Scientific Report

First name / Family name	George Almpanidis
Nationality	Hellenic
Name of the *Host Organisation*	Norwegian University of Science and Technology (NTNU)
First Name / family name of the *Scientific Coordinator*	Heri Ramampiaro
Period of the fellowship	01/10/2012 to 31/09/2013

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

During my ERCIM fellowship period I studied machine learning and information retrieval techniques for mining collaboration filtering in online social networks. In specific, my research involved latent model algebraic and statistical based clustering methods for online social data. The goal was to integrate content and context information in complex networks and perform data mining using high order heterogeneous coclustering.

Since the availability of quality social data that also include content is limited I focused on smaller datasets that correspond to vertical domains, such as the professional sports. As a case study the National Basketball Associaton of USA was selected, since it features certain interesting properties: it is a star-based system, which globally promotes fan participation and user-generated content through general-purpose online social networks, while at the same time it is a well-formed hierarchical organisation. Having an underlying semantic link network of actors (i.e. teams/players etc) alongside a set of broader friendship networks (i.e. Twitter, Facebook, YouTube) designates the need to assume a complex model rather than assuming random undirected graphs commonly employed in the social network analytics industry.

Consequently, I examined extensions of lower order representation structures and mathematical tools to higher order equivalents for dealing with real-world problems and information tasks in complex online social networks, such as ink user clustering, event detection, opinion mining, mood detection. Research was done in multigraphs as ways of modelling multi-relational, multi-aspect social data for domain specific clusters and on tensors as corresponding representation tools. Representing high-dimensional multi-aspect data as tensors allows a richer set of data features to be utilised for information tasks on a smaller scale and in a more focused setting. But as tensor-modelled data can also get large, depending on the complexity of the network model, efficient dimensionality reduction algorithms are still required in order to avoid memory constraints and overfitting problems. Accordingly, I explored efficient tensor clustering techniques and conducted preliminary research on exploiting the sparsity statistical distribution properties of social network representation tensors across different modes.

Nevertheless, recent changes in the cloud data API of general purpose online social services (i.e. Twitter and to a lesser extent YouTube) hindered the extraction of sufficient data mass as time/bandwidth rate limiting for accessing data objects and streams became stricter and some features became end-of-line (e.g. YouTube video responses).

Research was also done in recommender systems for link prediction and link recommendation tasks in online social networks. During the fellowship period I had the opportunity to expand my knowledge on data processing in cloud environments and I was introduced to hadoop and mahout and investigated the parallelism ability of popular machine learning algorithms for the above tasks.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Almpanidis, G. and Ramampiaro, H. "Data mining in heterogeneous social networks using latent variable models", under preparation.
A paper titled "Data mining in heterogeneous social networks using latent variable models" is under preparation and will be submitted as a conference paper in the following months. This paper talks about using tensor and graph -based approaches for employing latent variable model machine learning in the online social network setting. An efficient dimensionality reduction and tensor clustering method is employed for social data concerning the National Basketball Association.

Almpanidis, G. and Ramampiaro, H. "Sentiment analysis on Twitter using graph based methods", under preparation.
A conference paper titled "Sentiment analysis on Twitter using graph based methods" is also pending. This paper is concerned with alleviating the intrinsic noise (due to small size and use of domain specific slang) of the textual Twitter content by semi supervised learning methods that exploit link information (through URL, media, hashtag entities, retweets, etc), social relations, and profile similarity in order to improve the sentiment labelling of tweets.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

I have participated in the second ABCDE seminar on 26 and 37 of October 2012. The event was held in INRIA in Nice, France. On this seminar I participated on a range of non-scientific events presented my research interests, past and anticipated work in the area of web information retrieval, data mining, and recommender systems.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

I visited the VTT Technical Research Centre of Finland in Espoo from 9th to 13th of September 2013. There I met with the researchers of the Data Mining Group, led by Renne Tergujeff and gave a presentation talk of my research work on information retrieval and machine learning. I also gathered information on relevant projects that VTT is working on and expanded my knowledge on research and development activities from a non-academic perspective.

I visited the Swedish institute of Computer Science (SICS) in Kista-Stockholm from 16th to 20th of September 2013. I met with the researchers of the Industrial Applications and Methods Lab (IAM), led by Dr. Daniel Gillblad, and discussed about scientific and research opportunities in the field of data mining and the applicability of big data analysis in the mobile network industry. During my stay I had the opportunity to exchange ideas with the researchers of the group about potential research collaborations. Under the guidance of Prof. Anders Holst I gave a seminar talk with the title "Data mining for online social networks".