Project no. 34721

TAGora

Semiotic Dynamics in Online Social Communities

http://www.tagora-project.eu

Sixth Framework Programme (FP6)

Future and Emerging Technologies of the Information Society Technologies (IST-FET Priority)

D4.3 Set of software simulators implementing the best performing modeling schemes and the ensuing control strategies

---

Period covered: from 01/06/2006 to 31/05/2008
Start date of project: June 1st, 2006
Due date of deliverable: July 15th, 2008
Distribution: Public

Date of preparation: 31/05/2008
Duration: 36 months
Actual submission date: June 20th, 2008
Status: Final

Project coordinator: Vittorio Loreto
Project coordinator organisation name: PHYS-SAPIENZA
Lead contractor for these deliverables: PHYS-SAPIENZA

Contents

List of Figures

Chapter 1

Introduction

In the following we briefly describe two simulators developed by the consortium. The first is a tool for network analysis, which has been extended for tag co-occurrence analysis. The second is a simulator of tag streams, based on the modeling activity performed in TAGora. Simulator as available from the TAGora web portal: `http://www.tagora-project.eu/data/#simulators`

Chapter 2

Software simulators

2.1 Epistemic dynamical tagging model

A generative tagging simulator has been developed in order to integrate both the background knowledge and the influence of previous tag assignments. The simulator is based on the recently proposed epistemic dynamiic model for tagging systems (Dellschaft and Staab, 2008), which successfully reproduces characteristic properties of tag streams and even explains effects of the user interface on the tag stream. The simulator reproduces two widely accepted algorithm for producing artificial tag stream, statistically similar to the real datasets found in collaborative tagging communities. It is written in Java and is available online via the TAGora website, where additional documentation can be found. Along with the software, an archive containing all generated tag streams, the software simulator and the technical report is provided. A README file describes how to start the software simulator and which files are contained in the archive with the artificial tag streams. Fig. 2.1 shows a screenshot of the simulator.

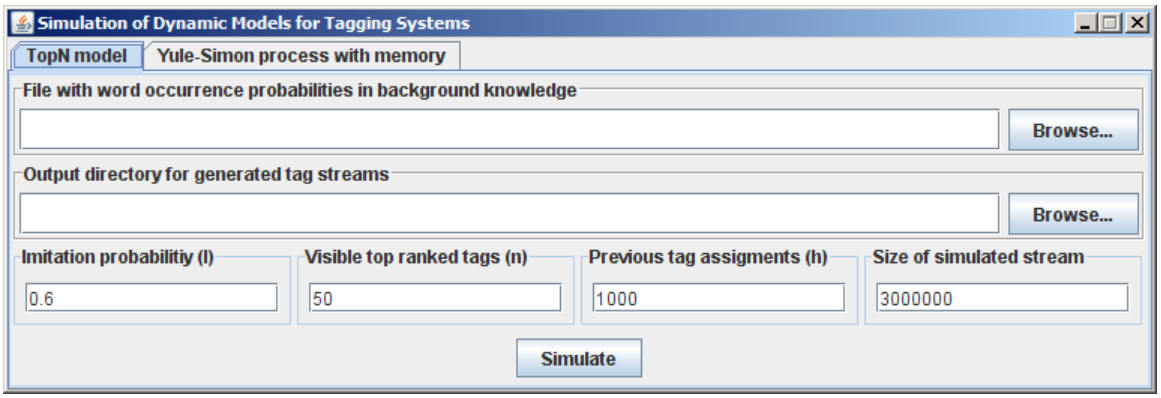

Figure 2.1: Screenshot of the simulator interface.

2.2 NET

NET is a software developed by Vito D.P. Servedio of the Physics Department at University "Sapienza" of Rome, Italy. It is thought to help researchers in the field of complex networks to analyse the statistical properties of complex networks. NET provides tools to both analyze and generate random graphs: it can run multiple realizations of network generation models and perform multiple realization (ensemble) statistics; alternatively, it can perform single and multiple statistics on networks read from files.

NET generates networks according to the “rich-get-richer” model introduced by Barabasi and Albert (Barabási and Albert, 1999), or to the fitness model introduced by Caldarelli et al. (Caldarelli et al., 2002; Servedio et al., 2004).

Current work is devoted to introduce tools more specific to folksonomy. In particular, we implemented the generation of tag co-occurrence networks. The simulator can load a stream of posts and generates the corresponding tag co-occurrence network. This could be used with post streams coming from real datasets, as well as from model simulations. Two examples are implemented directly in NET, i.e. an uncorrelated stream of posts, where tags are chosen accordingly to a generic fat tailed distribution, and post length are determined following a post length distribution, which mimics the observed post length distribution. This kind of streams are useful as a null modem, to compare with those coming from more complex models. For instance, the model recently introduced in (Cattuto et al., 2008) is simulated by NET in order to generate a synthetic tag co-occurrence network.

Below network generation, NET includes several analysis tools for directed, undirected and weighted networks. It performs several tasks including graph reduction according to the minimum betweennes criterium, spectral analysis (requires GSL and Lapack libraries), calculation of the degree distribution and correlations, clustering coefficient, site and edge betweenness, node pair distance and cluster dimension. NET is written in C, and a graphical frontend is available (see Fig. 2.2), requiring the installation of the QT 3.0 libraries. A direct interface with the common XmGrace mathematical plotting software is provided. Networks generated by NET are displayed by the graphical visualization software Graphviz and Grip.

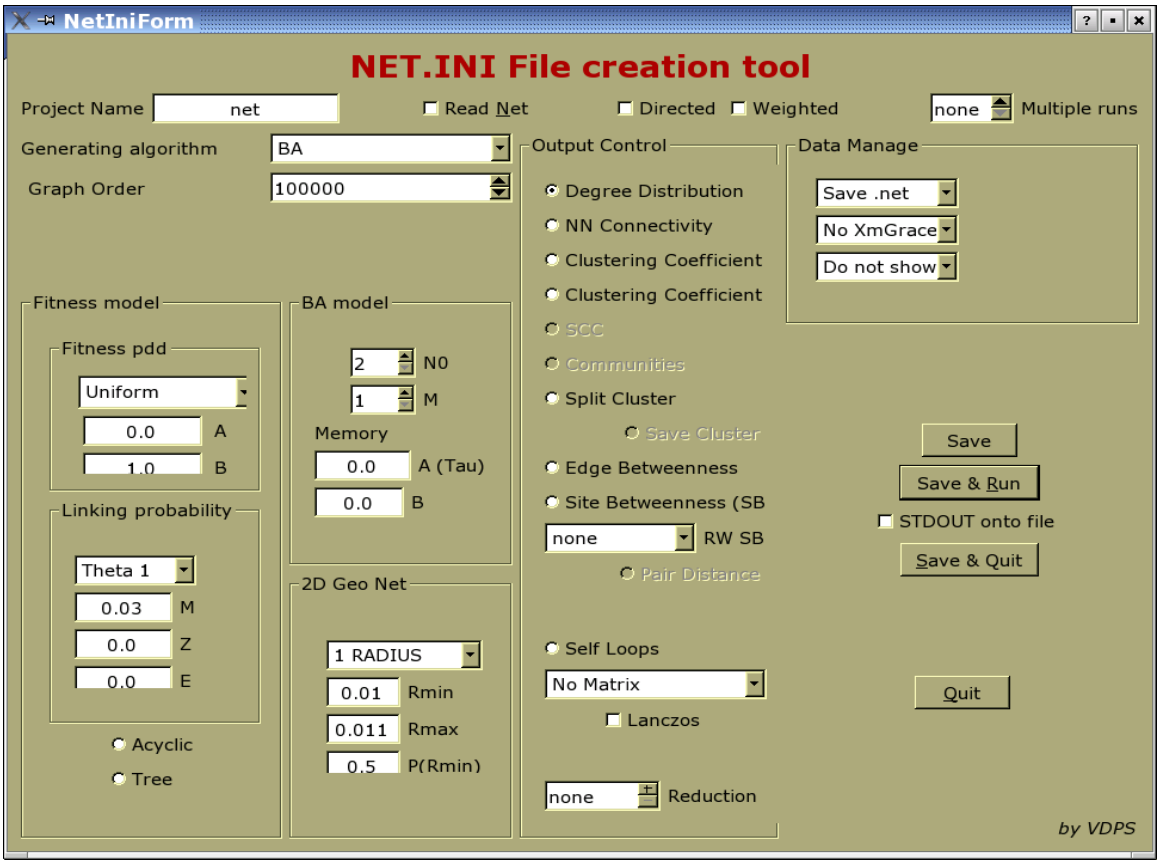

Figure 2.2: Screenshot of the NET interface.

Bibliography

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999. URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/9910332.

G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Mu noz. Scale free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89:258702+, 2002.

C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, and V. Loreto. The collective dynamics of social annotation. 2008.

Klaas Dellschaft and Steffen Staab. An epistemic dynamic model for tagging systems. In *HYPERTEXT 2008, Proceedings of the 19th ACM Conference on Hypertext and Hypermedia*, 2008.

Vito D. P. Servedio, Guido Caldarelli, and Paolo Buttà. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 70 (5), 2004. doi: 10.1103/PhysRevE.70.056126. URL http://dx.doi.org/10.1103/PhysRevE.70.056126.