



WT2.14: Universidad Carlos, III de Madrid, Computer Science Department, Madrid, Spain Report

Activities performed during the visit

in Wrocław University of Technology

period: 21.07.2014 -14.09.2014.

author: Radosław Michalski



Personal Information

Dr. Jose María Alvarez Rodríguez, faculty member of Universidad Carlos III de Madrid, Spain, visited Wrocław University of Technology, Poland, in the period from 2014-07-21 to 2014-09-14 in order to stay for a long-term research visit.

Information about Seminar

Dr. Jose María Alvarez Rodríguez gave a number of seminars during his visit. Starting from describing his research background through how the scientific career looks like in Spain up to summarizing his scientific activities at the end of the visit.

Description of scientific activity in details

This research visit was programmed in the framework of collaboration between the Carlos III University of Madrid, Spain and the Wrocław University of Technology in the ENGINE project. Before starting the secondment a research proposal made by Jose María Alvarez was sent to the colleagues in the Wrocław University of Technology as a first contact to build on it. More specifically, the research proposal was entitled as: “Community detection in graph-based structures. Study and application in social network analysis and quality analysis of critical systems.” The abstract of this proposal is also presented below:

Last years have seen a growing interest in the improvement and extension of existing graph-partition and graph topology analysis techniques with the aim of getting a better understanding of complex systems such as social networks/on-line communities, biology systems or critical systems development (traceability). In general, one of the main features of graph-based structures to represent complex systems lies in the (implicit/explicit) creation of communities or clusters. In this sense and according to the distribution of nodes and edges as well as the graph density in certain areas enable us the possibility of inferring and activating independent information areas defined by a set of particular features that can be 1) static or explicit ones and 2) dynamic or implicit ones. In the same way, these features are expected to be extracted by a local or a global analysis of the underlying graph structure. In both cases, graph partition, following the aforementioned approaches and features, is a complex task and although some techniques have been designed for different scopes, a general solution is still under study. That is why this proposal seeks for making progress in the state-of-the-art by studying the key-elements for community detection and graph partition. To do existing techniques such as *“hierarchical, partitional or*



spectral clustering” and others based on statistical inference or in the algorithm of “Girvan and Newman” will be reviewed to thus design and implement a new solution that allow to get better results in terms of accuracy and time. On the other hand and with the aim of validating the purposed approach, a complete test bed will be designed to assess the behaviour of the algorithm in two main scenarios: 1) analysis of on-line communities and more specifically community detection in social networks (an implicit graph comprising thousands of nodes and edges) in which both pre-defined and inferred features (e.g. based on sentiment analysis techniques) will be used to perform graph partition and to detect on-line communities. As an example, the technique should be able to detect interaction between communities and detect hostile or violent groups. Another possible benefits from the above approach may be finding influential nodes or groups or predicting groups evolution. This part of the research work will be supervised by the Prof. Kazienko. On the other hand, the second scenario comprises the analysis of critical systems development (e.g. engineering systems in aerospace, automotive, etc.) in which interconnected data and information are continuously being generated such as requirements, models, changes, test cases, quality metrics, etc. The goal of this second scenario is to apply the designed technique to enable system traceability and quality analysis by inspecting the implicit graph structure with the aim of ensuring the safety criteria. This part of the research work will be supervised by the Prof. Llorens. As conclusion the study and development of this research work is expected to be a key driver from both points of view: basic and applied research. That is why we are purposing here a new solution to tackle a problem that can be found in two different and real scenarios such as social network analysis and critical systems development.

In order to prepare the visit and accommodate the topic of the proposal and the background of the visitor, some conversations were kept between Jose María Alvarez y Radoslaw trying to find the best research cluster within the research group within the Wrocław University of Technology (hereafter WR). In this very beginning preparation of the stay, both researchers agreed about the possibility of making some changes with the aim of boosting at the most the collaboration between both parties. In this sense, Radoslaw as representative of the research group Poland always showed his willingness to find the most proper and adequate research lines. Once a draft of the research activities was fixed, all administrative and accommodation issues were prepared and organized by the WR avoiding in this way very time consuming activities for a visitor in a foreign county.

On the other hand and taking into account the activity of the WR research group organizing different events such as the “The First European Network Intelligence Conference (ENIC 2014)” or the summer school “Machine Learning for Social Media Analysis 2014 (MLSMA'2014)” and their schedule in the dates that Jose María Alvarez was going to be there (between July and September), the WR research group also suggest Jose María Alvarez to submit a paper to this conference and participate in the Summer School. Finally and due to the academic activities of Jose María Alvarez he could not attend both events but a first paper was accepted to be presented in the ENIC 2014 Conference in a very close topic related to the initial research proposal.

After arriving in Wrocław the 21st of July, 2014 the research stay started the day after, Jose María Alvarez was received in the building of the WR research group and everything was ready to start work such as a private room, a mug and office stuff, see Figure 1. Jose María was introduced to all members of the team and he could enjoy of this very warm welcome, see Figure 2. Taking into account that some of the members of the team were involved in different projects, research lines, etc. and some of the them were in different conferences and events we agreed to organize a kick-off meeting one week after to thus present all on-going activities and evaluate real possibilities of collaboration in some of the



research lines carried in the WR research group. The first week then continued with the preparation of this meeting as main activity and the first steps in collaboration through short talks in different moments such as coffee breaks, lunch, etc. Here it is important to remark that all members were very willing to explain their on-going research.

After that and during the kick-off meeting, some of the members of the WR research group (one per research line) presented in short presentations the motivation, on-going work and outcomes (Tomasz, Piotr, Radoslaw and Prof. Kazienko). More specifically, Prof. Kazienko as main researcher also commented the working methodology, objectives in the short/mid/long term and the regular schedule such as a shared lunch all Wednesdays. After these first introductory presentations (multi-relational machine learning, social influence, social network analysis, etc.), Jose María Alvarez presented their background and experience making emphasis in some of the preliminary potential research lines. This helps to discuss about how to align the initial research proposal to some of the existing research activities. In this sense Prof. Kazienko suggested that the most adequate research line would be the one led by Tomasz in multi-relational machine learning due to the fact the Jose María Alvarez had purposed something to study networks multimode and multilayer (WR members rapidly detected the context of the underlying problem). This first meeting was about 3 hours and it served as a very good icebreaker for collaboration. As another outcome of the meeting was the need of gathering data about development of software projects to being able to apply the techniques designed and developed by the WR group.



Figure 1 Room for a visiting researcher.

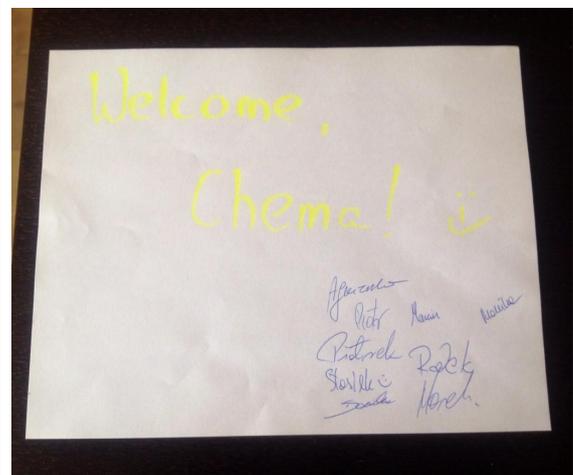


Figure 2 Warm welcome.

Next days and weeks continued with some short discussion to focus the data required to apply multirelational machine learning in the context of software systems. Due to the fact that the information about the development of software projects is not publicly available and after checking different data sources such as the statistics of FP7 projects, etc. they decided to use the information available through the API of Github (user, repositories, milestones, issues, commits, collaborators, etc.) as an input for applying the information fusion techniques. Once a target “datasource” such as Github was fixed, Jose María Alvarez designed and developed a configurable crawler to extract data and statistics from Github creating a dataset. This dataset was commented in a meeting with Tomasz and he designed a first approach to exploit this information through the information fusion techniques. After that the approach was also discussed with Prof. Kazienko that agreed on following in that research line because it seemed a very interesting possibility to find new hidden relationships in this dataset. Nevertheless the prototype to create the dataset needed some changes in order to create ego-networks around some selected users or type of project. Furthermore the information about events in the social network was also considered as relevant for further experiments but so far the dataset only contain some “static” features. On the other hand and with the aim of boosting this collaborative experiment they agreed to target a deadline in a journal in which the experiments could be submitted. For this purpose, the Information Fusion journal was selected and the 15th of November was fixed as deadline to submit the paper. More specifically, the target schedule would be:

1. Update of the dataset (Jose María Alvarez). First two weeks of October, 2015.
2. Check and validate the dataset. First test using the information fusion techniques. (Tomasz and Adrian). Next two weeks.
3. Update of the dataset (Jose María Alvarez, if necessary)
4. Full execution of the experiments. Next two weeks. (All)
5. Paper writing (All)

After the research stay some conclusions and outcomes are presented in Table 1 and commented below:

- The initial draft of a research proposal is completely required to build on it and finally select a research field. It is also very useful to focus the collaboration when the group has different research lines such as the WR research group. Thus once all people is introduced it is possible to easily detect potential collaborations.
- An introductory presentation to the visitor background and experience it is also required to align the collaborative research to the members of the research group.
- It is completely necessary to flow in the stream of the group. Taking into account the working methodology of the research group, the incoming visitor must participate as another member. In this sense, members of the WR research group helped a lot to introduce Jose María Alvarez in all shared activities in both senses professional and personal.

- From a scientific point of view, the research stay has enabled the possibility of applying the existing information fusion techniques of the WR research group to a new area such as the study of development of software projects. Thus both research lines have been boosted. On the one hand, the WR members has found a new scenario in which their techniques can help to understand why some software projects are successful or not and, on the other hand, Jose María Alvarez has learnt (in a beginner level) new techniques for information exploitation that can be applied to other contexts of software and systems engineering. Furthermore the exchange of ideas and discussions in the use of different tools and APIs have enriched both parties enabling new thinking on how to solve existing problems.
- In order to ease the execution of experiments the WR research group has also offered the possibility of using its super computing infrastructure.
- From a personal point of view, it is completely required that the visitor is open to participate as another member of the group. In this sense, the WR research group (all its members) has demonstrated its willingness to introduce new people in the regular schedule of activities.
- This stay can be considered as a first step of a very promising collaboration in science and research mixing two areas of complex network analysis and development of complex systems.

Table 1 Summary of activities.

Type	Description	Link
Presentation	Background, experience and initial proposal	http://slides.com/josemariaalvarez/mee-research-lines#/
Presentation	Creation of a dataset from Github	http://slides.com/josemariaalvarez/smartgit-ds#/
Prototype	Github crawler to serialize data in different formats (csv and Neo4J)	https://github.com/chemaar/smartgit
Presentation	Research and Science in Spain	http://slides.com/josemariaalvarez/research-science-in-spain#/
Presentation	Presentation at ENIC 2014 conference	http://slides.com/josemariaalvarez/the-skillrank-and-scales#/
Paper	Short contribution to a paper led by Tomasz, Adrian and Prof. Kazienko	Kajdanowicz T., Kazienko P., Popiel A., Alvarez-Rodriguez J.M.: Layer reduction and decision fusion in relational classification for multi-layer social networks. WISE 2014 - The 15th International Conference on Web Information System Engineering, IWCSN 2014 - International Workshop on Computational Social Networks, Thessaloniki, Greece, 12-14 October 2014, Lecture Notes in Computer Science LNCS, Springer, 2014, pp. .
Dataset (on-going)	Adaptation of the Github crawler to create an ego-network	https://github.com/chemaar/smartgit

<p>Paper (on-going)</p>	<p>Target journals and special issues to contribute according to the on-going research</p>	<p>http://www.journals.elsevier.com/information-fusion/call-for-papers/information-fusion-technologies-for-social-big-data/ (15/11/2014)</p> <p>http://listserv.acm.org/scripts/wa-acmplpx.exe?A2=ind1409&L=seworld&F=&S=&P=8860 (30/11/2014)</p> <p>http://www.journals.elsevier.com/information-fusion/call-for-papers/special-issue-on-big-data-deal-with-information-fusion/ (15/02/2015)</p> <p>http://listserv.acm.org/scripts/wa-acmplpx.exe?A2=seworld;89338af5.1409 (01/03/2015)</p>
-----------------------------	--	---

Information referring to the intellectual property

(the generally binding law in this area in the visited country and procedures of patenting);

Description of the procedure of cooperation between universities and SME in visited country

(how it is organized, the sources of funding, the opinions about drawbacks and strengths of existing solution).

REMARK: Apart from this information also a program of the visit and the presentation in electronic version should be given to the project office (please send the report to Monika.rok@pwr.wroc.pl). Please respond to the points 1-5 for outgoing visit and points 1-3 for incoming visit.