Monitoring Large Scale Complex Information Systems

Stefano Leonardi

Università di Roma ‘‘La Sapienza“

BIGG meeting
28 - 29 November 2006
ETSI - Sophia Antipolis

The DELIS challenge

- Understand the structure, self-organization and dynamics of larg scale information systems
- Provide methods, techniques and tools for controlling and optimizing such systems
- Apply market and biological inspired methods in order to organiz and understand the competition between agents
- Two benchmark applications:

  Decentralized, self-organized Web search engine based on a p2 architecture

  p2p Management platform for telecommunication transport networks

- *Subproject 1; Monitoring, Visualizing, and Analyzing Large Dynamically Evolving Information Systems.*
- *Subproject 2; Structure and Self-Organization of Evolving Networks.*
- *Subproject 3; Large Scale Optimisation.*
- *Subproject 4; Game Theoretic and Organizational Economics Inspired Approaches*
- *Subproject 5; Biology-Inspired Techniques for Organic IT.*
- *Subproject 6; Data Management, Search, and Mining on Internet-scale Dynamical Evolving Peer-to-Peer Networks.*

CTI, CUNI, ETHZ, EPFL, MPII, RWTH Aachen, Telenor, TILS, TU Wroclaw, TUM, UCAM-DAE, UCY, UDRLS, UniBO, UniKarl, UPB, UPC, UPF, RAL, UP

- Decentralized, self-organized, huge size, dynamic.
- The process of network formation reflects the pattern of interacti between agents in the system.
- Well known patterns are inspired by biological and economic paradigms.
- They are in general aimed to optimize a measure of individual benefit while ensuring good functional properties
- Well known paradigms are:
  - preferential attachment with several notions of centrality
  - geometric distance
  - copying and mutation
  - selfishness in route selection

Structural changes

- Emerging of new communication targets, Network faults
- Selection of neighbours, Change in transmission radius
- Selfish local changes: Re-routing on less expensive paths, Copying patterns from neighbours, Antagonize by learning

Behavioural changes vs Structural Changes

- Models information propagation
- Agents (robots, viral/anti-viral) moving in the net
- Agents that meet are involved in a local game

Models of structural changes

- Geometric graphs with controlling parameters (radius, degree)
- Sparse overlay graphs
- Net formation procedure: selection of neighbors
- Evolving networks
- Graphs with prescribed degree sequence Models

Models of Behavioural change

- Models of local interaction between neighbors, e.g. models for infinite lattices used in statistical physics
- Concurrent random walks on graphs. Particles that meet chang their state, appear/disappear

- A large body of work has concentrated on monitoring and modeling the evolution of large scale complex information systems:
  - The Internet topology (Scale free, Power law, Barabasi, Faloutsous,..)
  - p2p systems (navigability, small world, Kleinberg, Watts and Strogatz,)
  - The World Wide Web (Bow-tie structure, Broder et al,)
  - Open content repositories, e.g. Wikipedia, the blogspace
  - Large scale complex social networks

- Structural analysis of complex networks is a fundamental ingredient for:
  - Ranking Web documents (link-analysis)
  - Spam detection
  - Improve efficiency of overlay networks (navigability, load balancing, ….)
  - Building reputation systems
  - Search p2p networks
  - Mining the structure of network communities (identify cluster users, content, concepts)
  - ……

Limitations of current approaches

- Most of the work has concentrated on a static analysis of the structure of complex networks
- The characterization of the structure of these networks has bee based on the study of few samples
  - The Web: Altavista, Webbase, Internet Archive,…
  - Internet topology: AT&T, CAIDA and Oregone Route View
  - Small size samples of p2p networks

We miss a comprehensive and scalable set of methodologies f tracking the temporal evolution of large scale complex systems

- Define a whole set of concepts, methods and algorithms for monitoring the dynamic of complex networks, i.e. alignment, compression of temporal information, incremental crawling, ….
- Goal: Reconstructing the baseline features of the dynamic from accurate and limited sampling
- Distributed and scalable monitoring and storage of data
- Continuous update of the characteristics of system's components

From monitoring to classification

Build a large scale trustable classification systems is a key aspect in several key applications (e.g. distributed Web search, p2p data management for telecommunication services, ….)

- Describe each component by a limited number of features, e.g., topological properties, contents, performances, feedback from transacting peers
- Classify by clustering in a multi-dimensional feature space. Identi components with near common features.
- Relate structure and evolution of a physical, a p2p and a semanti network
- Implement sophisticated distributed algorithms for gathering glob statistics (trust values, performance indexes).
- Distributed secure computation (DHT, majority voting, etc…)

Modeling and predicting the temporal evolution

- Recognize features in an early stage of the life of network components.
- Detect emerging trends vs well established phenomena
- Temporal information is fundamental to assess data quality, reputation and relationships.
- Compute network properties and statistics on a data stream.
- Build automatic on-line classifiers, use techniques from machine learning (decision trees, reinforcement learning, …)

Now that we start to “understand” the structure of complex information systems, we should design mechanisms for controlli their behavior.

- Selfish cooperative and non-cooperative optimization is a way to go: game theory + algorithms + mechanisms + convergence techniques + stability.
- "local implies global" principle should be examined carefully together with ….

  "selecting a few nodes controlled by a central authority to guide t evolution in a desired direction”
- Design incentive mechanisms

Evolutionary aspects of network formation

- Drop assumptions of unbounded rationality and global knowledge
- How a large population agents can compute or learn an equilibria?
- Play a repeated game in which they use (stale) information about current network conditions updated at regular intervals

Construction of cooperative P2P networks

- Protocols that do not require explicit reputation information
- Give incentive mechanisms to limit the influence of malicious pe inspired by market and biological principles.
- Evolutionary approach translated in selecting peers with higher performance that push non cooperative nodes at the border of th network
- Ensure highly cooperative social utility

Grand Challenges

1. Monitor and model the temporal evolution of large scale comple information systems

2. Build a scalable and trustable classification systems in open distributed environments

3. Design mechanisms to control the behavior of complex information systems in the presence of selfish cooperative and non-cooperative components

Thank you!