Web Graph: Learning Models for Prediction and Evolution Monitoring
LEMPEM
Marie Curie Reintegration Grant
M. VAZIRGIANNIS

1. FINAL PUBLISHABLE SUMMARY REPORT – ADDITIONAL MATERIAL

Dimensionality reduction Algorithms

As web data are voluminous and distributed we have argued that dimensionality reduction techniques should be employed to ensure efficient processing of the data. Data & web mining tasks results are usually improved by reducing the dimensionality of data. This improvement however is achieved harder in the case that data lay on a non linear manifold and are distributed across network nodes. Although numerous algorithms for distributed dimensionality reduction have been proposed, all assume that data reside in a linear space. In order to address the non-linear case, we introduce, a novel distributed non linear dimensionality reduction algorithm, particularly applicable in large scale, structured peer-to-peer networks. Apart from unfolding a non linear manifold, our algorithm is capable of approximate reconstruction of the global dataset at peer level a very attractive feature for distributed data mining problems. The main conclusion is that dimensionality reduction can credibly decrease the complexity and partially increase the quality of data mining in large and distributed data. The obtained results show the suitability and viability of our approach for knowledge discovery in distributed environments as the web inherently is.

Briefly, the *results* of the research performed so far can be summarized in the following:

- Three (3) publications in international peer reviewed journals among of high quality and impact.
- Nine (9) publications in the proceedings of international conferences among which WWW2008, ECML/PKDD 2009, SIAM/SDM 2009, IEEE/PAKDD 2010, IEEE ASONAM 2011)
- Three (2) invited talks in industrial research labs (INRIA/LUCENT - Paris, Telefonica/Barcelona)
- Three invited tutorials in international conferences (French data base conference -2010 BDA/Toulouse, Brazilian database conference SBBD 2010/Belo Horizonte Brasil)
- One lecture in the host institution
- Two invitations for the Google EMEA faculty summit (2008 and 2011)
- A competitive R&D research project where the host institution and the fellow will develop a service based brokering system for the contractor (a tourism related broker industry).

More details are available at the fellows web site: http://www.db-net.aueb.gr/michalis

The fellow has developed *collaborations* with both *industrial* and *research* institutions. With some of them there are already published joint articles. Finally the fellow serves as the programme co-chair of the ECML/PKDD 2011 conference (Athens 2011), served as Data Mining Chair (for the IEEE/ICDE 2011 conference) and was invited to *participate in the Scientific Programme Committee* of more than eleven international conferences and workshops in the area of data mining and machine learning

The potential of the project's results is promising as the researcher has developed collaborations with the web search & telecom industry (i.e. LUCENT ALCATEL, EXALEAD, Google research). The industrial partners are interested in the research results produced in the area of localized and real time web personalization. Also the project results in the area of web services and data management have been

exploited in a local industrial R&D competitive proposal that resulted into a project in the area of web service brokers for car reservations.