Computations on Collaboration Spotting Datasets

Richárd Forster

CERN

With a huge amount data getting collected every day, visual analytics is getting more important to help the users understand their data. The CERN developed Collaboration Spotting is a tool providing generic infrastructure to visualize different types of data. To achieve this a multistep process is computing the visual maps of the available data, that involves community detection and visual graph organization. Community detection has become an important operation in numerous graph based applications. It is used to reveal groups that exist within real world networks without imposing prior size or cardinality constraints on the set of communities, thus being able to recognize the different connected brain areas. Despite its potential, the support for parallel computers is rather limited. This is largely because the algorithm is irregular and the underlying heuristics imply a sequential nature. The community detection algorithm used is the Louvain method, that is a multi-phase, iterative heuristic for modularity optimization. It was originally developed by Blondel et al. (2008), the method has become increasingly popular owing to its ability to detect high modularity community partitions in a fast and memory-efficient manner. To parallelize this solution multiple heuristics are used, that were first introduced by Hao Lu et al. (2015). For graph organization the ForceAtlas algorithm is used that was originally developed for the Gephi toolkit. This method is responsible to assign coordinates in a 2D space for every node in such a way that they will not overlap on each other. These methods are implemented in C++ and CUDA and thanks to the GPUs the performance can be increased with a factor of 5.