Supporting Visual Exploration of Interorganization Collaboration

Salli Hukkinen and Tomi Kauppinen

Department of Media Technology, Aalto University School of Science

Abstract. The increasing amount of linked data from organizations supports studying of underlying phenomena such as collaboration networks. Yet, the size of data, heterogeneous representations, and different granularity levels create challenges for supporting these tasks. In this paper we study whether and how Information Visualization can be used to support making sense of organization data. Our contribution is to study via a case study the visualization of implicit connections. The promise is that these serve as an evidence of intercollaboration between organizations. The key finding is that three levels of granularity, none of which were explicitly in the data, are needed to make sense of the data. Two of these levels came out as a result in the evaluation via a user study.

1 Introduction

The publication process of linked data is nowadays well understood but the need for new insights on how to use linked data has been raised [4]. The Linked Universities alliance[1] is an initiative to publish open, and linked data of universities. The idea is that data can be connected and reused worldwide for research and education purposes. The Linked Universities initiative has made possible the development and study of new linked data applications [4,**?**].

The Linked Open Aalto Data service[2] is Aalto University's effort to publish linked university data. In this paper we take this data into focus and create an information vizualization application to make sense of it. Because of its network nature, linked data is suitable for exploring social networks. In exploratory social network analysis, visualizations play an important role [7,2]. Information visualization plays an especially important role when dealing heterogeneous data [10]. We too aim to make sense of organization data via a visual approach. Our study experiments with an explorative data visualization of co-publishing networks in Aalto University. This *making sense* journey is exploratory, the starting point for the exploration being Aalto University's mission to be a university "where art and science meet technology and business". To support the mission, Aalto University is organized as a unity of six schools that represent different disciplines. We turned the vision statement into a question: "Where do interdisciplinary collaborations take place in Aalto University?" This is a valid question because

[1] http://linkeduniversities.org/lu/

[2] http://data.aalto.fi

the management needs to have an up-to-date picture of how the university is doing in terms of its strategy.

The challenge is that data on potential occurrences of collaboration (like on publications, projects, and courses) reside in distinct databases that are maintained by different schools. These registers do not contain explicit data on collaboration, even if transformed to RDF and published as linked data. In this sense the data itself is incomplete for answering the question about collaboration. Our approach to overcome this is by making implicit links to be present— links that are embedded in data on organizations, people, and publications.

The paper is structured as follows. Section 2 describes the approach to the problem of visualizing organizational Linked Data and explains the methods in use. Section 3 describes the evaluation setting which supported to improve the visualization. Section 4 explains the results we obtained and provides a discussion. Section 5 outlines the related work and Section 6 provides the concluding remarks.

2 Approach for Explorative Data Visualization

2.1 Linked Data about a University

We approach the question "where organizational collaboration takes place" by studying publications. Academic research efforts usually always target at publishing, thus publications co-authored by researchers from different organizations can be seen as an evidence of organizational collaboration. For this we decided to make use data on people, organizations, and publications to construct an interorganizational collaboration network. This kind of data is available in the Open Linked Aalto Data service[3].

Publication data is originally retrieved from two publication registers, one shared register maintained by the four technology schools (whose abbreviations are ELEC, ENG, SCI, and CHEM) and one maintained by the School of Business (BIZ). We limited the study to cover years 2008–2011. For the most recent years, publication data may be incomplete because articles are updated to the registers with a delay that can be up to 2 years. Therefore, we extend the study only up to 2011 to ensure that the data for the studied years are complete. Our analysis covers 13756 papers published during years from 2008 to 2011 in the four technology related schools of Aalto and in the School of Business. The publications of the School of Arts and Design (ARTS) are not available in the Linked Open Aalto Data service. We have access to their publications which are co-authored with some other Aalto school, but we do not have data on papers authored exclusively in the School of Arts and Design.

People data comes from the public Aalto People portal[4] where members of Aalto University personnel can create a profile. This means that the data is maintained by the users. Data about organizations, including the type of

[3] http://data.aalto.fi

[4] http://people.aalto.fi

organization and references to its parent and child organizations, are maintained in the same Aalto People portal database. Our analysis covers 1133 people who currently[5] both work in Aalto University and have a public profile in the Aalto People portal.

2.2 Visualizing Data

Because the aim is to visualize organizations and co-publications between organizations, a method for presenting graphs is appropriate where nodes represent organizations and links represent co-publishing activity between organizations. For this we made use of D3 (Data-Driven Documents) JavaScript library [1]. D3's force-directed graph layout (Force Layout) implementation is suitable for visualizing graphs and has been used in many examples online that can be used as reference [6].

The d3 force-directed graph layout appearance can be tuned with several parameters. The parameter values of our visualization evolved throughout several iterations. Of the parameters, *nodes*, *links*, and *size* were the fundamental parameters that needed to be specified. *Links* and *nodes* define what data is bound to the links and nodes of the force-directed graph, and *size* sets the force layout size in pixels. These parameters could not be used to tune the appearance of the visualization. The central parameters that were used to define the appearance of the graph were *charge*, *linkDistance*, and *gravity*. *Charge* determines how strongly nodes attract or repel each other. *LinkDistance* determines the length of a link between two nodes. *Gravity* defines how strongly nodes gravitate towards the center of the layout.

The first visualization version to which user feedback was collected is presented in Figure 1. In this graph, each node is a department of Aalto University. Aalto University is organized as six schools from different disciplines under which the departments and other units operate. Links represent co-publishing activity between two Aalto University departments that do not belong to the same school. Thus the visualization illustrates the cross-disciplinary co-publication network in Aalto univeristy during one year which in this Figure 1 is 2010. The number of this type of co-publications in a department determines the color of the respective node, and the count of such co-publications is visible on each node. University- or school-level organizations are not shown in this figure.

At this stage, the idea of visualizing co-publications received positive feecback from Aalto Institutional Relations unit which develops and maintains partnerships between the university and enterprises. After seeing the first version, their comment was that they would like to see a visualization where school nodes can be expanded to reveal departments, and collapsed again to a single school node, by clicking the node in the same manner as the force-directed graph showing the number of Aalto People portal users [7].

---

[5] At the time of writing this paper in February 2014

[6] `https://github.com/mbostock/d3/wiki/Force-Layout`

[7] See `http://data.aalto.fi/peopleforce.html`

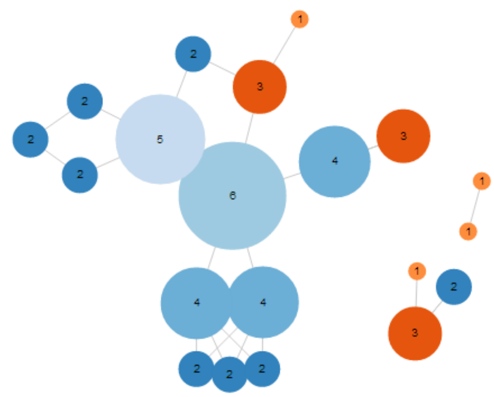

Fig. 1. Co-publication visualization—version 1

2.3 Improving the Visualization

In the next version, node coloring was coded according to school. This version is illustrated in Figure 2. A tooltip was implemented which shows the department name when the node is hovered over. Aalto People users application was referenced in the implementations of color-coding and tooltip. By tuning the parameters of the force-directed graph, the nodes settled themselves in a more informative way: the connections between nodes are easier to perceive when nodes don't overlap.

The final keystone in discovering a way to present departments grouped to schools, which was suggested in the initial user evaluation, was Ger Hobbelt's example of a Force Layout[8] which let's you "click to group and bundle nodes". In Hobbelt's method, data is first read to a "group map" which is a JavaScript object that contains each group's name and the list of nodes that belong to the group. Data is also read to two JavaScript arrays, *nodes* and *links*, that are given to the D3 Force Layout. The D3 Hull Layout is used to draw a lightly colored cluster that connects the nodes of one group (in other words, the convex hull of the node set). The graph is initialized with nodes collapsed in their groups. When a node circle is clicked, the respective group is expanded, the group's child nodes appear, and the convex hull that connects the group's nodes appears. Figures 3 and 4 present the result after Hobbelt's example was imitated (all schools are collapsed in Figure 3 and then expanded in FIgure 4). In the figures, abbreviations such as "ELEC" or "SCI" are the abbreviations of Aalto University schools. Departments, which are child organizations of schools, are presented with the entire name of the department, such as "Department of Media Technology".

[8] http://bl.ocks.org/GerHobbelt/3071239

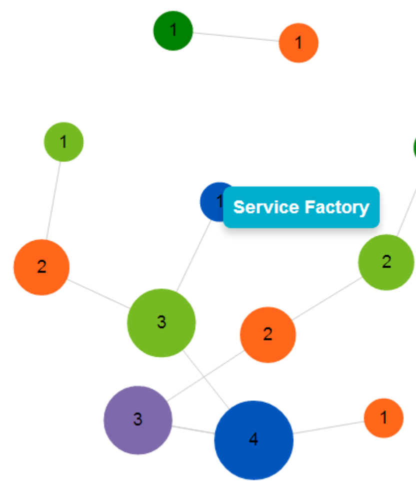

Fig. 2. Co-publication visualization—version 2

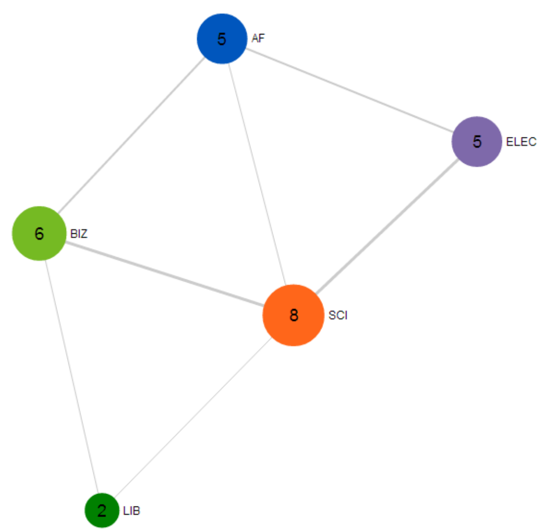

Fig. 3. Co-publication visualization—version 3. Schools collapsed.

So far the examples have illustrated the co-publication network in one year (year 2010 in Figure 1, and year 2011 in Figures 2, 3 and 4). We also want to examine the total number of co-publications during the years 2008-2011. This is possible because the application retrieves all data at once when it is initialized. When data is retrieved from data.aalto.fi, it is stored in a JavaScript object called *articles*. The *articles* object is used like a directory where years are keys and each key points to an array of distinct articles published in that year. Each

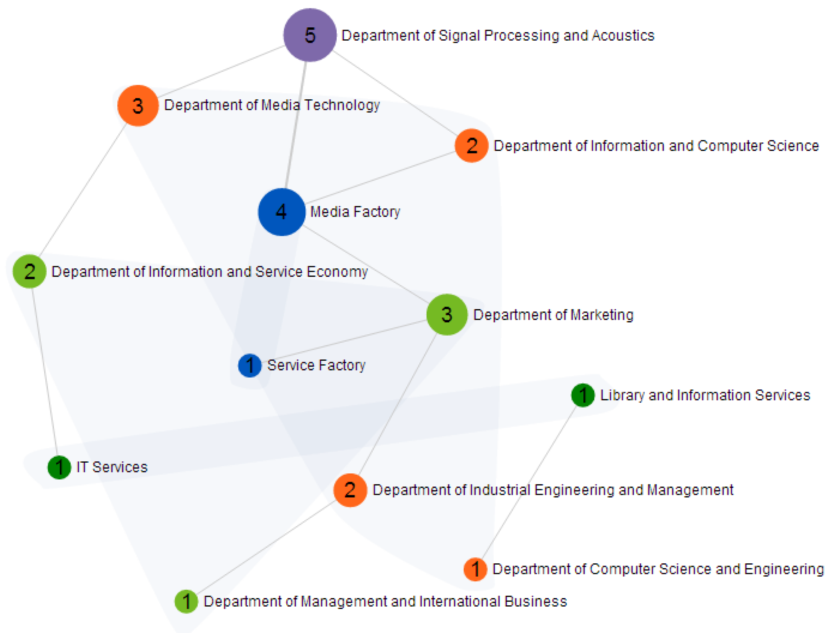

Fig. 4. Co-publication visualization—version 3. Schools expanded.

article has information about the contributing organizations. When a button is clicked in the application, the year of the button is used as key to get the right articles from the *articles* object. Nodes and links of the Force Layout are then re-computed from that year's article list. When the button "All years" is clicked, no single year is used as a key; instead, all entries of the articles object are read one after the other, and the force layout is re-computed using the article lists from all years (see Figure 5).

Three development ideas from the evaluation round were implemented (Figure 6). First, co-publications between departments of the same school were added to the visualization. Previously, only cross-school collaboration was illustrated; in the final visualization, all co-publishing collaboration between departments is shown, even within schools. Second, a tooltip was added that lists the titles of the articles that a department has published together with another Aalto department. The tooltip appears when the mouse is hovered over a node. Third, the total count of co-publication results was added to the header. Visualizing the higher number of co-publications between two organizations with a stronger link was dismissed from the final version. Including all cross-department co-publications increased the number of nodes, and various links sizes made the appearance messy in this case.

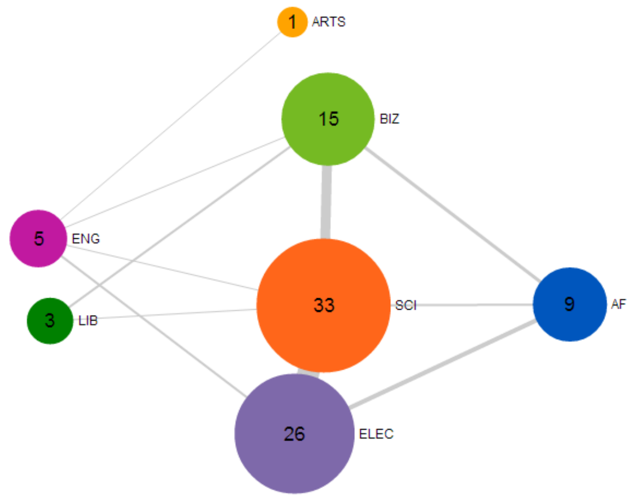

Fig. 5. Co-publication visualization—version 3. Aggregate of all years with schools collapsed.

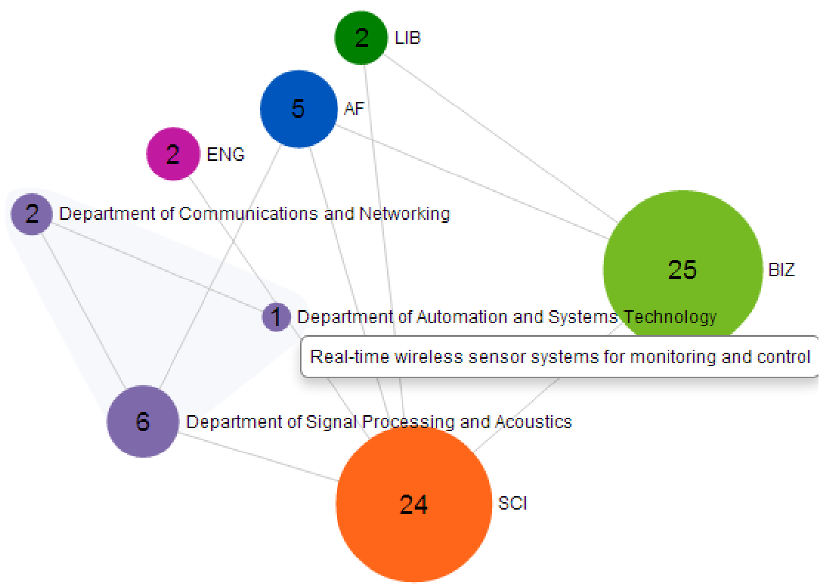

Fig. 6. Co-publication visualization—version 4. Aggregate of all years - ELEC expanded and other schools collapsed.

3 Evaluation

The evaluation was done as heuristic evaluation with the small modification that experts in the data or the domain were engaged instead of usability experts. The benefits of a heuristic evaluation are that it can be done at any stage of development and results can be used in an iterative process where the prototype is developed step by step based on evaluations [5]. Evaluators were allowed to inspect the visualization and feedback was collected through interviews in the evaluation. The evaluation was thus based on qualitative data.

The visualization was first shown to two employees from the Aalto International Relations unit and one from Aalto Institutional Relations unit. These units have an interest in understanding the collaboration networks of Aalto University because their role is to support the partnerships and the institutional collaboration of the University. The evaluators raised many suggestions on how the visualization could be improved. The essential finding was that evidence of research collaboration should be more explicitly visualized. An example feedback of this is the following quote:

> Is it possible to see research collaboration within schools too? It would be interesting to see the collaboration between the Department of Architecture and other departments of the School of Arts, Design and Architecture.

Next, the visualization was evaluated by the HR director from the top executive management team of Aalto University. The visualization received a positive and enthusiastic welcome. The visualization provided new understanding to the HR director:

> A nice surprise is the size of ELEC in this visualization. I expected SCI to dominate. [. . .] Sad is the absence of CHEM.

Data quality was discussed with the HR director. The visualization was able to reveal flaws in data. An example of a flaw that came up is that some departments are coded under a wrong school in Aalto Open Linked Data. Also, the visualization revealed that service providers such as IT Services and Library Services are coded as similar departments as the actual research departments. The kind of organizational data analysis conducted with our visualization would be more illustrative if different types of organizations were explicitly separated in the data, i.e., if research departments were separated from service providers.

4 Results and Discussion

The aim of this study was to experiment with a linked data application that combines and visualizes organizational data from different sources, to explore and make sense of siloed organization data—i.e. data from different separate sources—and to create new insights on interorganizational networks with the visualization. The visualization revealed a total of 148 papers that have been published by two or more Aalto departments during the years from 2008 to 2011. We were able to visualize the implicit links between resources and thus create information that was otherwise "hidden". We showed that with a visualization new insight can be derived from incomplete and siloed data.

The result of visualizing siloed organization data was a better understanding of the overall picture of organizational collaboration. Value to end-users was created as the visualization managed to suprise, inspire, and invoke new questions in users. Raising new questions means that the visualization had the power to provide new insight and to take the discussion to a next level and to more meaningful questions. The visualization managed to surprise users, which means it challenged existing understanding about the studied issues. At the same time, implementing the ideas from user feedback and thus answering the newly aroused questions was relatively easy. Indeed, the visualization became a tool for iterative exploration of organization data.

The iterative visual data exploration process led us to develop a visualization prototype that has the power to visualize three different levels of interorganizational collaboration: school-level collaboraton, department-level collaboration, and finally, topic-level collaboration through the titles of co-publications between organizations. An interesting avenue for further experiments will be to study the collaboration networks between different universities to find out which are the most interconnected universities.

Future trends are the increasing availability of public data and increasing demand for big data analysis [8]. Analytics will play an huge role in organizations. But to support management decision-making, the understanding created with analyses needs to be communicated effectively and essential information needs to be extracted and crystallized. A visualization is often the most effective way to do this. In this experiment we learned that a visualization can also facilitate an explorative data analysis. Involving users in the process gave valuable input and allowed the analysis to dive deeper in the topics that are relevant to the users.

The advantage of our web-based visualization tool is that it can be like a dashboard that gives an up-do-date view on the current data. In the evaluation, the organizational data visualization was seen as a potential management tool. A member of the executive management of Aalto University was inspired by the visualization and expressed the desire to show the tool to the entire management team. An idea developed during evaluation that if the visualization was published online it could function as an icentive for Aalto personnel to update their Aalto People profiles. The people data that the visualization uses are updated by the users of Aalto People portal. The possibility to make a visible impact on a published figure by reporting ones own work would be a motivation to update profiles in Aalto People portal.

The power of visualization has been recognized in story-telling too [12] and one idea is to use the visualization as a tool for communicating Aalto University mission and strategy. The visualization presented in this study has a narrative element: it presents a network for four consecutive years and thus illustrates the development of a research collaboration network.

There are some limitations in our data. A challenge with using people data that originates from Aalto People portal is that all members of Aalto university personnel have not created their own profile in the portal. Another challenge is that those who have a profile do not necessarily mention their organization in the profile. Last, people's links to organizations are their current affiliations. Even when publications from previous years are visualized, they can only be linked to their authors' current organizations. Publication data in turn is limited to publication archives of the four technology schools and the School or Business. Archives of the School of Arts and Design are not available in Linked Open Aalto Data service.

There are some further limitations in our approach for combining data. The following two assumptions are made in the study: (1) People are identified with first name and last name. If two people have the same first and last name, they are interpreted as one person. This is a fairly significant assumption because in a university with 5000 members of personnel[9] there are very likely some people with the same first and last name. (2) Publications are identified with their title. If two articles have the same title, they are counted as one article. This was not considered to be an issue since research publications are given detailed and descriptive names so finding two articles with the same name is extremely unlikely. No such incidents came into our knowledge in the experiment.

Last, the analysis was limited in terms of how many co-authors from one paper are examined. A paper might in principle have any number of co-authors. However, in practice analyzing more than five co-authors per article did not bring new results and only made the application slower. Testing with the years 2010–2011, analyzing the first five co-authors revealed all relevant publications. When exploring the year 2009 or earlier years the main problem was clearly the outdated people data. Therefore, the application was implemented so that it inspects the affiliations of the first five co-authors of publications.

Having acknowledged the limitations of our data, an important further finding was that a visualization can effectively reveal flaws and limitations of data. Thus it provides understanding on how data could be improved.

5 Related Work

We have described an exploratory visual data analysis process and shown how a visualization can provide new understanding about organizational networks and illustrate different levels of interconnectedness. Our visualization was able to arouse discussion and new questions about these networks. In a way, the visualization determined the direction of the analysis. Thus it is a very interactive analysis tool.

Other scholars have discovered the power of visualization to serve as an analytical tool and data exploration tool. Earlier visual explorative data analysis

[9] http://www.aalto.fi/fi/about/

experiments have studied biomedical data [3], geodata [6] , and medical data [11]. An explorative data visualization that is worth mentioning is NameVoyager [13]. It is a web application that visualizes baby name popularity and its development over time in the U.S. The application became instantly popular when it was published. NameVoyager [13] shows that a data visualization can be much more than a mere "task-oriented problem-solving activity" [14]. Wattenberg [13] explains that the application is used in a highly social way: the interactive visualization is so compelling that it is used more like a game than a plain visualization. Users also share and discuss online the findings they make with the visualization tool. For this reason, Wattenberg and Kriss [14] describe this behavior as "social data analysis". Also in Neumaier and Hinkenjann's experiment [11] with medical data analysis, a visual interactive data exploration tool invoked more discussion among a team of analysis than traditional spreadsheet analysis. Iterative exploration—for instance NetLens [9]—has been tested also earlier. Our approach deals is different in a sense that it deals with online Linked Data.

What is more, with NameVoyager [13], surprising patterns in name trends have been identified, such as the booms and declines of some familiar names. These findings are remarkable because they provide new viewpoints to things that are taken for granted. Thinking about a familiar-seeming name without using NameVoyager, the first question in mind is not "when was the name invented" because the popularity of the name seems self-evident. The visualization could reveal that seemingly stable popularity is not stable at all. Similarly, our visualization tool invoked questions such as "what about higher level organizational collaboration" and "which departments do not co-publish with other Aalto departments". A visualization can answer questions that would otherwise be unanswered or never even asked.

6 Conclusions

More and more data is worthless if it does not provide new understanding and new insights for people. When it comes to linked data, visualization can be the way to make tangible and understandable the information that is embedded in links. Visualization makes it easier to explore and make conclusions about linked data. Visualization can be a way to see things that would otherwise not be seen and to get people to ask questions that would otherwise not be asked. A visualization can also increase the level of interest people have in data. The question where cross-disciplinary research happens in Aalto Universty could not be answered comprehensively due to limitations in the data, but was it the point of this study anyway? After all, strong indication was obtained about the most active departments in this area, and most importantly, discussion was engendered about how to make available all the missing information to get the full picture of co-publishing in Aalto University.

While the greatness of linked data is that it can be combined from different sources, this experiment has pointed out that it is also the very challenge with linked data. Knowing if two things are the same is difficult if the data doesn't explicitly indicate so. In principle linked data has the needed features for doing this, so the purposes for which linked data can and will be used rest on the providers of linked data. To really be linked, a linked data resource should indicate what resources are the same with itself.

References

1. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.
2. Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj. *Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences)*. Cambridge University Press, January.
3. Janez Demsar, Gregor Leban, and Blaz Zupan. FreeViz - An intelligent multivariate visualization approach to explorative analysis of biomedical data. *Journal of Biomedical Informatics*, 40(6):661–671, DEC 2007.
4. Mathieu d'Aquin. Putting linked data to use in a large higher-education organisation. *Interacting with Linked Data (ILD 2012)*, page 9, 2012.
5. C. Forsell. Evaluation in information visualization: Heuristic evaluation. In *Information Visualisation (IV), 2012 16th International Conference on*, pages 136–142, July 2012.
6. Julia Gonschorek and Lucia Tyrallová. Geovisualization and geostatistics: a concept for the numerical and visual analysis of geographic mass data. In *Computational Science and Its Applications–ICCSA 2012*, pages 208–219. Springer, 2012.
7. Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. Nodetrix: A hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, November 2007.
8. Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Enterprise data analysis and visualization: An interview study. In *IEEE Visual Analytics Science & Technology (VAST)*, 2012.
9. Hyunmo Kang, Catherine Plaisant, Bongshin Lee, and Benjamin B Bederson. Netlens: iterative exploration of content-actor network data. *Information Visualization*, 6(1):18–31, 2007.
10. D.A. Keim. Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):1–8, Jan/Mar 2002.
11. Nils Neumaier and André Hinkenjann. Explorative analysis of medical study data. *VR*, 5:237–240, 2005.
12. Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.
13. Martin Wattenberg. Baby names, visualization, and social data analysis. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 1 –7, 23-25 2005.
14. Martin Wattenberg and Jesse Kriss. Designing for social data analysis. *IEEE Trans. Vis. Comput. Graph.*, 12(4):549–557, 2006.