Project Number 001907

DELIS
Dynamically Evolving, Large-scale Information Systems

Integrated Project
Member of the FET Proactive Initiative Complex Systems

Deliverable D2.1.2

Report on Foundational Research within
Workpackages WP 2.1 and WP 2.2
Start date of the project: January 2004
Duration: 48 months
Project Coordinator: Prof. Dr. math. Friedhelm Meyer auf der Heide
Heinz Nixdorf Institute, University of Paderborn, Germany
Due date of deliverable: December 2005
Actual submission date: January 2006
Dissemination level: PU – public

Work Package 2.1: Phase Transition and Self-organization in Complex Networks
Work Package 2.2: Dynamics and Stability of Networks
Participants: Università Degli Studi di Roma “La Sapienza” (UDRLS), Italy
Computer Technology Institute, Patras (CTI), Greece
Universitat Politècnica de Catalunya (UPC), Spain
Charles University, Prague (CUNI), Czech Republic
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Université de Paris-Sud XI (UPS), France

Author of deliverable: Lefteris Kirousis (kirousis@ceid.upatras.gr)
1 Introduction

It is by now a widely accepted fact that Random Graph Theory is undergoing an important shift, both with respect to its techniques, as well as with respect to the questions posed. In the classical theory of Erdős and Rényi [10, 11], where each pair of vertices in the graph is connected by an edge with a given probability, the graphs obtained had no hierarchy and their vertices were essentially undistinguishable. Also the issues at stake were parameters like colorability, connectivity, diameter etc., but certainly did not concern dynamic properties like robustness or resilience. On the contrary, graphs that are of interest to modern research, either natural (sociological or biological) or artificially constructed, have a rich structure, are complex, and can be classified into an assortment of different types. Additionally new features like community structure and physical realizability acquired importance. Last but perhaps most important, this different approach necessitated less rigorous and more flexible techniques borrowed from physics or the biological sciences.

This New Random Graph Theory was initiated at about the turn of the last century by the pioneering work of e.g., Watts and Strogatz [25, 26], and Barabási and Albert [1]. Random graphs thus defined have properties, like power law distribution of the degrees, or freedom from scale, and small-world properties, i.e. small diameter, that reflect the complex interaction patterns of the complex systems they are intended to capture.

Also, this novel paradigm lead to an assortment of distinct graph models, each suitable for particular point of view of analysis or tailored for a particular application. Despite this pot pourri of different models however, a unifying thread for research is emerging where the fundamental questions seem to include:

- How local and random or individually motivated connections between nodes affect the global properties of the resulting network?
- What are the global properties that are important from the point of view of large and complex networks, especially information networks?
- How abrupt changes in global properties, i.e. phase transition phenomena, result as a consequence of gradual changes in the parameters that characterize the policy of local connections?

Foundational research within SP2 aims at answering questions as above. We have already made important advances towards this goal, having developed and studied new models, proved important results about them, defined novel notions of importance for understanding complex networks and pin-pointed cases of sudden change on the macroscopic level brought up by smooth microscopic alterations.

We outline below the fundamentals of this research.

2 Geographic models and modes of communication

The unifying theme in this area of research the global communication structure of a network comprised of individual nodes randomly scattered in 2d or 3d and establishing local communication links based on criteria like closeness, connectivity degree etc.

2.1 Connectivity between agents scattered on a mesh or graph

The Barcelona site [22] studied connectivity issues related to agents scattered on a mesh or more generally an arbitrary graph. Imagine that we have a set of w agents (for example robots), with cardinal movement (N/S/W/E), sampling several city levels (ambient noise, carbon monoxide, ozone, humidity, etc.). The robots move around the city. At regular steps of time, they stop, take their
samplings, broadcast to the others, and randomly continue or change direction. The agents communicate with radio frequency, using a simple gossip protocol. One of the agents has a secondary station with sufficient power to relay the information of all agents to a base station. The agents are deployed uniformly at random through the intersections of the street. It is assumed that the streets of the city are modeled in a grid-like pattern. The first question studied is the threshold for connectivity between the agents; i.e. given the size of the grid and the maximum distance of broadcast (as function of the grid size), the minimum number of agents needed to insure connectivity is estimated. As a second step, the way that connectivity evolves in a dynamic setting is examined. Finally the same questions are raised when the underlying path is not a grid but an arbitrary path.

2.2 Connectivity between nodes establishing communication based on distance

Caldarelli and Servedio (Rome, see also D2.2.1) analyzed multi-hop ad-hoc networks in the frame of geometric random networks. An ad-hoc network may be defined as an undirected network by considering two nodes as adjacent whenever their effective transmitting areas overlap. It is shown with simple probabilistic reasonings that a so constructed network is connected if the effective transmitting area scales as \((\log N - \log \log N)/N\) with the number \(N\) of nodes, irrespective of the area shape. It was found that in case of circular transmitting areas, the number of nearest neighbors needed for connectivity scales as \(4 \log N\). The undirected ad-hoc network was generalized to the case of effective transmitting radii with value extracted with a probability distribution density and the main network statistical properties were calculated. Ad-hoc directed networks with link pointing towards those nodes falling inside the transmission area of another node were examined. Statistical properties of the network in case of the non trivial case of transmission radii obeying a given random distribution were calculated.

Another important aspect of local communication based on distance is when in addition to distance also the relative orientation of the two agents is taken into account. Pioneering work in this area was performed in Barcelona in collaboration with Patras. This work captures the global net structure emerging from local broadcasting of nodes, within a certain communication radius and within a certain angle of broadcasting. They have phase transition phenomena with respect to their connectivity and chromatic number [7].

Considerations like the ones outlined in the above subsections lead to important approaches to problems related to device discovery and scatternet formation in multi-hop Bluetooth networks [8].

3 Small-world, clustering and weighted links

These are two properties shared by many complex information networks, however they often seem to exclude each other at least in theoretically constructed models. Researchers in Lausanne and Rome devoted considerable part of their effort in this issue:

3.1 Physical realizability of Small-World networks

One feature shared by most real networks, i.e. the small-world (SW) [26] property, involving a high degree of interconnectedness both at local and global level. That is, for every node, most nodes close to it should also be close to each other (high clustering coefficient), and every pair of nodes is separated, on the average by a number of links that grows at most logarithmically with the network size (small diameter). The SW phenomenon is relevant for communication and technological systems such as the Internet [21] and electronic circuits [12]. SW properties are of great relevance for communication systems: SW networks are particularly efficient for message passing protocols that rely only on the local (i.e. nearest neighbors) knowledge of the network available to each node [17].
Yet, despite their relevance and widespread occurrence, there is still no clear understanding of the minimal requirements necessary to realize SW networks in physical space.

The simple rule to build a SW network is to take a lattice (nodes ordered on a grid) and to supplement it with a few long range connections (shortcuts). Therefore, the relevant parameters to build a SW network are: the number of nodes (which, for fixed density, determines the linear size \( L \) of the system, \( N = L^D \)); the physical dimensionality \( D \) (one-, two- or three-dimensional lattices being the most interesting cases); the shortcut length distribution \( q(l) \).

It is shown both analytically and numerically [23] that SW networks can be realized only if shortcut lengths \( l \) are taken from a power-law decaying distribution, \( P(l) \sim l^{-\alpha} \) for large values of \( l \), with \( 0 < \alpha < D + 1 \). The main argument to obtain such result is that to cross-over from a simple Euclidean world to SW behavior a critical number \( N^* = p^*N \) of nodes must be supplemented with shortcuts (the SW behavior being then fully realized for \( p > p^* \)), and \( N^* \) must be large enough to guarantee the presence of at least a finite number of shortcuts. Stated otherwise,

\[
p^* L^D \int_{(1-\epsilon) L}^{L} q(l) dl = O(1)
\]

and consequently \( p^* \sim L^{\alpha-D-1} \), which implies that only if \( \alpha < D + 1 \) the number of shortcuts is subextensive (the superextensive case being trivially small-world, with the complete graph as a limit).

Interestingly, the consequences of this result is that \( D \geq 2 \) it is possible to realize SW networks using a finite average shortcut length per node (for \( 2 < \alpha \)) making SW networks physically realizable.

### 3.1.1 Evaluation of cluster robustness

A characteristic of the topology of complex networks is their community structure. In real-world networks, it is common to have small sets of nodes highly connected with each other but with only a few connections with the rest of the network. Finding clusters of a network is a crucial point in order to understand its internal structure. A large amount of clustering algorithms have been recently developed, each of them attempting to find reasonably good partitions [15, 4]. In most cases these algorithms identify non-overlapping clusters, assigning each node to a given cluster (“hard-clustering”). However the resulting clustering is sometimes questionable, especially for nodes that “lie on the border” between two clusters. We define such nodes as unstable nodes.

Defining and identifying unstable nodes is closely related to the problem of evaluating the robustness of the clustering. Recently several non-deterministic clustering algorithms have been developed [24, 16]: since every application of the algorithm is going to provide slightly different clustering partitions, the robust cores of clusters will be identified because they are identified by every realization, whereas the rest of the clusters, that can change from one algorithm run to another, can be classified as unstable nodes. In this context we introduced a way to have stochastic outputs from deterministic algorithms, that usually run much faster than stochastic ones [14].

The basic idea is to add a small noise on the weights of edges: an edge is not just present or absent, but it also carries a variable which represents its weight. At every application of the deterministic clustering algorithm, we re-assign the noise over the network. Consequently, after a number of clustering realizations, nodes that have been consistently assigned to the same cluster are defined as stable nodes, whereas nodes that often switch from one cluster to the other in different noise realizations are unstable. With this method we have identified ambiguous words in a French synonym dictionary, helping its editors to prune the dictionary. Applied to physical networks of molecular conformational changes, the method has helped identify stable conformations and transition states, i.e. conformations that are on saddle points between minima of the potential energy.

The method that we propose is not tailored to a specific clustering method, but rather it can be applied with any deterministic clustering algorithm able to deal with weighted networks.
3.2 Weighted links

In Paris [3, 2, 5] researchers studied the following issues:

- Models for weighted networks, in particular study of the combined effects of weights and spatial constraints.
- Analysis of the structure of networks through a rate equation approach (collaboration with UPC).
- Study of the percolation in networks with inhomogeneous links.

4 Preferential attachment models

One of the most prevalent models for complex networks is, of course, the Barabási and Albert model of preferential attachment. Several issues pertaining to this model were studied within the framework of SP2.

Nikoloski, Deo and Kucera (CUNI) defined the degree correlation—correlation of the degrees in a pair of adjacent nodes—for a random graph process in [19]. They showed by determining asymptotically the joint probability distribution for node-degrees, $d$ and $d'$, of adjacent nodes, that the model of Barabási and Albert does not generate degree-correlation.

As an application of the above ideas, Nikoloski, Deo and Kucera (CUNI) studied in [20] the worm propagation problem as a random process that creates a complex system of interacting agents (worm copies) over a scale-free graph, representing real-world networks. Additionally they developed a pair-approximation (correlation) model of worm propagation that employs the salient network characteristics order, size, degree distribution, and transitivity. They confirmed the validity of this model by comparing its numeric solution to results from our individual-based simulation. Eventually they demonstrated that the network structure has considerable impact on the propagation dynamics when the worm uses local propagation strategies.

For the model of preferential attachment, Georgiadis and Kirousis (CTI) studied deliberate attacks on the infrastructure of large scale-free networks in [13]. These attacks are based on the importance of individual vertices in the network and the concept of centrality. Some measures of centrality however, as is betweenness, have disadvantages that do not facilitate the research in this area. They showed that with the aid of scale-free network characteristics such as the clustering coefficient we can get results that balance the current centrality measures, but also gain insight in the workings of these networks.

5 Other models

Besides the geographic and preferential attachment models, other models, suitable for particular purposes, were studied.

5.1 Intersection graphs

Intersection graphs arise when nodes sharing some common attributes, chosen from a totality of possible attributes, are connected together. From another viewpoint, intersection graphs capture the global, emerging net structure when local nodes choose at random certain “channels” to use for communication. Two nodes can communicate directly if they both have chosen at least one common channel. They have phase transition phenomena with respect to the existence of very long (Hamiltonian) Paths and Cycles [9].
5.2 Graphs with a given degree sequence

One approach to generate graphs with predefined degree distributions, similar to those found in real systems, had been proposed by Newman et al. [18]. In this approach, instead of modeling a random graph with an edge generating process with a specified attachment rule at each step (generative models), one assumes instead that the graph is uniformly random conditional on its degree sequence. Certainly, this static approach is more restricted in comparison to the generative approach. However, it is often more suitable for the analysis of graph properties like clustering, distribution of component sizes, resilience, thresholds etc.

Here [6], we restrict our study to an important subclass of random graphs with a given degree sequence: that of random regular graphs, i.e., graphs where all vertices have the same degree. This work concerns a much studied graph parameter, that of the chromatic number of a graph. The chromatic number is defined to be the smallest number of colors needed to color the vertices of a graph, given that no two adjacent vertices get the same color. In network terminology, the chromatic number is the smallest number of classes that one can partition the nodes of a network, so that each class comprises only of mutually isolated nodes.

6 Conclusions

Prima facie one may think that the foundational work described above comprises only of unrelated bits and pieces. However, central pieces of the jigsaw puzzle are already in place and the total picture is emerging. In particular:

- We now understand quite well connectivity or percolation when randomly placed nodes establish communication links based on geographic criteria. These criteria could be quite diverse, including distance, broadcast area or number of nearest neighbors. Also the initial placement of nodes could be on 2d, or a mesh or even on an arbitrary underlying graph.

- We have identified and studied various phase transition phenomena having to do with long paths and chromatic number. Also we further developed rigorous techniques for establishing sharp transition intervals.

- We have advanced the study of small-world graphs leading to the better understanding of important graph parameters like clustering and connectivity.

All in all, we believe that foundational research within the framework of SP2 has already placed a distinctive mark on the effort to develop a novel Random Graph Theory that better suits the theoretical needs for the understanding of large and complex communication networks.

References


