Internship: Characterizing the spatiotemporal structure and dynamics of e-science social networks

Context

Complex networks are expected to capture important characteristics of biological, social, information, and technological systems. A *complex system* consists of many interacting units, whose collective behavior cannot be explained from the behavior of the individual units alone. *Complex networks* are a special type of complex systems that can be represented with graphs whose structure is irregular, complex and dynamically evolving in time. The main focus of present research is moving from the analysis of small networks to that of systems with thousands or millions of nodes, and with a renewed attention to the properties of networks of dynamical units.

Computational grids provide new natural examples of large-scale complex networks emerging from collective behavior. Moreover, computational grids feature multiple levels of interactions. An interesting question is thus whether these networks will exhibit properties similar to those of social networks, or original ones, which would be the specific signature of e-science. An operational question is the creation of generative models appropriate for forecasting future graph structure.

Internship

As a first step towards answering these questions, the internship will characterize the spatiotemporal structure of the graphs created by co-access to files. Historical data from one year (2008-2009) of the GRIF component of EGEE/EGI are available, and data are continuously recorded.

The requested work will cover the following three areas.

- Descriptive static analysis: the goal is to characterize the fixed-time snapshots of the trace.
- Dynamic analysis: characterize the temporal evolution of the snapshots.
- Generative models: in the limits of the internship, the requested work is an evaluation of the adequacy of the models in the literature through qualitative analysis and experimental testing.

For the two first parts, the methodology described in [1] will be followed. The characterization of self-similarity, if any, can exploit the strategy described in [2]. It can be expected that serious scalability issues might appear, offering the opportunity to an algorithmically-oriented part of the internship: the characterization must follow the rate of graph re-shaping, thus falls in the area of streaming in the sense of real-time data mining of massive datasets.

[1] Jurij Leskovec. Dynamics of Large Networks. PhD Thesis.
http://www.cs.cmu.edu/~jure/pubs/thesis/jure-thesis.pdf

[2]Zhou, Jiang, Sornette. Exploring self-similarity of complex cellular networks: the edge-covering method with simulated annealing and log-periodic sampling. *Physica A* 375, 741-752 (2007). http://arxiv.org/abs/cond-mat/0605676

Related projects

The internship

- is proposed by the INRIA TAO project http://tao.lri.fr
- is part of the Grid Observatory project http://grid-observatory.org, related to EGEE and EGI (European Grid Initiative) http://www.eu-egee.org.

Contact

Cecile Germain-Renaud, Professor, Laboratoire de Recherche en Informatique, Université Paris-Sud cecile.germain@lri.fr

Location

Laboratoire de Recherche en Informatique, Université Paris Sud, centre d'Orsay http://www.lri.fr