Finding Interesting Events in Twitter and Communities in Social Networks

Mots clés :
- Directeur de thèse : Mauro SOZIO
- Co-encadrant(s) :
- Unité de recherche : Laboratoire Traitement et Communication de l'Information
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)

When events such as the elections to the European Parliament or the recent crisis in Ukraine unfold, social network users of Google Plus, Facebook or Twitter engage in intense social activity by establishing new friendship links and posting tweets related to the event. One could represent tweets and social networks as graphs (called entity and social graphs, respectively), where nodes represent entities in a tweet (such as ‘Obama’ or ‘Crimea’) or users in a social network, while edges measure the strength of the correlation between the corresponding entities in Twitter or denote friendship links in a social network. We can then observe that such events trigger the quick emergence of dense subgraphs in the entity and social graph, that is, subgraphs containing a relatively large number of edges. In this PhD proposal, we aim at developing efficient algorithms for automatically finding interesting events in Twitter and communities in social networks by finding dense subgraphs in the entity and social graph. This poses several non-trivial challenges, given the sheer size of the graphs described above. Moreover, such graphs are highly dynamic as many friendship links can be established within a relatively short time window and up to 5 thousands tweets per second are posted in Twitter. Therefore, it is crucial to develop algorithms that quickly compute new solutions as the input graph changes over time. In order to cope with the sheer size of the input graph, the PhD candidate will implement these algorithms in modern architectures such as MapReduce, multi-core architectures etc. and evaluate their effectiveness on large real-world graph.

Résumé du projet de recherche (Langue 2)

One of the main technical challenges that the PhD candidate will have to face is to cope with the sheer size of real-world graphs. To this end, the PhD candidate will develop and adapt efficient algorithms into modern architectures such as MapReduce and multi-core architectures. It is also crucial to develop efficient algorithms that compute quickly new solutions as input graph changes over time.

Informations complémentaires (Langue 1)

The PhD candidate will have the chance to work in collaboration with several top research institutes and companies worldwide, such as Google, Yahoo! (USA), Max Planck Institute for Informatics (Germany), and university of Rome "La Sapienza" with which the director of the PhD thesis have strong collaborations.