Visualizing Collaborations and Online Social Interactions at Scientific Conferences for Scholarly Networking

Laurens De Vocht, Selver Softic, Anastasia Dimou, Ruben Verborgh, Erik Mannens, Martin Ebner, and Rik Van de Walle

In: Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume, 1053–1054, 2015.

http://doi.acm.org/10.1145/2740908.2742030

Visualizing Collaborations and Online Social Interactions at Scientific Conferences for Scholarly Networking

Laurens De Vocht* Selver Softic† Anastasia Dimou* Ruben Verborgh*
Erik Mannens* Martin Ebner† Rik Van de Walle*

*iMinds - Multimedia Lab
Ghent University, Belgium
{firstname.lastname}@ugent.be

†IICM - Social Learning Group
Graz University of Technology, Austria
{firstname.lastname}@tugraz.at

ABSTRACT

The various ways of interacting with social media, web collaboration tools, co-authorship and citation networks for scientific and research purposes remain distinct. In this paper, we propose a solution to align such information. We particularly developed an exploratory visualization of research networks. The result is a scholar centered, multi-perspective view of conferences and people based on their collaborations and online interactions. We measured the relevance and user acceptance of this type of interactive visualization. Preliminary results indicate a high precision both for recognized people and conferences. The majority in a group of test-users responded positively to a set of statements about the acceptance.

1. INTRODUCTION

Social media used by researchers resulted in the emergence of alternative scientific networks beyond the traditional co-authorship and citation networks. However, the various ways of scientific interaction, including these with collaboration tools (e.g. Mendeley[1], ResearchGate[2] and social media (e.g. Twitter [3]) are reflected, but remain distinct from the scholar networks formed in the frame of their publications. Co-authorship, citation and social media based networks are rarely associated, let alone combined in a single visual interface. Social media captures an aspect of conferences that proceedings do not, they reflect the "talk" and networking that goes on during and in-between presentations.

Researchers, as other twitter users, tend to adopt hashtags to create threads of communication around a certain topic, e.g. *#SemWeb* or *#savesd15*. When used appropriately, searching for these hashtags returns messages that belong to the same conversation (even if they do not contain the same keywords). Results are promising concerning the compliance between Twitter hashtags and URIs, and detecting concepts and entities valuable to be treated as new identifiers [3, 5]. Applying semantic modeling for Twitter data led to identifying hashtags as good resolvers for the retrieval of information and a solid interlinking base with the rest of the Linked Data Cloud [6]. For this kind of data, an exploratory visualization scenario to academic metadata is applicable and useful [1]. Exploratory visualization is the process of creating maps and other interfaces while dealing with relatively unknown data [4].

Our visualization finds its application in "scientometrics", the study of measuring and analysing science, technology and innovation [7]. The novelty in our approach in this context lies in combining Twitter data with co-authorship and conference data [3]. It specifically relates to the challenge of detecting interesting people in a community of interest where it is useful to have a common research focus and thereby using Twitter as a real-time source. This includes identifying how a researcher's network is structured through collaborations (i.e. co-authorship) and how this is reflected in online interactions and who is joining the conversation that might be relevant, before, during and after conferences. Furthermore we verified the relevance of the content presented to the user and validated the acceptance of the way it was visualized.

2. VISUALIZING SOCIAL AND BIB DATA

Aligning event data (COLINDA[4]), social media data (Twitter) and publication data (DBLP[5]) forms the foundation for combining recognized conferences tags and Twitter accounts in a single visualization. This is driven by the result that conferences and people can be accurately recognized and interlinked with corresponding authors [2]. In our approach, exploratory analysis methods are used to visualise the network around researchers. Our visualisation achieves aligning traditional research networks and networks as they emerge based on data from social media, providing a unique perspective of researchers multi-modal interactions.

The screenshot in figure 1 depicts the network of a researcher. The scholar is centered with the blue node and around it are other

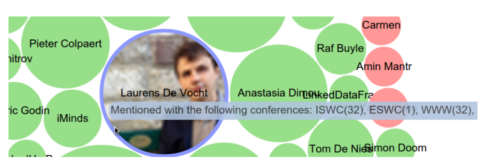

Figure 1: The scholar is centered in the middle and the network is visualized in nodes around the central (blue with picture) node.

related people (the more co-mentions, the more nearby they are positioned). The size of the scholar is in the middle between the

[1] http://www.mendeley.com
[2] http://www.researchgate.net
[3] http://www.twitter.com
[4] http://colinda.org
[5] http://dblp.uni-trier.de

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742030.

minimum and maximum size of a node. The more publications someone coauthored with the scholar, the bigger the node. After the researcher has signed in with their Twitter account on ResXplorer they can check recent interactions. A video is available at `http://youtu.be/QopnPvWIFzw`. A tooltip displays facts about the collaborations (e.g. co-authorships and mentions), i.e. the number of mentions for a specific conference and the the number of co-publications.

We interlinked research oriented datasets such as DBLP and COLINDA with data from social media containing information about conferences and social profiles of researchers. This last data was extracted on-the-fly just before generation of the visualization, this way the visualization always shows the latest results. We used common vocabularies (such as FOAF[6], SIOC[7], SWRC[8] , and the Dublin Core[9]) to annotate tweets from user profiles as Linked Data. We filtered the latest 200 tweets from the user's timeline and home-timeline, to find those containing matching hashtags corresponding with conference abbreviations (provided by COLINDA). Furthermore, each user that is mentioned in a tweet or is the creator from a tweet is identified and linked as a person. For the interlinking, we used same techniques as the ones we used before to align scholar profiles with data from web collaboration tools [2]. After extracting and converting the tweets, we identified mentions and hashtags which we interlinked with the scholar's bibliographic record (on DBLP). We matched each scholar together with one of their co-authors, if mentioned in a tweet, with the scholar's bibliographic record. The result is a graph of people containing links to the conferences and co-occurrences of mentions.

3. PRELIMINARY RESULTS

For the evaluation of the visualization we were interested in two aspects: **relevance**, by observing the *precision* and *recall* of the visualization, essential to validate that the presented information is sufficiently applicable to the user; and **acceptance**, a user survey to verify that the visualization is *usable*, *useful* and might lead to more *effective* scholarly networking.

3.1 Relevance

To measure precision and recall we asked *10* researchers in computer science (who visited and contributed to at least one computer science related conference in 2014 and use Twitter) to complete a set of tasks where they indicated how they judged the visualized nodes. The visualization for all test-users combined resulted in *217* recognized people and *29* recognized conferences. We got a high precision for conferences (*0.97*), because almost all detected conferences were correct. There was still a low recall (*0.56*), as many conferences were missing according to the users. We noted a moderate recall (*0.83*), a relatively large number of people in the network (co-authors especially) were missing because they were not mentioned together - however users expected them in the visualization. The precision for recognizing people is high (*0.92*): people who where connected indeed belonged to the researchers network or the users considered adding them to their network.

Noteworthy is that overall test-users in total, discovered *19* (out of *217*) people they considered adding to their network. Not everybody was presented the way test-users expected: they indicated that *37* people were missing. This implies that the coverage does not extend to people that do no have a Twitter account. However, users could increase the number of people in their visualization by actively tweeting and interacting with the people they consider relevant. This implies that for conferences where Twitter is not common, the results are definitely less interesting for the users.

3.2 Acceptance

We created a set of statements by applying the Technology Acceptance Model to measure effectiveness, usefulness and usability. The same *10* test-users completed the survey by answering the questions on a Likert-scale from 1 to 5, indicating the degree to which users agree with the statement or find it likely. The users response is varied, but on average and the majority (scores between 3 and 5) agreed or found most of the statements likely for all of the parameters measured. A small portion of the users were not immediately convinced by the usefulness (scores between 1 and 3), mainly because for them the visualization returned few results. Some indicated this was because they were only passively using Twitter, or Twitter was not used at all during the recent conferences they visited.

4. CONCLUSIONS AND FUTURE WORK

We found that the information presented in the visualization has high precision, both for people and conferences. However, the recall is moderate for people and low for conferences. This is due to the many missing conferences, typically because Twitter was not used very often or because there was a lot of noise (unrelated tweets) preventing the detection of the relevant context hashtag. The strength of the visualization lies in its fairly effective mapping as perceived by users. The more scholars use Twitter and use it to interact with others in the context of conferences, the more relevant results they will see in the visualization. We consider to extend the number of tweets taken into consideration, to obtain larger networks, especially for users who tweet often. However, relevancy does not tell everything, there is also a serendipity factor involved, for example showing a few targeted people to users might lead to an overall better acceptance without requiring to increase the amount of relevant results presented.

5. REFERENCES

[1] L. De Vocht, A. Dimou, J. Breuer, M. Van Compernolle, R. Verborgh, E. Mannens, P. Mechant, and R. Van de Walle. A visual exploration workflow as enabler for the exploitation of Linked Open Data. *Proceedings of the 3rd Workshop Intelligent Exploration of Semantic Data*, 2014.

[2] L. De Vocht, S. Softic, E. Mannens, M. Ebner, and R. Van de Walle. Aligning web collaboration tools with research data for scholars. In *Proceedings of the of the 23rd int'l conference on World Wide Web Companion*, pages 1203–1208, 2014.

[3] L. De Vocht, D. Van Deursen, E. Mannens, and R. Van de Walle. A semantic approach to cross-disciplinary research collaboration. *iJET*, 7(S2):22–30, 2012.

[4] M.-J. Kraak. Exploratory visualization. In *Encyclopedia of GIS*, pages 301–307. Springer, 2008.

[5] D. Laniado and P. Mika. Making sense of Twitter. In P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, and B. Glimm, editors, *International Semantic Web Conference (1)*, volume 6496 of *Lecture Notes in Computer Science*, pages 470–485. Springer, 2010.

[6] S. Softic, M. Ebner, H. Mühlburger, T. Altmann, and B. Taraghi. @twitter mining# microblogs using# semantic technologies. *6th Workshop on Semantic Web Applications and Perspectives*, pages 1–9, 2010.

[7] A. F. Van Raan. Scientometrics: State-of-the-art. *Scientometrics*, 38(1):205–218, 1997.

[6] `http://xmlns.com/foaf/spec/`
[7] `http://rdfs.org/sioc/spec/`
[8] `http://ontoware.org/swrc/`
[9] `http://dublincore.org/documents/dcmi-terms/`