Technology Transfer in Computing Systems

D3.7: Individual TTP7 abstract

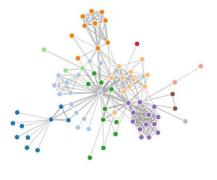|	
Project no.:	609491
Funding scheme:	Collaborative project
Start date of the project:	1st September 2013
Duration:	36 months
Work programme topic:	FP7-ICT-2013-10
Deliverable type:	Report
Deliverable reference number:	ICT-609491 / D3.7
WP and tasks contributing:	WP 3 / all
Due date:	30/09/2015
Actual submission date:	30/09/2015
Responsible Organization:	UPC
Dissemination Level:	Public
Revision:	1.0

TETRACOM D3.7: Scalable Community Detection on the Cloud (SCDC)

David Dominguez (Sparsity Technologies), Arnau Prat (DAMA-UPC), Josep LLuis Larriba-Pey (DAMA-UPC)

During the last years, the analysis of complex networks has become a hot topic both in research and industry. Social, biological, information and collaboration networks are typical targets for such analysis, just to cite a few of them. Among all the tools used to analyze these networks, community detection is one of the most relevant. Communities, also known as clusters, are often referred 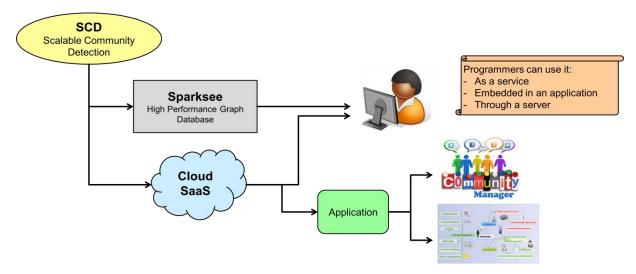to as vertices with a high density of connections among them and with very few connections with the rest of the graph. Community detection provides valuable information about the structural properties of the network, the interactions among the agents of a network or the role the agents play inside the network.

The use of community detection services in the Cloud and as a functionality in Graph Databases is becoming important for companies to discover the bonds between the different entities that they represent through a graph. In addition the use of fast algorithms for such a purpose is important, provided that they have a minimum quality that can be proven not only empirically but also theoretically.

In this TTP, we have focused on productizing the Scalable Community Detection (SCD)[1] algorithm that has demonstrated to be the fastest and most accurate Community Detection technique for Social Networks. With the objective to offer this technology to the public and our customers, we have created two implementations of SCD, as shown in the following figure: one through its implementation in the core of Sparksee, Sparsity's Graph Database (see www.sparsity-technologies.com) and another one through an open service through the web page (see http://www.communitydetectionhub.com/cdh/) which provides a service that can be used interactively through the web page or via a Rest API.

The two implementations are already published, one through the new Sparksee v5.2 and one through the web service provided. The advantages of both implementations are that they are available to the technologists and to the wider public because they can be accessed as technology or as a web service in the cloud.

---

[1] Arnau Prat-Pérez, David Dominguez-Sal, Josep-Lluis Larriba-Pey: High quality, scalable and parallel community detection for large real graphs. WWW 2014: 225-236

The technologies implemented have been widely announced through the following means:

- The publication of our latest Sparksee version v5.2 and the corresponding blog entry for it at www.sparsity-technologies.com and http://sparsity-technologies.com/blog/sparksee-5-2-new-release/ .
- The publication of a web service and the corresponding blog for it at http://sparsity-technologies.com/blog/scalable-community-detection-cloud-scd-made-product/ .
- The publication of posts and news in different social networks:
  - Twitter through our coporate account and a number of tweets through it @SparsityTech.
  - LinkedIn through our corporate account and a number of posts about the Tetracom project: https://www.linkedin.com/company/sparsity-technologies/comments?topic=5935406106779082752&type=U&scope=1298811&stype=C&a=GJ9O
  - Facebook though our corporate account and a number of posts about the Tetracom project: https://www.facebook.com/saved/?collection_token=784105512%3A586254444758776%3A92&cref=35

The business model for the two implementations of SCD is:

- The implementation in Sparksee is provided as a new functionality at no extra cost to our customers.
- The cloud implementation is provided as a freemium service. Graphs of 2M nodes or less are analysed for free, larger graphs we charge on a linear cost model with a ceiling cost.