Tagora – Emergent Semantics in Social Online Communities

Vittorio Loreto

TAGora (http://www.tagora-project.eu) is a project sponsored by the Future and Emerging Technologies program of the European Community (IST-034721) focussing on the semiotic dynamics of online social communities.

1 The vision

The widespread diffusion of access to the Internet is making possible new modalities of interaction between Web users and the information available online. The new vision of the Web regards users not only as producers or consumers of information, but also as architects of the information on the Web, which gets shaped according to criteria closely related to the meaning of information, the semantics of human agents. In this perspective the Web is becoming an infrastructure for "social computing", that is, it allows to coordinate the cognitive abilities of human agents in online communities, and steer the collective user activity towards predefined goals.

An approach to information management that has become wildly popular during 2005 (in a matter of a few months), is *collaborative tagging*. The central idea is that users interested in organizing and sharing a certain kind of resources (digital photographs, web pages, academic papers, and so on), use a web application to associate free-form keywords – called "tags" – with the content they're interested in. Such associations are personal, but globally visible to the user community. At the system level the set of tags, though determined with no explicit coordination, evolves in time and leads towards patterns of terminology usage that are shared by the entire user community. Hence one observes the emergence of a loose categorization system – commonly referred to as *folksonomy* – that can be effectively used to navigate through a large and heterogeneous body of resources. Tags act as a sort of "semantic glue" bringing together resources and users in a time-dependent and truly complex architecture, providing an unexpected bottom-up realization of the vision of the Semantic Web.

Overall, the collaborative character underlying many Web 2.0 applications puts them, very naturally, in the spotlight of complex systems science, since the problem of linking the low-level scale of user behavior with the high-level scale of global applicative goals is a typical problem tackled by the science of complexity: understanding how an observed emergent structure arises from the activity and interaction of many globally uncoordinated agents. The large number of users involved, together with the fact that their activity is occurring on the Web, provide for the first time a unique opportunity to monitor the "microscopic" behavior of users and link it to the emergent properties of Web 2.0 applications.

2 Scientific and Technological Objectives

The project is articulated in four main areas whose activities are strongly intertwined. The initial phase of the project will deal with collecting actual data from existing, live systems and analyzing them with a variety of formal tools, eventually inferring models that are able to capture the essential features of the emergent dynamics, and explain how they might arise from the interactions of single agents. The inferred models of the emergent dynamics will be subsequently used to develop simulations that will allow the formulation of design strategies targeted at attaining a specific global behavior.

Emergent metadata. The initial phase of the project will deal with collecting actual data from existing, live systems. Several online communities are readily accessible over the web: for a selected set of these systems, tools will be developed and deployed to harvest the relevant data, metadata and temporal dynamics, and to store the acquired information in a form amenable for data analysis.

Data analysis of emergent properties. Examining quantitative aspects of folksonomy is a highly important area of research. Our objective is the set up of several protocols of data analysis to be performed on the raw data sets. A data analysis protocol is defined by: (1) indicating a specific quantity / observable / estimator suitable of a quantitative measure on the raw data sets; (2) acquiring the existing software tools, or developing new specific tools, needed to perform the measure; (3) extracting the relevant statistical information characterizing the analyzed data sets.

The aim of the data analysis is to identify and quantify emergent properties of the system in study, i.e. properties that can not be simply inferred from the behavior of the single agent. Beyond suggesting the collection of new or more refined raw data, the results of the data analysis will be used to (i) identify general features common to the different systems in study, (ii) characterize/discriminate the specific features of different systems in study, (iii) orient the modelling phase of the research project (see below), (iv) providing benchmarks to test/improve existing systems or to suggest the creation of new more performing systems.

Modeling and simulations. The objectives of this research area are twofold: *Understanding complexity:* develop