

From this we can assume the following:

- (1) An online service that is not trusted will be less used, and thus have less impact, than a service that is trusted.**

Furthermore, predicting trust has been done by several papers in social network analysis. A challenge with such an approach is gathering of data on trust, as it is not easy to extract such data using a web crawler, web site logs or similar.