VISUALIZATIONS OF SLOVENIAN SCIENTIFIC COMMUNITY

Mario Karlovčec, Dunja Mladenić, Marko Grobelnik, Mitja Jermol
Jozef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel: +386 1 4773419; fax: +386 1 4251038
e-mail: mario.karlovcec@ijs.si

ABSTRACT

Using advanced analysis techniques new useful insight into data can be achieved. This paper addresses a problem of gaining insights in the data on scientific collaboration on a National level, where data can be seen as a graph with researchers and research content. Two existing visualization techniques were applied on data about scientific community in Slovenia: collaboration diagram and competence map. Collaboration diagram gives a clear overview of collaborations for a selected researcher, while competence map shows semantically grouped research content the researcher has worked on.

1 INTRODUCTION

Different types of data about scientific community are being collected on National level in many countries. This includes data about researchers, their projects and organizations. With more advanced analysis techniques user can gain deeper understanding of different aspects of scientific community. There are at least three types of users of such data: (1) researchers – they are interested in finding possibilities for collaboration, comparing achievements of colleagues, observing profiles of organizations; (2) investors – they want compare investments and results of projects, directions of development in science, etc. (3) companies – want to see latest achievements in science which they can commercialize, search for experts in particular fields, etc.

In the next chapter detail description of the data is given, following two different visualization techniques: collaboration diagram and competence map.

2 DATA

The data that we have used for evaluation was provided by Slovenian Research Agency (ARRS), and can be publicly accessed with SICRIS (SlovenIan Current Research Information System , 2000) information system managed by Institute of Information Science, Maribor. Main entities contained in the data are: researchers, research organizations and research projects. In 2011, there are 13897 researchers who have taken part in Slovenian Research Agency projects since 1998, or whose active status has been either registered or confirmed by the research organizations. Researchers/experts not being funded by the Slovenian Research Agency are also able to join SICRIS if they provide the data. Second entity – research organizations contains list of all (901) research organizations inside the country carrying out projects partly financed by Ministry of science from 1995 onwards. The complete data, however, is available only for the organizations that responded to the invitation of 14 June 1999 and submitted the required data. Third entity is research projects and it contains data on 5389 projects partly financed by Slovenian Research Agency from 1998 onwards. In the future it will also include data on any other research projects that the researchers from the database will wish to present.

3 COLLABORATION DIAGRAM

Collaboration diagram[1] is a type of sociogram that visualizes collaboration between people. **Sociograms** serve as simple visual illustrations in helping people to make sense of the underlying network information (Stuart K. Card, 1999); they are the most important instrument, basis for visualization of social networks and analysis of social ties structure (Wouter de Nooy, 2005).A **social network** is a collection of actors (such as people, organizations or other social entities) and relationships among the actors, indicating the way in which they are connected socially (such as friendship, trade or information exchange). **Social network analysis** is a methodological approach to understanding the structure of such networks, by means of mapping and measuring these relationships (Wasserman S., 1994).

3.1 Main Characteristics of The Diagram

Proposed collaboration diagram is different from other sociograms that represent collaboration between people. The main difference is focusing on a single actor (researcher in our case). This means that the chosen researcher is placed in the middle of the diagram (central researcher) and all others, with which he is collaborating, are placed radially around

[1] The usage of term collaboration diagram in the context of SNA should be differentiated with one type of UML diagram the same name

him with the distance relative to intensity of collaboration. With this drawing convention, we lose information about collaboration between researchers that are not central, but increase readability for a single researcher. The argument for this approach is assumption that user is not interested in exploring collaboration of multiple researcher in the same time, but is always focused on a single researcher. Even if this assumption is wrong and user wants insight into collaboration of multiple researchers in the same time, because of too many nodes and connections between them, the diagram would not be clear and the information would not be communicated effectively. Furthermore the diagram would not clearly show the collaboration of the single researcher, which was maybe in primary focus of the user.

As given in (Weidong Huang, 2005) the major concern in network visualization is effectiveness, it includes two issues: readability and communication. Readability can be strongly affected by network layout. Among many factors, edge crossing has been widely accepted as a major aesthetic affecting the ease of reading. In proposed approach for drawing collaboration diagram, crossing edges are completely avoided. Second issue of effectiveness – communication, refers to communicating exactly the same information which is intended to convey with the visual representation of the underlying network (Ulrik Brandes, 1999). Radial layout with a central actor is inteded to clearly communicate structure of collaboration.

In (Weidong Huang, 2005) authors compare five sociogram drawing conventions: circular, radial, hierarchical, group and free. The radial layout (figure 1) had moderate user ratings and task performance. But the radial layout used in these experiments differed from our approach for drawing collaboration diagram. Single actor was placed in the center, but the radial layout of the others was not clear. To emphasis this layout, a circle was drawn in the edge level of the network. Observations of results in referenced research show that the subjects did not realized nodes were actually arranged radially. A big circle around the diagram did not make viewers read radially, it caused additional confusion.

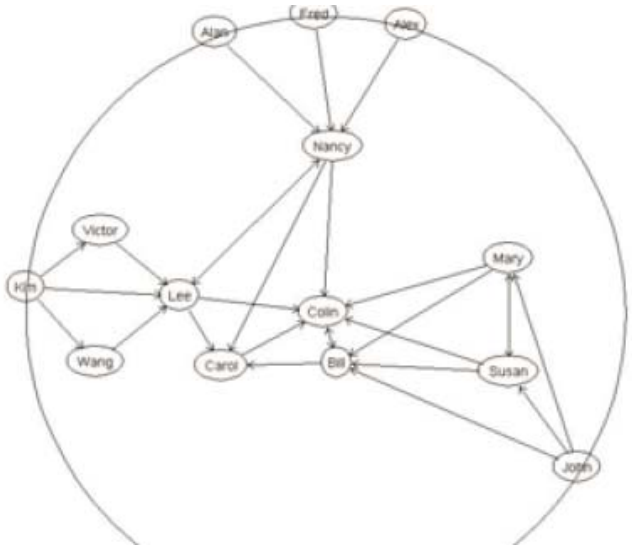

Figure 1: *Radial layout used in experiments in* (Weidong Huang, 2005)

3.2 Process of Creating Diagram

Proposed approach for drawing collaboration diagram can be performed in the following steps (Figure 2):

1. Place the chosen actor in the center of the network
2. Divide the maximal allowed radius number of different groups of collaboration count to get the distance between circles
3. Draw circles around the central actor so that each circle is a multiple of distance calculated in previous step
4. Draw actors of each group of collaboration count on a corresponding circle. The angle of the first actor of a group is determined randomly, while the others are evenly spaced on the circle.

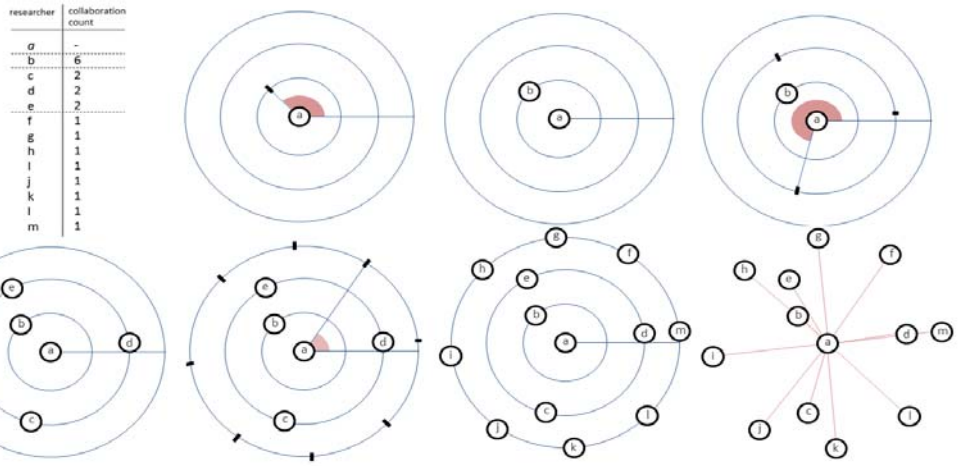

Figure 2: *Steps in creating collaboration diagram*

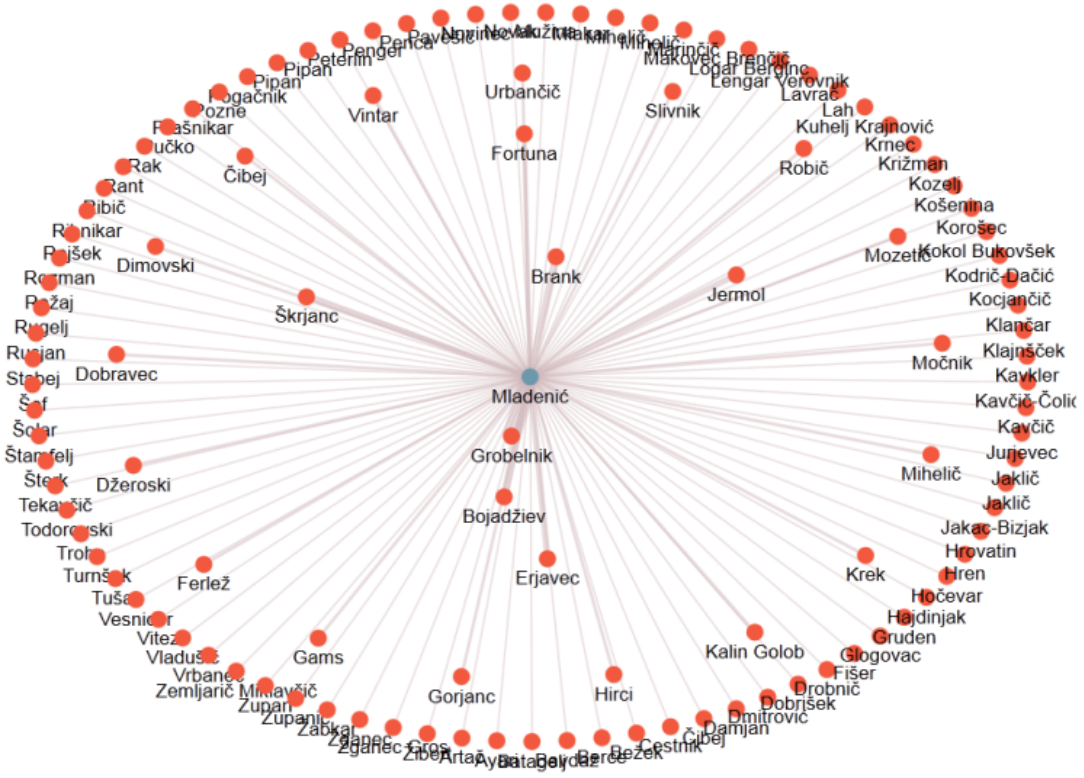

Figure 3 - *Collaboration diagram of a researcher*

An example of collaboration diagram on our data is shown in Figure 3. Selected researcher (Mladenić) is located in the center of the diagram and marked with different color than the other researchers she collaborates with. The diagram shows that the selected researcher collaborated with a number of other researchers on some projects. The greater the number of common projects with the selected researcher, the greater is the proximity to the center. For instance, Grobelnik, Brank, Bojadžiev, Erjavec, Jermol have been the most intensive collaborators and are thus places closer to the center of collaboration diagram.

4 COMPETENCE MAP

Competence map is a diagram used to represent intellectual achievements, skills or knowledge of a person, organization or some other entity. Proposed competence map is directed to a single person, which is in our case a researcher. Competence map of a researcher is based on scientific projects on which she has worked on. To build a competence map we have used Document Atlas (Blaz Fortuna, 2011). Document Atlas is a utility for visualizing large corpora of text documents using Latent Semantic Indexing and multidimensional scaling for positioning the corpus on a 2D plane. Input for creating competence map of a researcher are titles and when available, abstracts and keywords of projects on which he has worked on. The output are points which represent project, which are positioned on the 2D plane in such a way that it is easily to recognize most important groups of competences researcher is working on.

An example of a competence map on our data is shown in Figure 4. It can be seen that the researcher has collaborated on several projects, six of them (in the middle cluster) involving data analysis, two projects targeting Slovenia (the left up cluster), two involving semantic technologies (the right most cluster), three projects related to military topics and two projects not belonging to bigger clusters.

Figure 4: *Competence map of a researcher*

5 CONCLUSION

This paper describes visualization techniques applied on data about scientific community in Slovenia. We have shown how these visualizations enable new insights into the data. Two visualization techniques were used: collaboration diagram and competence map.

Collaboration diagram clearly communicates collaboration of a researcher. Novel approach, which focuses on a single actor, is used to construct the diagram. In this way, information contained in the diagram is decreased, but its effectiveness and ability to amplify users' cognition is increased.

Competence map shows projects of a researcher, which are visualized in the way that the user can recognize most important groups of competences researcher is working on. It is build using Document Atlas utility. Described visualization techniques enable deeper understanding of different aspects of scientific community in Slovenia.

Further visualizations will be developed in the future work, in order to provide additional knowledge to researchers, investors into scientific work and users of scientific achievements in practice.

6 ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency, the IST Programme of the EC under PASCAL2 (IST-NoE-216886) and MetaNet (ICT-NoE-249119).

References

[1] Blaz Fortuna, D. M. (2011). *Document Atlas*. Retrieved 8 4, 2011, from http://docatlas.ijs.si/

[2] Stuart K. Card, J. D. (1999). *Information Visualization Using Vision to Think.* San Francisco: Morgan Kaufmann Publishers, Inc.

[3] Ulrik Brandes, P. K. (1999). Centrality in Policy Network Drawings. *Proc. 7th Intl. Symp. Graph Drawing*, (pp. 250-258).

[4] Wasserman S., F. K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.

[5] Weidong Huang, S.-H. H. (2005). *Layout Effects: Comparison of Sociogram Drawing Conventions.* Sydney: University of Sydney.

[6] Wouter de Nooy, A. M. (2005). *Exploratory Network.* Cambridge: Cambridge University Press.