A B A K Ó S
PUC Minas Instituto de Ciências Exatas e Informática

DocGenealogy – Visualizing the doctoral advisors and mentors genealogic tree

Model magazine Abakós ICEI the PUC Minas

David Fernandes[1] e Elizabeth Carvalho

Resumo

DocGenealogy é um projeto de mineração de dados que permite a visualização e exploração do relacionamento de assessoria / orientação de doutorados através de um grafo interativo. A Wikipedia é conhecida por sua natureza colaborativa informal, com base em uma forma livre e não estrutural de produzir informações. Esse é um dos seus pontos fortes, na verdade, a abertura a todo tipo de colaborações. A DocGenealogy utiliza os dados disponíveis da Wikipedia sobre assessoria e orientação para doutorados para descobrir e rastrear as relações existentes entre conselheiros ou mentores e seus alunos. O ponto de partida do processo de mineração de dados é sempre um cientista bem conhecido. Este processo continua, até encontrar os conselheiros e mentores relacionados com renome co-relacionados sendo construído iterativamente o grafo de relacionamento. Além disso, o DocGenealogy também mostra visualmente como as pessoas estão distribuídas por *alma mater* e área de conhecimento.

Palavras-chave: Visualização de informação. Visualização de grafos. Recuperação de informação. Mineração de dados visual. Wikipedia.

[1] Mestrando em Tecnologias e Sistemas Informáticos, pela Universidade Aberta.

[1] Professora Auxiliar na Universidade Aberta e investigadora do pólo CIAC/UAb

Abstract

DocGenealogy is a data mining project that allows the visualization and exploration of the doctoral advisement/mentoring relationship through an interactive graph chart. Wikipedia is known for its informal collaborative nature, based on a somewhat loose and non-structural way of producing information, and one of its strengths is, in fact, the openness to all sorts of collaborations. DocGenealogy uses the Wikipedia's available data on doctoral advisement and mentoring to find out and track the existing relationships between advisors or mentors and their students. The data mining process starting point is always some well-known scientist. This process continues, till finding all the ultimate reputed co-related advisors and mentors, and builds iteratively the relationship graph. Besides this, DocGenealogy also shows visually how people are distributed by alma mater and field of knowledge.

Keywords: Information visualization. Graph visualization. Information retrieval. Visual data mining. Wikipedia.

1 INTRODUCTION

Information visualization (InfoVis), the study of transforming data, information, and knowledge into interactive visual representations, is very important to users because it provides mental models of information. The boom in big data analytics has triggered broad use of InfoVis in a variety of domains, ranging from finance to sports to politics (Liu et al., 2014). There are different kinds and types of visualization techniques that can be used to map visually the data.

The study of complex activities such as scientific production and software development often requires modelling connections among heterogeneous entities including people, institutions, and artefacts (Heer & Perer, 2014). Graphs have been widely used to visualize complex relationships such as the trade relationships among cities and the friend relationships in social networks. A graph consists of nodes and edges where the nodes represent entities and the edges represent the relationships among the entities/nodes. For example, graphs can show vivid trade routes by representing cities as nodes and trade relationships as edges. Large data sizes often bring out the visual clutter problem in graphs (Zhou et al., 2013). A node-link diagram visualization technique is frequently used to map graphs.

DocGenealogy ([http://address/]) is a web-based application that allows easily finding out the existing relationship between doctoral advisers and mentors along the time. It offers an interactive dynamic force-direct node-link diagram to assist end-users on this task besides other charts to support a better information insight.

The rest of the article is structured as follow: we will give first the scientific background to this work, with special focus on visual techniques used to map visually graphs, and then we will describe the DocGenealogy conceptual model and its prototype, including its data model, source, design and technology. Finally, we will discuss the main conclusions and outputs, besides future work to be developed.

2 BACKGROUND

Information visualization is concerned with (Moere et al., 2012) exploiting the cognitive capabilities of human visual perception in order to convey meaningful patterns and trends hidden in abstract datasets. As data has steadily become more complex in terms of its size, dimensionality and time-variance, the field has been challenged to create

new techniques that are more sophisticated, and to develop objective evaluation methods that are able to benchmark these different techniques against each other.

Visualization is suitable when there is a need (Munzner, 2014) to augment human capabilities rather than replace people with computational decision-making methods. The design space of possible Vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations. Vis design is full of trade-offs, and most possibilities in the design space are ineffective for a particular task, so validating the effectiveness of a design is both necessary and difficult.

Graphs in general form one of the most important data models in computer science because many problems and domains can be modelled as graph structures. Just to name a few, there are automata in theoretical computer science, flow networks such as pipes and roads, digital and non-digital social networks, computer networks such as the Internet, networks of companies and financial transactions, chemical reaction chains and molecular interactions, epidemic spreads of diseases in communities, or correlations of controlled variables in experiments (Beck, Burch, Diehl & Weiskopf, 2014).

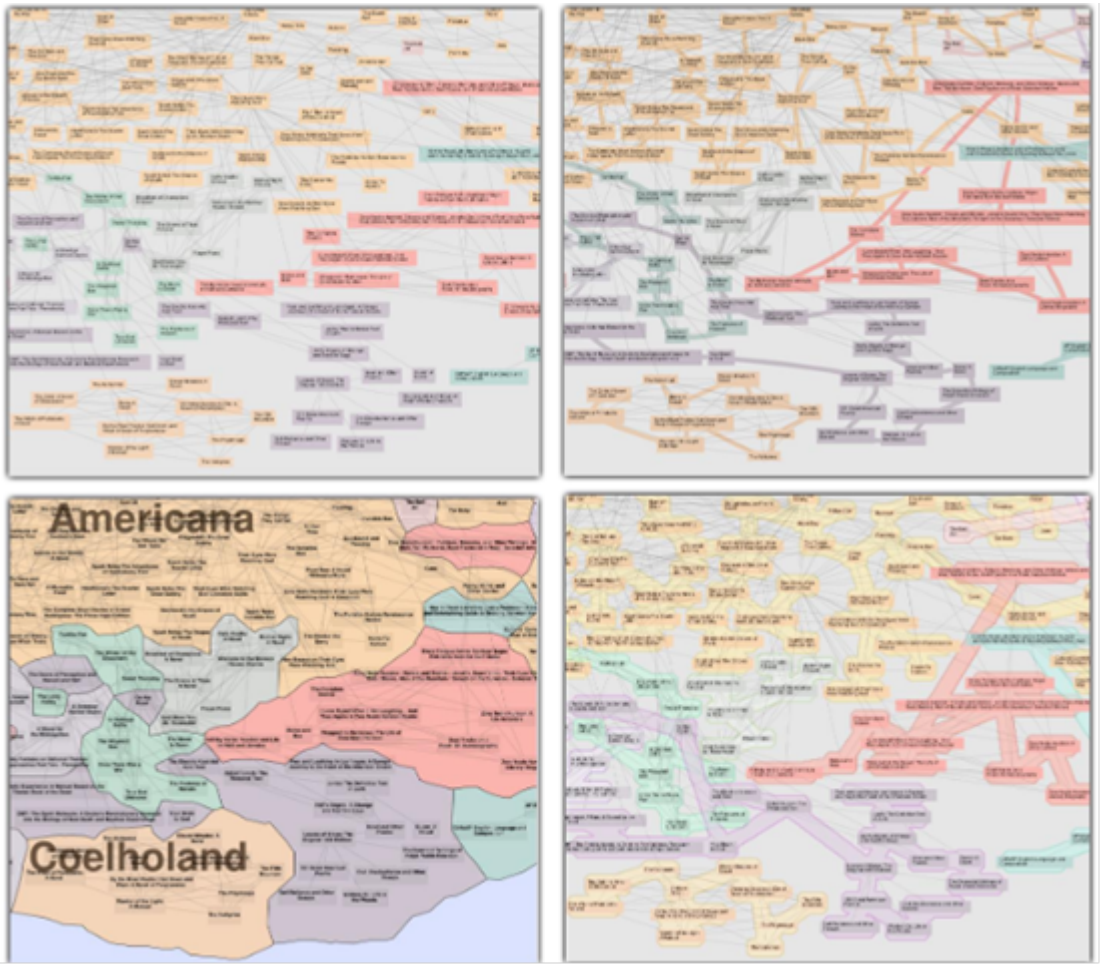
Researchers have created a diversity of visualization techniques for networks and graphs. Two common representations used in social network analysis are node-link (Jianu, Rusu, Hu, & Taggart, 2014) diagrams and adjacency matrix views (Kang, Lee, Koutra & Faloutsos, 2014). Hybrids of the two have also been proposed (Rufiange, McGuffin & Fuhrman, 2012; Henry, Fekete & McGuffin, 2007). These approaches organize elements according to the linkage structure of the graph creating a combination between these visual techniques and even others.

An alternative approach is to plot network data according to the attributes of the nodes, as in a scatter plot or so-called semantic substrates. Shneiderman and Aris (2006) proposed a network visualization layout based on a user-defined semantic substrate with node-links diagram as an underlying visualization. Semantic substrates are spatially non-overlapping regions that are built to hold nodes based on some category present in the dataset. This approach is well suited for assessing potential correlations between node attributes and network structure.

Others have researched means of dealing with large graphs in excess of tens of thousands nodes. Common strategies include filtering and aggregation. Dunne and Schneiderman (2013) to help address this problem introduced a technique called motif simplification, in which common patterns of nodes and links are replaced with compact and meaningful glyphs. Zinsmaier, Brandes, Deussen, & Strobelt (2012) proposed a technique that allows straight-line graph drawings to be rendered interactively with adjustable level of detail. The approach consists of a novel combination of edge cumulation with density-based node aggregation and is designed to exploit common graphics hardware for speed. Figure 1

illustrates some node-link diagrams.

Figure 1 - Four visualizations for viewing group information over node-link diagrams



Fonte: Jianu, Rusu, Hu, & Taggart, 2014.

Data Mining (DM) aims to extract useful knowledge from raw data (Cortez & Embrechts, 2013). Interest in this field arose due to the advances of Information Technology and rapid growth of business and scientific databases. These data hold valuable information such as trends and patterns, which can be used to improve decision making. Two important DM tasks are classification and regression. Both tasks use a supervised learning paradigm, where the intention is to build a data-driven model that learns an unknown underlying function that maps several input variables to one output target.

Liao, Chu and Hsiao (2012) reviewed the data mining techniques developed recently and several major kinds of data mining methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining. The techniques for mining knowledge from different kinds of databases, including relational, transactional, object oriented, spatial and active databases, as well as global information systems were also examined by them besides potential data mining

applications and some research issues.

Recently, applications of evolutionary algorithms (Mukhopadhyay, Maulik, Bandyopadhyay & Coello, 2014) have been found to be particularly useful for automatic processing of large quantities of raw noisy data for optimal parameter setting and to discover significant and meaningful information. Many real life data mining problems involve multiple conflicting measures of performance, or objectives, which need to be optimized simultaneously. Under this context, multi-objective evolutionary algorithms are gradually finding more and more applications in the domain of data mining since the beginning of the last decade.

3 DOCGENEALOGY PROTOTYPE

Many times it is interesting to find out the relationship between people. To achieve a master or doctoral degree, a student must have an advisor, who guides him through the research process and stands for him when the defense of his thesis takes place. However, it is very easy to lose the track of this relationship. Many of the students that were advised do not pursuit an academic career or do not develop relevant research, becoming notorious somehow. Others, in opposite, do.

The Wikipedia is a free encyclopedia, written collaboratively by the people who use it. It is a rich repository of information, not exhaustive, but holds a huge quantity of useful data. It contains, for instance, many articles and details about scientists, including the name of their notorious advisors or students. However, it is not easy to interpret the relationship between them.

Node-link diagram is one of the most suitable information visualization techniques to represent the relationship between data. Scientific mentors and advisors relationships can be easily interpreted using this kind of visual approach. A node-link diagram can be visually explored in many ways, besides interacted using different methods.

The next sections describe what is behind the DocGenealogy prototype. We will refer to its data model, source and mining process. We will also describe its design, interface and technology.

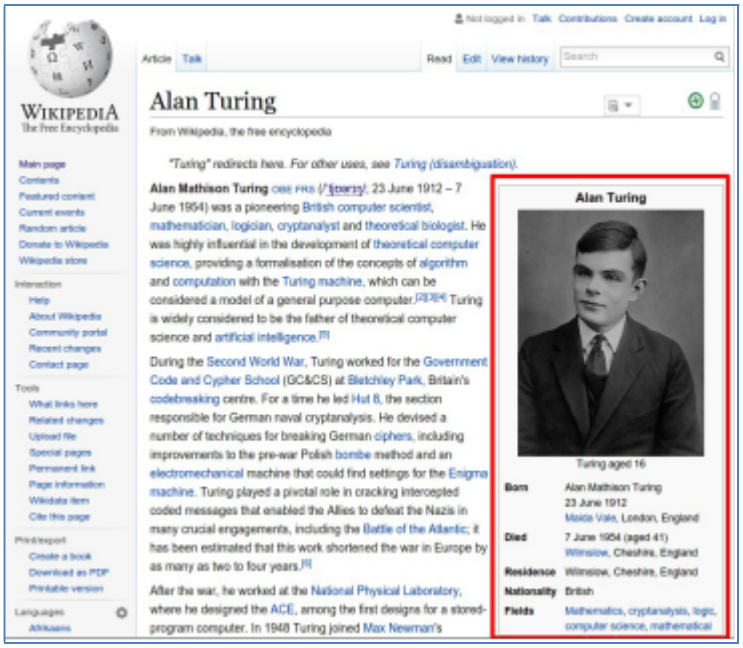
3.1 Data

An infobox of a (Tran & Cao, 2013) Wikipedia article generally contains key facts in the article and is organized as attribute-value pairs. Infoboxes not only allow readers to

rapidly gather the most important information about some aspects of the articles in which they appear, but also provide a source for many knowledge bases derived from Wikipedia. However, not all the values of the attributes of an infobox are updated frequently and accurately. Although not mandatory, the use of infoboxes is quite common and about 33% of Wikipedia articles contained an infobox (figure 2 illustrates an infobox inside a Wikipedia page).

For this work we are interested in the Infobox scientist which has the following definition template described in figure 3. Highlighted in red, we show the attributes that were necessary to extract from each page, namely: the name, the birth and death dates, the nationality and citizenship, the alma mater, the advisors of the scientist and his notorious students.

Figure 2 - Infobox (inside the red rectangle) within a Wikipedia page example.



Fonte: Wikipédia.

Figure 3 - Wikipédia infobox scientist attributes.

```
{{Infobox scientist
| honorific_prefix =
| name             =
| honorific_suffix =
| native_name      =
| native_name_lang =
| image            =        <!--(filename only, i.e. without "File:" prefix)-->
| image_size       =
| alt              =
| caption          =
| birth_date       =        <!--{{birth date |YYYY|MM|DD}}-->
| birth_place      =
| death_date       =        <!--{{death date and age |YYYY|MM|DD |YYYY|MM|DD}} (death date then birth date)-->
| death_place      =
| death_cause      =
| resting_place    =
| resting_place_coordinates =  <!--{{coord|LAT|LONG|type:landmark|display=inline,title}}-->
| other_names      =
| residence        =
| citizenship      =
| nationality      =
| fields           =
| workplaces       =
| patrons          =
| education        =
| alma_mater       =
| thesis_title     =        <!--(or  | thesis1_title =  and  | thesis2_title = )-->
| thesis_url       =        <!--(or  | thesis1_url  =  and  | thesis2_url  = )-->
| thesis_year      =        <!--(or  | thesis1_year =  and  | thesis2_year = )-->
| doctoral_advisor =        <!--(or  | doctoral_advisors = )-->
| academic_advisors =
| doctoral_students =
| notable_students =
| known_for        =
| influences       =
| influenced       =
| awards           =
| author_abbrev_bot =
| author_abbrev_zoo =
| spouse           =        <!--(or | spouses = )-->
| partner          =        <!--(or | partners = )-->
| children         =
| signature        =        <!--(filename only)-->
| signature_alt    =
| website          =        <!--{{URL|www.example.com}}-->
| footnotes        =
}}
```

Fonte: Wikipédia.

Table 1 describes the data that is used by DocGenealogy. Most of it is mined. We have a total of seven dimensions and three measures (the totals and counting). The totals of advisors or mentors per alma mater and knowledge fields are calculated. The counting of direct relationships from and to an advisor or mentor is also computed.

Table 1 - Data dimension/measures and classification.

Dimensions/measures	Classification
Name	Nominal
Birth date	Numeric and temporal
Death date	Numeric and temporal
Citizenship	Nominal and geo-referenced
Nationality	Nominal and geo-referenced
Alma mater	Nominal
Doctoral advisor	Nominal
Totals per alma mater	Numeric, and quantitative
Totals per knowledge field	Numeric, and quantitative
Count of relationships	Numeric, and quantitative

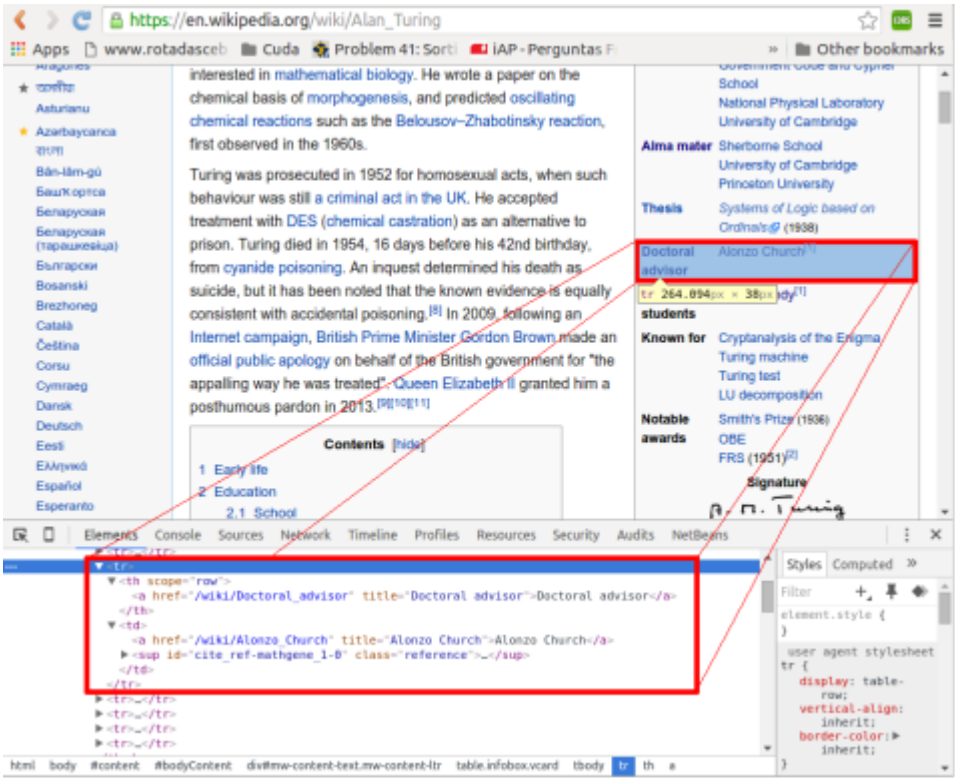
The structure showed in figure 3 is used in the Wikipedia Editor as a template for article editing. The result, and the information that we have access to, as in any other type of dynamic web page, is pure HTML, which obeys to another type of structure. The data-mining tool that we implemented, which may be considered a simplified web crawler, reads the HTML of a page, extracts the relevant information needed and, making use of the URL's of

the doctoral advisors and students, recursively obtain more information about the graph.

The recursive mining uses a depth-first search strategy (Liu & Motoda, 2012), starting on a given scientist's Wikipedia page (Alan Turing was chosen for this work) and crawling the Wikipedia pages of its doctoral students first, downstream, and doctoral advisors afterwards, upstream. In order to respect the Wikipedia best practices for automated content crawling (Wikipedia, 2015), we limited the request frequency to 1 page request per second.

An example for Alan Turing's infobox HTML regarding his Advisors and Students (all the other attributes were extracted in a similar way) is showed in figure 4. In order to put some limit on the time and space needed to gather the data, we established two limiting values: one on depth, for the maximum number of levels to crawl, up and downstream, and total elements for the maximum number of scientists to crawl.

Figure 4 - HTML view of a scientist infobox.



Fonte: Wikipédia.

3.2 Visualization

Any interactive multimedia project must follow certain steps in order to facilitate their development and implementation. It is also important to consider the contents and the interface to be chosen to provide an interactive structure that meets the original objectives of data exploration.

Because we wanted to represent the existing relationship between doctoral advisers and mentors along the time, we assumed that a dynamic node-link diagram visual representation should be the more suitable. We also wanted to give some flexibility to navigate and explore the graph. Another sensitive issue is how the graph will behave while being explored (like a zoom or selection directly on it) and its usability aspects.

When designing an algorithm for drawing any node-link diagram (not just trees), one must consider three categories of often-contradictory guidelines: drawing conventions, constraints, and aesthetics (Ward, Grinstein & Keim, 2010). Conventions may include restricting edges to be either a single straight line, a series of rectilinear lines, polygonal lines, or curves. Other conventions might be to place nodes on a fixed grid, or to have all sibling nodes share the same vertical position. Constraints may include requiring a particular node to be at the center of the display, or that a group of nodes be located close to each other, or that certain links must either go from top to bottom or left to right. Each of the above guidelines can be used to drive the algorithm design. Aesthetics, however, often have significant impact on the interpretability of a tree or graph drawing, yet often result in conflicting guidelines. Some typical aesthetic rules include: minimize line crossings, maintain a pleasing aspect ratio and minimize the total area of the drawing

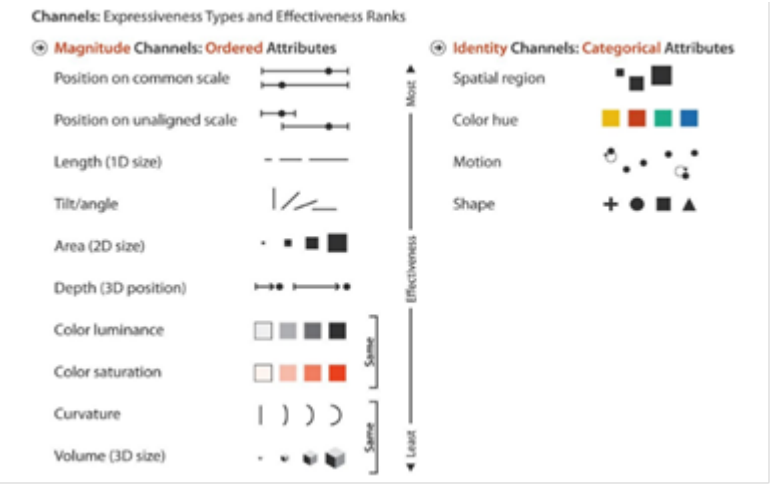
Force-directed algorithms (Kobourov, 2012) are among the most flexible methods for calculating layouts of simple undirected graphs. Also known as spring embedders, such algorithms calculate the layout of a graph using only information contained within the structure of the graph itself, rather than relying on domain-specific knowledge. Graphs drawn with these algorithms tend to be aesthetically pleasing, exhibit symmetries, and tend to produce crossing-free layouts for planar graphs. We considered this approach ideal to represent the existing relationship between doctoral advisors and mentor along the time.

The graph nodes can also offer additional information about the age of the advisor or mentor, or even if he is still alive or not. The color attribute is used to map this information. The end-user can drag, filter, select (by name, alma mater institution or knowledge field) and zoom the graph to have a better information insight. Annotation is also available.

Besides the relationship, we also assumed that the distribution per alma mater or field of knowledge would be also helpful to the end-user. The graph itself is not the best way of showing this information. Ordered bar charts would easily show these distributions.

Although most of the data is directly extracted from the infobox, some of it, has to be computed, like the number of students or the total per alma mater. The visual encodings where chosen taking into consideration the order of their effectiveness according the data attributes (Munzner, 2014). Figure 5 illustrates this order. Table 2 summarizes the visual encodings that were considered for the existing dimensions and measures in the data.

Figure 5 - Visual encodings.



Fonte: Munzner, 2014

Table 2 - Chosen visual encodings

Dimensions/measures	Visual encoding	Type
Name	Text	Nominal
Birth date	Node colour	Quantitative
Death date	Node colour	Quantitative
Citizenship	Text	Nominal
Nationality	Text	Nominal
Alma mater	Text	Nominal
Doctoral advisor	Text	Nominal
Number of students	Node diameter	Quantitative
Number of mentors/advisors	Node border width	Quantitative
Totals per *alma mater*	Bar size and gradient color	Quantitative
Totals per knowledge field	Bar size and gradient color	Quantitative
Count of relationships	Node size	Quantitative

Hue color associations have been the topic of significant study in psychology literature and recent supports the notion that the formation (Labrecque & Milne, 2012) and activation of color associations can be understood through models of semantic memory such as associative network. Although these studies are restricted in the number of colors and types of emotions and associations they test, the effects of colors remain relatively consistent across studies, which provides some empirical evidence of a systematic relationship between color and emotions and psychological functioning. Color is a relevant visual attribute that we use to map some of our quantitative data, such as the birth and death date or the alma mater or knowledge field distributions values. Finally, nominal data is mapped with text while size denotes the count of relationships or number of students.

The end-user interface is composed by two different types of panels. The main panel depicted in Figure 6 shows the interactive graph and presents filtering and thematic map selection tools. Each one of its input boxes gives auto-completion aid (figure 7) as soon as a

first letter is typed. It allows filtering by a combination between:

- Name: which allows the highlight of a particular scientist giving its name;
- *Alma mater*: allows the filter of scientists of a particular university;
- Fields: that allows the filtering of scientists by field of knowledge.

Figure 6 - Graph visualization, filters and thematic map selection.

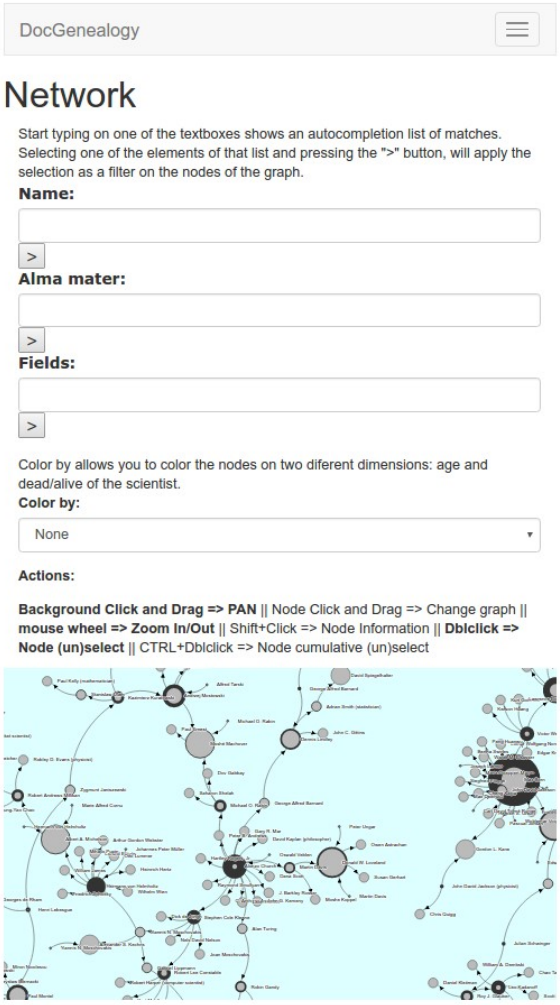
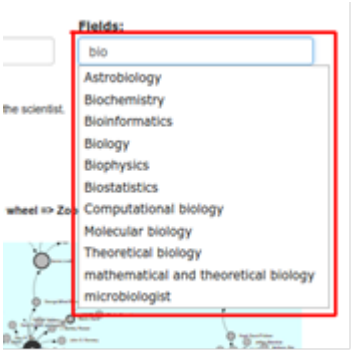


Figure 7- Auto-completion help.



A combo box labelled “Color by” is presented and allows the selection of two different dimensions of the data: age and dead/alive condition, both reflecting in the coloring of the nodes as seen in Figure 8. A second type of panels shows static frequency histograms on two attributes of the data: *Alma mater* and field of knowledge. Figure 8 shows the one regarding *Alma mater.*

Figure 8 - Thematic map on age.

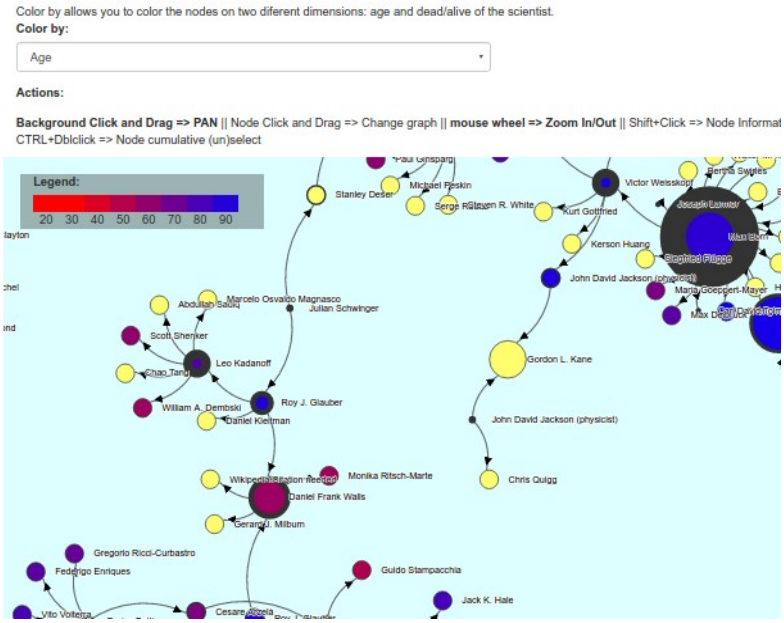


Figure 9 - Histogram of frequencies of scientists per university.

3.3 Technology

As for the data mining/web crawling process we have implemented a specific java application that executes direct http requests. The HTML content of each scientist page is

then explored using XML XPath queries (W3C, 2015), extracting the relevant data and organizing it in a JSON data structure suitable for direct consumption by the visualization framework.

As for the visualization and graphic interaction we used D3.js (Bostock, 2016a), namely its histogram functions and force-layout graph representation (Bostock, 2016b). D3.js is a JavaScript library for manipulating documents based on data. D3 helps to bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

Finally, a website was created ([http://address]) to show case the visualization and interaction features using plain HTML/CSS/JS. This first version of the prototype was used to run some initial evaluation tests.

4 MAIN CONCLUSIONS AND FUTURE WORK

We adopted the guidelines proposed by Lam et al. (2012) to define how the visualization process was going to be designed and conducted. Because our main goal is to assess a visualization tool's (or more specific, its data representations) ability to support visual analysis and reasoning about data to the end-users, we adopted the Evaluating Visual Data Analysis and Reasoning (VDAR) scenario.

In this case, outputs are both quantifiable metrics such as the number of insights obtained during analysis or subjective feedback such as opinions on the quality of the data analysis experience. The potential end-users is identified as general public, who wants to find out the relationship between doctoral advisors and mentors.

Because the study of how a visualization tool may support analysis and reasoning is problematic and the products of an analysis are hard to homogenize and quantify, we assume that a case study should be further designed to properly evaluate the effectiveness of DocGenealogy.

This paper introduced the DocGenealogy web-application. It is still an on-going work. It will be improved and extended to visualize the result of the mining of other types of Wikipedia's infoboxes contents, and thus support visually its interpretation and insight.

We have already conducted preliminary tests to evaluate its performance and its overall visual quality both in data mapping and aesthetics level besides its interface usability. Although we had a very positive global result, because our main target is to know its effectiveness to support the visual analysis and reasoning about the relationship between doctoral advisors, mentors and their students, we are presently designing a case study to test

and evaluate it more broadly and with a more fine grain in terms of its efficacy to support information insight.

REFERÊNCIAS

Beck, F., Burch, M., Diehl, S., & Weiskopf, D. (2014). The state of the art in visualizing dynamic graphs. EuroVis STAR.

Bostock, M. (2016a). Data-driven documents. https://d3js.org/. Accessed: 2016-1-15.

Bostock, M. (2016b). Force layout. https://github.com/mbostock/d3/wiki/Force-Layout. Accessed: 2016-1-15.

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. Information Sciences, 225, 1-17.

Dunne, C., & Shneiderman, B. (2013, April). Motif simplification: improving network visualization readability with fan, connector, and clique glyphs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3247-3256). ACM.

Heer, J., & Perer, A. (2014). Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. Information Visualization, 13(2), 111-133.

Henry, N., Fekete, J. D., & McGuffin, M. J. (2007). NodeTrix: a hybrid visualization of social networks. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1302-1309.

Jianu, R., Rusu, A., Hu, Y., & Taggart, D. (2014). How to display group information on node-link diagrams: an evaluation. Visualization and Computer Graphics, IEEE Transactions on, 20(11), 1530-1541.

Kang, U., Lee, J. Y., Koutra, D., & Faloutsos, C. (2014). Net-ray: Visualizing and mining billion-scale graphs. In Advances in Knowledge Discovery and Data Mining (pp. 348-361). Springer International Publishing.

Kobourov, S. G. (2012). Spring Embedders and force directed Graph Drawing Algorithms, university of Arizona.

Labrecque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: the importance of color in marketing. Journal of the Academy of Marketing Science, 40(5), 711-7.

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical studies in information visualization: Seven scenarios. Visualization and Computer Graphics, IEEE Transactions on, 18(9), 1520-1536.

Lee, B., Isenberg, P., Riche, N. H., & Carpendale, S. (2012). Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. IEEE Transactions on Visualization and Computer Graphics, 18(12), 2689-2698.

Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. Expert Systems with Applications, 39(12), 11303-11311.
Liu, H., & Motoda, H. (2012). Feature selection for knowledge discovery and data mining (Vol. 454). Springer Science & Business Media.

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. The Visual Computer, 30(12), 1373-1393.

Liu, H., & Motoda, H. (2012). Feature selection for knowledge discovery and data mining (Vol. 454). Springer Science & Business Media.

Moere, A. V., Tomitsch, M., Wimmer, C., Christoph, B., & Grechenig, T. (2012). Evaluating the effect of style in information visualization. IEEE transactions on visualization and computer graphics, 18(12), 2739-2748.

Munzner, T. (2014). Visualization Analysis and Design. CRC Press.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2014). A survey of multiobjective evolutionary algorithms for data mining: Part I. IEEE Transactions on Evolutionary Computation, 18(1), 4-19.

Rufiange, S., McGuffin, M. J., & Fuhrman, C. P. (2012, February). TreeMatrix: A hybrid visualization of compound graphs. In Computer Graphics Forum (Vol. 31, No. 1, pp. 89-101). Blackwell Publishing Ltd.

Shneiderman, B., & Aris, A. (2006). Network visualization by semantic substrates. Visualization and Computer Graphics, IEEE Transactions on, 12(5), 733-740.

Steichen, B., Carenini, G., & Conati, C. (2013, March). User-adaptive information visualization: using eye gaze data to infer visualization tasks and user cognitive abilities. In Proceedings of the 2013 international conference on Intelligent user interfaces (pp. 317-328). ACM.

Tran, T., & Cao, T. H. (2013). Automatic Detection of Outdated Information in Wikipedia Infoboxes. Research in Computing Science, 70, 183-194.

Ward, M. O., Grinstein, G., & Keim, D. (2010). Interactive data visualization: foundations, techniques, and applications. CRC Press.

W3C (2017). Xml path language (xpath). https://www.w3.org/TR/xpath/. [Online; accessed 30-April-2017].

Wikipedia (2017). Wikipedia robots.txt. https://en.wikipedia.org/robots.txt. [Online; accessed 24-March-2017].

Zhou, H., Xu, P., Yuan, X., & Qu, H. (2013). Edge bundling in information visualization. Tsinghua Science and Technology, 18(2), 145-156.

Zinsmaier, M., Brandes, U., Deussen, O., & Strobelt, H. (2012). Interactive level-of-detail rendering of large graphs. Visualization and Computer Graphics, IEEE Transactions on, 18(12), 2486-2495.