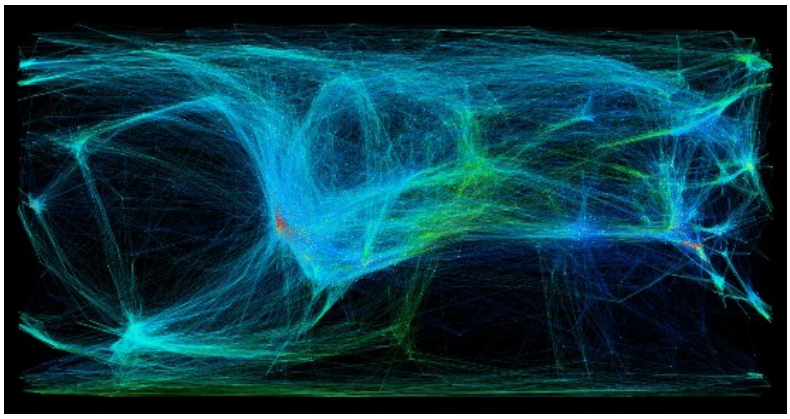
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing Group

Prof. Roger Wattenhofer
http://www.dcg.ethz.ch

SA/DA/MA: Analyzing Large Networks



Many systems can be usefully represented as networks. Examples include the Internet, the World Wide Web, social networks, citation networks, semantic networks, among many others. Understanding the structure of such networks is very important for the development of new and the maintenance of the existing applications. Interesting network features include clusters of strongly related nodes (community structure), shortest paths, flows, and many others.

There exists a wide variety of techniques to analyze such features. However, most of them are not suitable for very large networks. In this thesis we are interested in developing algorithms to analyze large real world networks, comprised of millions of nodes. Therefore, the student must be prepared to work with large databases, external memory algorithms, and possibly parallelization.

As a test bed the student will be able to use rich datasets, which have already been gathered. Examples of available data are *LiveJournal* user and friendship link structure, *Wikipedia* article and topic category graphs, *Last.fm* music playlist and friendship information, and *DBLP* publication and citation records. *LiveJournal* is a social networking and blogging site. *Last.fm* records what people listen to on their PCs, and then presents an array of interesting things based upon their tastes. *DBLP* is a publication database with approximately eight hundred thousands of papers in Computer Science.

In a first step of your thesis your task is to study related work in the area, such as clustering and labeling algorithms for large graphs. Then, the goal is to develop an own or adapt an existing algorithm that constructs a good clustering or a good labeling for large real world graphs in reasonable time for practical applications. The solution will be tested on one or several of the available datasets.

Interested? Please contact us for further details!

Kind of work: Theoretical and practical.

Contact:

1. Olga Goussevskaia: golga@tik.ee.ethz.ch, ETZ G61.4, phone 044 632 70 05
2. Kuhn Michael: kuhnmi@tik.ee.ethz.ch, ETZ G61.4, phone 044 632 77 30
3. Roger Wattenhofer: wattenhofer@tik.ee.ethz.ch, ETZ G63, phone 044 632 63 12