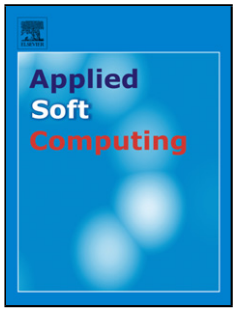

Title: Improving Data Exploration in Graphs with Fuzzy Logic and Large-Scale Visualisation

Author: id="aut0005" orcid="0000-0001-5688-2039" author-id="S1568494616306731-96ed5a1c6030e2d9cdc61fd5c4f80b11"> Miguel Molina-Solana id="aut0010" orcid="0000-0001-5347-8700" author-id="S1568494616306731-087f2faa053f8835959688327fb2db75"> David Birch id="aut0015" author-id="S1568494616306731-7bca59a5edda9f4bc159cc3eb5fbc5f0"> Yi-ke Guo

PII: S1568-4946(16)30673-1
DOI: http://dx.doi.org/doi:10.1016/j.asoc.2016.12.044
Reference: ASOC 3983

To appear in: *Applied Soft Computing*

Received date: 4-5-2016
Revised date: 19-12-2016
Accepted date: 20-12-2016

Please cite this article as: Miguel Molina-Solana, David Birch, Yi-ke Guo, Improving Data Exploration in Graphs with Fuzzy Logic and Large-Scale Visualisation, *<![CDATA[Applied Soft Computing Journal]]>* (2017), http://dx.doi.org/10.1016/j.asoc.2016.12.044

Improving Data Exploration in Graphs with Fuzzy Logic and Large-Scale Visualisation

Miguel Molina-Solana, David Birch, Yi-ke Guo

Data Science Institute, Imperial College London, UK

Abstract

This work presents three case-studies of how fuzzy logic can be combined with large-scale immersive visualisation to enhance the process of graph sensemaking, enabling interactive fuzzy filtering of large global views of graphs. The aim is to provide users a mechanism to quickly identify *interesting nodes* for further analysis. Fuzzy logic allows a flexible framework to ask human-like curiosity-driven questions over the data, and visualisation allows its communication and understanding. Together, these two technologies successfully empower novices and experts to a faster and deeper understanding of the underlying patterns in big datasets compared to traditional means in a desktop screen with crisp queries. Among other examples, we provide evidence of how these two technologies successfully enable the identification of relevant transaction patterns in the Bitcoin network.

Keywords: Graph sensemaking, Fuzzy Logic, Data Exploration, Large-scale Visualisation, Graph Visualisation

1. Introduction

With the advances of computational and sensing systems, huge amounts of data are created and stored every day. In order to gain useful knowledge from them, these data needs to be processed and interpreted in proper and meaningful ways. These tasks are still primarily performed by humans in possession of domain knowledge and the mathematical and statistical tools at their disposal.

In this context, Data Visualisation (together with descriptive analytics) has been proved an effective way of gaining insights into data [1], and is consistently the first and last step in most data science projects [2]. According to Leigh and colleagues [3], visualisation serves three important roles during the scientific discovery process: quick verification of correctness of simulation models during development, quick delivery of simulation results by close integration with the model, and easiness of understanding by lay audiences.

Fuzzy logic and fuzzy set theory [4] rely on the fact that human beings employ mostly words in their computing and reasoning processes. These are appropriate formalisms to handle imprecise knowledge, and particularly, imprecise queries over the data [5]. That kind of queries is even more relevant when performing exploratory data analysis over big datasets, and has its maximum exponent in the query "which are the most interesting pieces of data?".

As it is well known, within the Fuzzy formalism, axioms and facts are not in general either true or false, but they may hold to some degree of truth. Fuzzy Logic offers then a sound theoretical framework for curiosity-driven questions in the initial stages of the Data Exploration process.

In summary, both Data Visualisation and Fuzzy Logic improve interpretability and understanding of the data at hand, facilitating its query and exploration. In Chen's words [6], "a visualisation process is a search process". A search of meaningful insights than can drive further and deeper research and analysis on the dataset. Our thesis in this work is that blending these two technologies is a sensible proposal to intuitively perform an informed data exploration on big datasets, that would otherwise be very difficult, if not impossible, to process and fully understand. This combination has also the advantage to enable data understanding by broader lay audiences.

Describing datasets as graphs and networks, and relying on existing techniques to analyse them, has proven a successful line of action for plenty of scenarios. However, making sense of large graphs remains a fundamental challenge (reaching performance and usability issues), with few tools that allow users to interactively explore, visualise, and understand large graphs [7].

This paper precisely describes successful examples of such exploration employing fuzzy queries (particularly, looking for *interesting nodes*) and leveraging the *KPMG Data Observatory* at Imperial College's Data Science Institute, the largest visualisation studio of its kind in Europe with 132 megapixels of display surface. The employment of a large-scale, high-resolution visualisation environment provides more effective data visualisation (with a global view) and greater insight into the data. By identifying interesting nodes, users are able to focus their attention

Email addresses: m.molina-solana@imperial.ac.uk (Miguel Molina-Solana), david.birch@imperial.ac.uk (David Birch), y.guo@imperial.ac.uk (Yi-ke Guo)