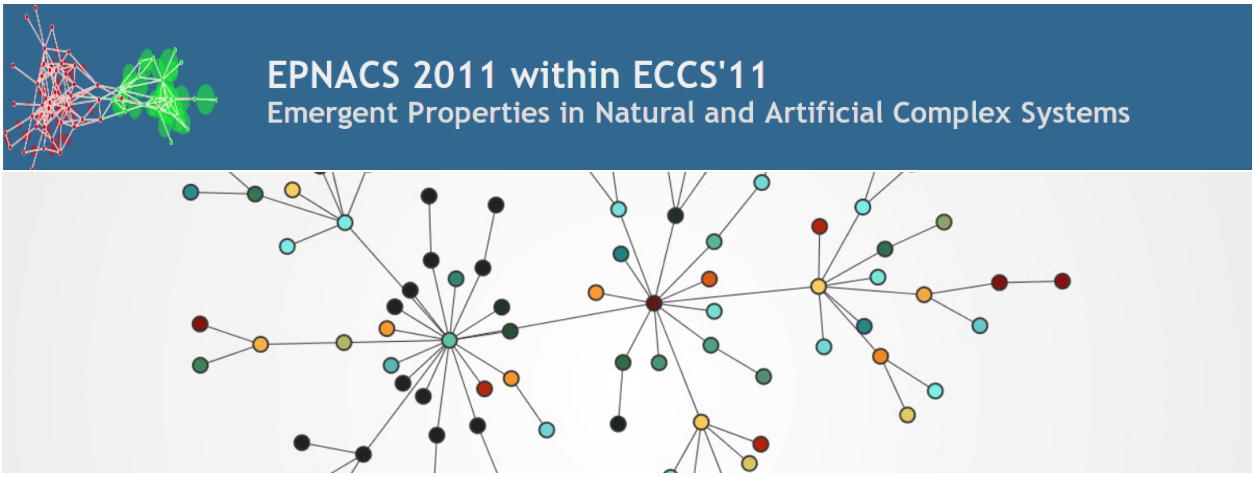
Proceedings of EPNACS 2011, Emergent Properties for Natural and Artificial Complex Systems, a satellite meeeting within ECCS'11

Moulay-Ahmed Aziz-Alaoui, Arnaud Banos, Cyrille Bertelle, Gérard H.E. Duchamp



EPNACS 2011

Emergent Properties for Natural and Artificial Complex Systems

A Satellite Meeting of

ECCS'11

European Conference on Complex Systems

September 15th, 2011

Vienna, Austria

Editors

M. Aziz-Alaoui
Arnaud Banos
Cyrille Bertelle
Gérard H.E. Duchamp

EPNACS 2011

Emergent Properties for Natural and Artificial Complex Systems

A Satellite Meeting of

ECCS'11

European Conference on Complex Systems

September 15th, 2011

Vienna, Austria

Editors

M. Aziz-Alaoui, Arnaud Banos, Cyrille Bertelle and Gérard H.E. Duchamp

Program Committee:

- Abdulrab, Habib, ISCN & INSA Rouen, France
- Ayesh, Aladdin, ISCN & De Montfort University, U.K.
- Aziz-Alaoui, M.A., ISCN & University of Le Havre, France
- Babkin, Edward, ISCN & Higher School of Economics (Nihzny Novgorod), Russia
- Badariotti Dominique, LIVe, Strasbourg, France
- Banos Arnaud, ISC-PIF & Géographie-cités UMR 8504, Paris, France
- Bertelle, Cyrille, ISCN & University of Le Havre, France
- Chen G., City University of Hong-Kong, China
- Duchamp, Gérard H.E., ISCN & Paris XIII University, France
- Elissalde B., ISCN & University of Rouen, France
- Ghnemat, Rawan, ISCN & German-Jordanian University, Amman, Jordan
- Guinand, Frédéric, ISCN & University of Le Havre, France
- Gulyas, Laszlo, ISCN & Eotvos University, Budapest, Hungary
- Iantovics, Barna L., ISCN & Petru Maior University of Tg. Mures, Romania
- Latapi Matthieu, LIP6, Paris France
- Li B.-L., California University, USA
- Lozi R., Nice University, France
- Lu Jinhu, Chinese Academy of Sciences, China
- Lucchini Françoise, ISCN & University of Rouen, France
- Mélançon Guy, Labri, Bordeaux France
- Olivier, Damien, ISCN & University of Le Havre, France
- Perna Andréa, ISC-PIF, Paris France
- Provitolo Damienne, Géosciences Azuz, Sophia Antipolis France
- Rozenblat Céline, University of Lausanne, Switzerland
- Rouquier Jean-Baptiste, ISC-PIF, Paris France
- Sanjuan Miguel A.F., University Rey Juan Carlos, Madrid, Spain

General Chairs:

- **M. Aziz-Alaoui**, ISCN & LMAH, University of Le Havre
  25 Rue Ph. Lebon, BP 540, 76058 Le Havre Cedex, France.
  Phone: +33 235 217 213
  email: aziz.alaoui@univ-lehavre.fr
- **Arnaud Banos**, ISC-PIF & Géographie-cités, CNRS
  13 rue du Four, 75006 Paris, France.
  Phone: +33 140 464 000
  email: arnaud.banos@parisgeo.cnrs.fr
- **Cyrille Bertelle**, ISCN & LITIS, University of Le Havre
  25 Rue Ph. Lebon, BP 540, 76058 Le Havre Cedex, France.
  Phone: +33 235 217 214
  email: cyrille.bertelle@univ-lehavre.fr
- **Gérard H.E. Duchamp**, ISCN & LIPN, University of Paris 13
  99 avenue Jean-Baptiste Clément, 93400 Villetaneuse, France.
  Phone: +33 149 402 862
  email: ghed@lipn.univ-paris13.fr

This Satellite Meeting is organized under the auspices of:

- ISCN, Institute of Complex Systems in Normandy, France
  http://isc-n.fr
- ISC-PIF, Complex Systems Institute, Paris-île de France, France
  http://iscpif.fr
- S4, European Research Group "Spatial Simulation for Social Systems"
  http://s4.parisgeo.cnrs.fr/
- RNSC, French Network of Complex Systems
  http://rnsc.fr

The organizers thank for their scientific and financial supports:

- LITIS, Laboratory of Computer Science, Information Processing and Systems, University of Le Havre, University of Rouen & Insa Rouen, France
  http;//www.litislab.eu
- LMAH, Laboratory of Applied Mathematics, University of Le Havre, France
  http://lmah.univ-lehavre.fr
- LIPN, Laboratory of Computer Science, Paris Nord, France
  http://www-lipn.univ-paris13.fr/
- Géographie-cités, Paris, France
  http://www.parisgeo.cnrs.fr/
- University of Le Havre, France
  http://www.univ-lehavre.fr
- University of Paris 13, France
  http://www.univ-paris13.fr
- CNRS, France
  http://www.cnrs.fr
- MESR, Ministery of High Education and Research, France
  http://www.enseignementsup-recherche.gouv.fr/
- Région Haute-Normandie, France
  http://www.hautenormandie.fr/
- Région Paris-île de France
  http://www.iledefrance.fr/

Preface

The aim of this 4th event from EPNACS series was to study emergent properties arising in natural and artificial dynamic networks. The meeting was concerned with multidisciplinary approaches aimed at exploring, visualizing, modeling and simulating large scale dynamic networks, in order to detect their emergent communities, to analyse self-organizing processes and to characterize their evolution. Linking the morpho-dynamics of complex networks with the dynamic processes they convey was, for example, one of the target questions this meeting addressed. Another key issue concerned the identification of general properties of these dynamic complex networks in various natural and artificial systems (urban networks, ecosystems, neuronal networks, etc).

Le Havre & Paris
July 29th 2011

M. Aziz-Alaoui
Arnaud Banos
Cyrille Bertelle
Gérard H.E. Duchamp

Contents

Part I

Dynamical Systems and Chaos

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

BURST SYNCHRONIZATION OF COUPLED OSCILLATORS: TOWARDS UNDERSTANDING THE INFLUENCE OF THE NETWORK TOPOLOGY

Nathalie Corson, Stefan Balev, M. A. Aziz-Alaoui *

Abstract.
This paper addresses the question of burst synchronization in networks of chemically coupled Hindmarsh-Rose neurons. After a brief description of the model and of an algorithm of numerical detection of burst synchronization, we present numerical experiments designed to give an insight on the influence of the network topology on the minimal coupling strength needed to obtain burst synchronization in the network. Two topological characteristics are studied: the network diameter and the in-degrees of the nodes. Our numerical simulations show that when the diameter grows, the network becomes more difficult to synchronize, while networks with bigger in-degrees of the nodes synchronize more easily.

Keywords. Neuron models, Dynamical systems, Synchronization, Complex networks

1 Introduction

Synchronization of two dynamical systems generally means that one system somehow follows the motion of another. A lot of research has been carried out and, as a result, showed that even chaotic systems could synchronize when they are coupled. Many researchers have discussed the theory, the design or applications of synchronized motion in coupled chaotic systems [1, 9, 12]. There are different synchronization regimes. Oscillators firing bursts exhibit *burst synchronization*, when they all fire the same number of bursts starting at the same moment.

The conditions on the network to make burst synchronization appear are weaker than the ones needed to observe a complete synchronization phenomena [2, 6, 10, 11, 13]. In particular, complete synchronization in nonlinearly coupled networks of Hindmarsh-Rose neurons necessitates equal in-degrees of all network nodes, which is biologically unrealistic. That is why in this paper we are interested in burst synchronization.

We focus on networks composed of Hindmarsh-Rose neuon models, given by (1).

$$(HR)\begin{cases} \dot{x} &= y + ax^2 - x^3 - z + I \\ \dot{y} &= 1 - dx^2 - y \\ \dot{z} &= \epsilon(b(x - c_x) - z) \end{cases} \quad (1)$$

*Nathalie Corson and M. A. Aziz-Alaoui are with LMAH, University of Le Havre, France. Stefan Balev is with LITIS, University of Le Havre, France. E-mails: nathalie.corson@univ-lehavre.fr, aziz.alaoui@univ-lehavre.fr, stefan.balev@univ-lehavre.fr

Parameters a, b and d are experimentally determined, c_x is the equilibrium x-coordinate of the two-dimensional system given by the first two equations of (1) when $I = 0$ and $z = 0$ and parameter I corresponds to the applied current. Finally, parameter ϵ represents the ratio of time scales between fast and slow fluxes across the membrane of a neuron. This HR neuron model can exhibit most of biological neuron behavior, such as *spiking* or *bursting*. With appropriate parameter settings, the HR model exhibits periodic behavior characterized by fast periods of spiking called bursts, followed by slow quiescent inter-burst periods, as shown in Fig. 1.

Hereafter, for all numerical experiments, we use HR system with the following coordinate changes, see [3], $y = 1-y$, $z = 1+I+z$, $d = a+\alpha$, $c = -1-I-bx_c$. Applying this transformation, we obtain,

$$\begin{cases} \dot{x} &= ax^2 - x^3 - y - z \\ \dot{y} &= (a + \alpha)x^2 - y \\ \dot{z} &= \epsilon(bx + c - z) \end{cases} \quad (2)$$

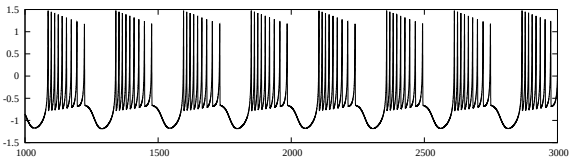


Figure 1: Time series (t, x) of (2) when parameters are fixed as in (5). For this set of parameters values, a HR system exhibit a periodic bursting behavior.

Let us consider a network composed by n HR neurons. These neurons are coupled by their first variable x_i with a coupling function modeling chemical synapses. A model of this network is given by

$$\begin{cases} \dot{x_i} &= ax_i^2 - x_i^3 + y_i - z_i - \sum_{j=1}^{n} c_{ji} h(x_i, x_j) \\ \dot{y_i} &= (a + \alpha)x_i^2 - y_i \\ \dot{z_i} &= \epsilon(bx_i + c - z_i) \end{cases} \quad (3)$$

for $i = 1, \ldots, n$, where h is the coupling function and $\{c_{ij}\}$ is the network adjacency matrix. When the neurons are

chemically coupled, the coupling function h is given by [3] and reads as

$$h(x_i, x_j) = g_{\text{syn}} \frac{(x_i - V)}{1 + \exp(-\lambda(x_j - \Theta))} \tag{4}$$

where g_{syn} is the coupling strength, Θ is the threshold reached by every action potential for a neuron. Parameter V is the reversal potential and must be larger than $x_i(t)$ for all i and all t since synapses are supposed excitatory. The parameters are fixed as follows throughout this paper,

$$a = 2.8,\ \alpha = 1.6,\ c = 5,\ b = 9,\ \epsilon = 0.001 \tag{5}$$

$$V = 2,\ \lambda = 10,\ \Theta = -0.25 \tag{6}$$

These parameter values are commonly used in the literature (see for example [3].

The rest of this paper is organized as follows. In section 2, we recall the algorithm we developed to detect burst synchronization phenomena in network of coupled oscillators, and in section 3 we apply this algorithm to networks of different topologies and sizes.

2 Algorithm of burst synchronization detection

We consider that a network of coupled neurons presents a burst synchronization behavior if the neurons fire bursts starting all at the same time. Unlike complete synchronization, burst synchronization is not easy to detect numerically. Even the distinction between fast and slow periods could be difficult, since the dynamics of single neurons could change in unpredictable way when the coupling force is slightly modified.

In this section we recall the main steps of a general algorithm of burst synchronization detection in networks of coupled oscillators [4]. Its application is not restricted to HR neurons, it can be used to detect burst synchronization in networks composed of any oscillators displaying burst behavior.

Our algorithm can be decomposed in four main steps. In order to detect burst synchronization, bursts of different neurons must be matched. To do this, one needs to determine the start time of each burst, and before detecting bursts, spikes must be detected first.

The first step of our algorithm is the detection of spikes. For our needs it suffices to find the local maxima of $x_i(t)$ for each neuron i. Thus each spike is associated to the time when the corresponding local maximum occurs. Fig. 2 shows an example of time pattern of bursts in a network of 5 neurons. Each spike is represented as a point on the horizontal line corresponding to the neuron in which the spike occurs.

Once all spikes localized, we need to determine the first spike of each burst. Since each burst is preceded by a quiescent period, the idea is to consider the inter-spike distances. When the distance between two consecutive spikes is large enough, the second spike is considered as the first of a new burst.

For certain coupling force values and network topologies, the behavior of individual neurons can change from bursting to spiking. An indicator allowing to distinguish between these two behaviors is the ratio between the smallest inter-burst distance and the largest inter-spike distance. This ratio close to one indicates spiking behavior. With our parameter settings, this ratio is about 4.2 for a single non-coupled neuron.

The next step of our algorithm is to match the bursts fired by different neurons. We define the distance between two bursts as the absolute value of the difference of their starting times determined at the previous step.

A necessary condition for burst synchronisation is that almost all bursts belong to n-tuples of matching bursts. This condition is measured by the ratio of the number of bursts belonging to matching n-tuples and the total number of bursts in the network, over a long time period.

Another condition for burst synchronization is that all the bursts within a matching group start in a small time interval. We measure the largest distance between two bursts in each n-tuple of matching bursts and we take the mean of this distance over all n-tuples in the observed period.

To recapitulate, there are three conditions for burst synchronization:

- The oscillators must have bursting behavior.
- The ratio between the number of matching bursts and the total number of bursts must be close to one.
- The average span of the bursts within each matching group must be below a given threshold.

3 Numerical Simulations

In this section, the algorithm presented in the previous section is applied to different kinds of networks. The numerical results are obtained using a code implementing Runge-Kutta 4 integration method and developed in CUDA C [8]. The code was executed on NVIDIA Tesla C2050 GPUs. This implementation allows us to reduce the running times two orders of magnitude compared to the equivalent CPU implementation and thus to simulate networks containing thousands of nodes in reasonable time. The post-processing (including the algorithm from the previous section) is implemented in Java. GraphStream library [5] is used to generate and to manipulate networks.

Our numerical experiments aim to answer the following question : What is the coupling strength g_{syn} (see (4)) needed to synchronize the bursts of all the neurons of a given network? The answer is not obvious in the general case and that is why we start by studying particular

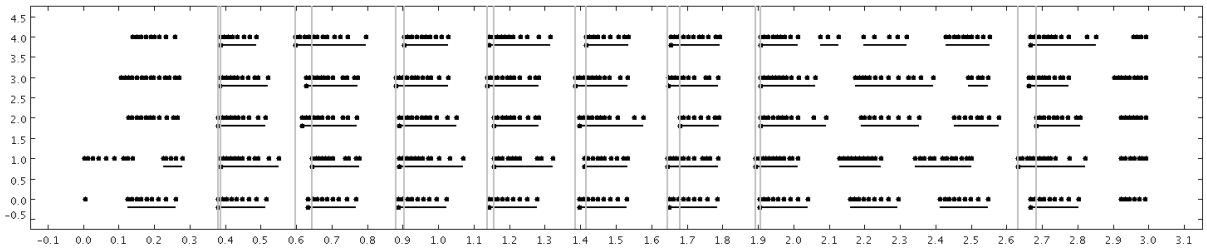


Figure 2: Illustration of the burst synchronization detection algorithm. The four steps of the algorithm are represented on this figure.

network topologies. These particular cases are chosen to give us an insight on the influence of different network parameters on the synchronization threshold.

3.1 Network topology

A necessary condition for synchronization of two nodes of a network is that either one of them must be influenced by the other one or both of them must be influenced by a third node. At network level this implies the existence of at least one "root" node from which all nodes can be reached.

We impose a second condition, which is the absence of cycles in the network. The reason for this restriction is that cycles could significantly modify the individual neuron behavior. In the presence of cycles, the bursting phenomenon could even disappear for certain coupling strength values. To illustrate this fact, let us consider the simplest cycle case, two neurons with bidirectional coupling. Fig. 3 shows that when the coupling strength grows the bursting is progressively transformed in spiking. From the moment when complete synchronization is observed, bursting behavior comes back but in different form and disappears again for very big coupling forces. In the presence of longer cycles, the individual behavior could be even more perturbed. Our experiments show that in acyclic networks, the bursting motion is more stable.

It is easy to see that in acyclic networks the root node is unique. Our first observation was that the coupling strength needed to synchronize a given node to the root node depends on the distance between them. In other words, networks with smaller diameter require less coupling strength to synchronize. To study the influence of the network diameter, we use networks constructed by levels. Level 0 contains the root node. Nodes of level l receive signal only from nodes of level $l-1$. Thus the distance between the root and all the nodes of level l is exactly l. An example of such a network is given in Fig. 4.

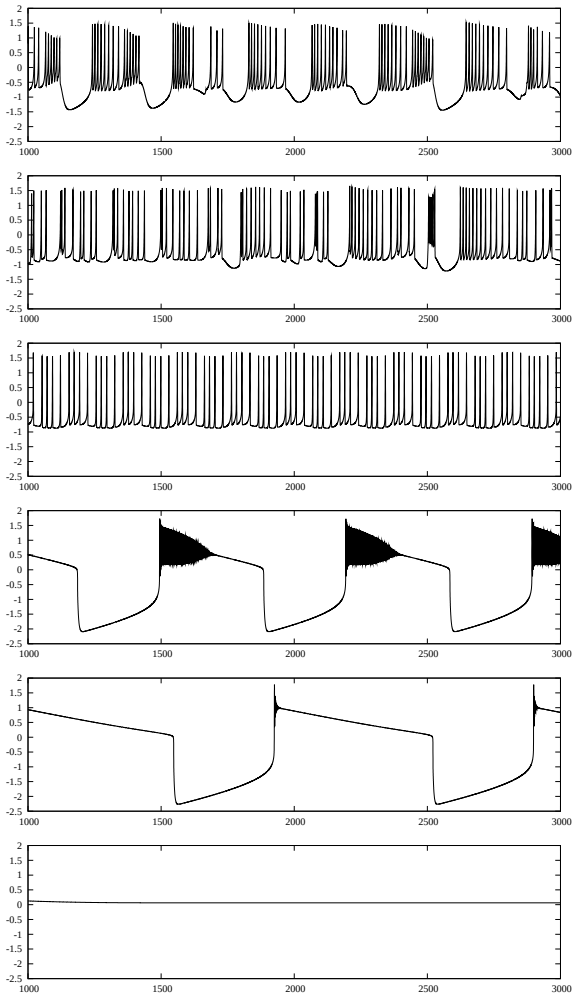


Figure 3: Times series (t, x) of a neuron (1) within a network with cycles for an increasing coupling strength (from top to bottom) 0.1, 0.5, 1.0, 1.3 and 3.0. Beyond a given value of the coupling strength, there is no more bursting behavior exhibited by the neuron.

3.2 Regular and quasi-regular networks

The simplest case of network constructed by levels is a chain network in which each level contains a single node connected to the node from the previous level. The results obtained for chain networks are summarized on Fig. 5. Fig. 5 (a)-(c) show the values of the three burst synchro-

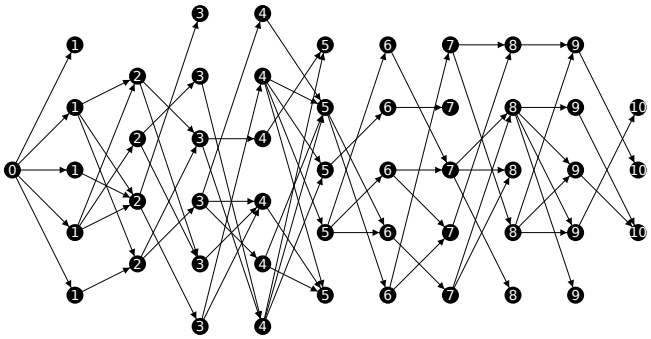


Figure 4: Example of a network constructed by levels containing 50 nodes distributed in 11 levels. Each node receives signal only from nodes of the previous level.

nization indicators defined in section 2. For each indicator we can see the presence of a threshold value of the coupling strength, above which the value of the indicator changes abruptly. These thresholds are shown as green lines on the figure. Fig. 5 (d) shows the three threshold lines drawn together. We can see that the first indicator, the ratio between the shortest inter-burst and the longest inter-spike distances, is predominant. In other words, when all the neurons have clearly expressed burst behavior, their bursts are synchronized. The same observation is valid for all the other network topologies we tested.

The experimental results show that the minimal coupling strength needed to synchronize a chain network scales linearly with the network diameter. In other words,

$$g(l) = Al + B \tag{7}$$

where $g(l)$ is the minimal coupling strength needed to synchronize the network up to level l. The values of coefficients A and B fitted by the least squares method to the simulation data are $A = 0.044$ and $B = 1.906$. Fig. 5 (d) shows also this linear approximation.

The same results are obtained for tree networks in which each node receives a signal from exactly one node from the previous level. Indeed, in these networks the in-degree of each node, except the root, is one and the behavior on each path from the root to another node is exactly the same as in a chain network.

The network diameter is not the only topological characteristics influencing the burst synchronization. The law observed for chain and tree networks is not preserved when we introduce nodes of different in-degrees. To illustrate the influence of the in-degree, consider a modified chain network as shown in Fig. 6(a). At a given level l we introduce k nodes each of them receiving signal from the node of level $l-1$ and sending signal to the node of level $k+1$. Thus, the in-degree of the root is 0, the in-degree of the node on level $l+1$ is k and the in-degrees of all the other nodes are 1.

Fig. 7 shows the burst synchronization thresholds for modified chain networks in which $k = 2$ or 5 nodes are introduced on level 20. When the network is synchronized up to level 20, there is no need to increase the coupling strength to synchronize the next several levels. In fact the behavior of the node on level 21 is roughly the same as if it was coupled with one node but with k times greater coupling strength. Nevertheless, this influence is only local, some more levels away the modification is progressively "forgotten" and the coupling strength joins again the linear law (7).

To study further the influence of the nodes in-degree, we consider a regular network shown in Fig. 6(b). In this type of networks all levels except level 0 contain the same number of nodes k and each node is connected to all the nodes from the previous level. In order to ensure the same in-degree of all nodes, the nodes of level 1 are connected to the root with k links. Fig. 8 shows the experimental results for this kind of networks. For the sake of simplicity, only the cases $k = 2$ and $k = 3$ are shown, but the results are similar for greater values of k. We can see that the coupling strength needed for burst synchronization grows linearly with the network diameter also in this case. Moreover, the simulation data fits to the lines

$$g_k(l) = \frac{Al + B}{k} \tag{8}$$

where $g_k(l)$ is the minimal coupling strength needed to synchronize a regular network with k nodes per level up to level l and A and B are the same coefficients as in the case of chain network. In particular, for $k = 1$, we obtain exactly (7). This result is not surprising, because a neuron coupled with k identical neurons with coupling strength g behaves as if it was coupled with a single neuron but with coupling strength kg.

3.3 Random networks

We have seen that the burst synchronization is influenced by two main characteristics of the network: the network diameter and the in-degree of the nodes. For regular networks in which the nodes in-degrees are the same, the coupling strength needed for burst synchronization grows linearly with the network diameter. In this section, we consider networks with heterogeneous in-degrees. They are always constructed by levels as described previously but each level contains a random number of nodes randomly connected to the nodes of the previous level. For our experiment we fixed 384 nodes randomly distributed on 64 levels. We then generated different random subsets of all possible links between adjacent levels in order to obtain different average in-degrees. Fig. 9 shows the experimental results for average in-degrees 1, 1.5, 2 and 3. The case of in-degree 1 corresponds to a tree and logically we observe a linear growth of the coupling strength with the network diameter. For bigger in-degrees, the coupling strength grows stepwisely with the network di-

ameter. This can be explained by looking again at Fig. 7. As we have seen, when a neuron receives more strength than it needs to synchronize, it synchronizes his successors several levels away. Fig. 9 shows also that the growth of the coupling force becomes slower when the in-degree is bigger. The places and the heights of the jumps in the lines are difficult to predict. In order to do this, an aggregate parameter as the average in-degree is not sufficient. The coupling strength needed to synchronize a neuron seems to depend not only on its in-degree, but also on the in-degrees of its direct and indirect predecessors.

4 Conclusion and perspectives

In this paper, we are interested in the minimal coupling force needed to obtain burst synchronization within a network of chemically coupled Hindmarsh-Rose oscillators according to the topology of networks. We study the influence of two topological network characteristics: the network diameter and the in-degrees of the nodes. We performed numerical tests on different network topologies in order to highlight the role of these characteristics.

Our experiments show that the coupling force needed to synchronize a network grows with its diameter. Inversely, when the in-degree of the nodes grows, the network becomes easier to synchronize. In the case of (quasi-)regular networks where (almost) all the nodes have the same in-degree, the coupling force grows linearly with the network diameter. In irregular networks with heterogeneous in-degrees we observe stepwise growth of the coupling strength due to the fact that some nodes receive stronger signal than others which they propagate to their successors.

In this paper, we present preliminary results which need to be confirmed by other numerical simulations. To make the general trends more evident, we have started by considering a restricted class of network topologies (acyclic networks with specific level structure). In future numerical simulations we will progressively remove these restrictions. It will be interesting to make a more detailed study of the influence of the in-degrees and to understand the evolution of the coupling force needed to synchronize a node as a function of its in-degree, but also of the in-degrees of its predecessors, the in-degrees of their predecessors, etc. Finally, a mathematical formalism describing the burst synchronization needs to be developed and used to justify theoretically our experimental observations.

References

[1] Aziz-Alaoui M.A. Synchronization of chaos, *Encyclopedia of Mathematical Physics*, Elsevier, (2006) 213-226

[2] Batista C.A.S., Lopes S.R., Viana R.L., Batista A.M., *Delayed feedback control of bursting synchronization in a scale-free neuronal network*, Neural Networks, Volume 23, Issue 1, (2010) 114-124

[3] Belykh I., Lange E., Hasler M., *Synchronization of Bursting Neurons: What matters in the Network Topology*, Phy. Rev. Lett.94, 18, (2005) 188101.1-188101.4

[4] Corson N., Balev S., Aziz-Alaoui M.A., *Detection of Synchronization Phenomena in Networks of Hindmarsh-Rose Neuronal Models*, ECCS'10 European Conference on Complex Systems, Lisbon, (2010)

[5] Dutot A., Guinand F., Olivier D. Pigné Y., *GraphStream: A Tool for bridging the gap between Complex Systems and Dynamic Graphs*, EPNACS: Emergent Properties in Natural and Artificial Complex Systems, (2007) http://graphstream-project.org/

[6] Han F., Lu Q., Wiercigroch M., Ji Q., *Chaotic burst synchronization in heterogeneous small-world neuronal network with noise, International Journal of Non-Linear Mechanics*, Volume 44, Issue 3, (2009) 298-303

[7] Hindmarsh J.L., Rose R.M., *A model of neuronal bursting using three coupled first order differential equations*, Proc. R. Sc. Lond. B221 (1984) 87-102

[8] NVIDIA CUDA^{TM}, *NVIDIA CUDA C Programming Guide*, (2010)

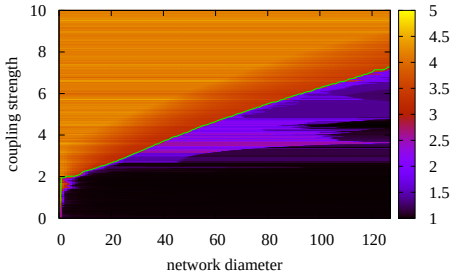
[9] Pecora L.M., Carrol T.L., *Synchronization in chaotic systems*, Phys. Rev. Lett. 64 (1990) 821-824

[10] Shi X., Lu Q., *Burst synchronization of electrically and chemically coupled map-based neurons*, Physica A: Statistical Mechanics and its Applications, Volume 388, Issue 12, (2009) 2410-2419

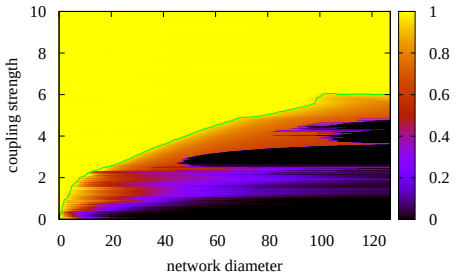
[11] Wang Q.Y., Lu Q.S., Chen G.,*Ordered bursting synchronization and complex wave propagation in a ring neuronal network*, Physica A: Statistical Mechanics and its Applications, Volume 374, Issue 2, (2007) 869-878

[12] Yamada T., Fujisaka H., *Stability theory of synchronized motion in coupled-oscillator systems*,Progress of Theoretical Physics 70 (1983) 1240-1248

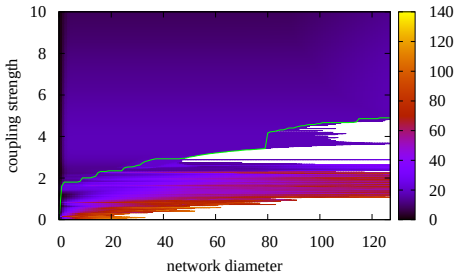
[13] Zheng Y.H., Lu Q.S., *Spatiotemporal patterns and chaotic burst synchronization in a small-world neuronal network*, Physica A: Statistical Mechanics and its Applications, Volume 387, Issue 14, (2008) 3719-3728



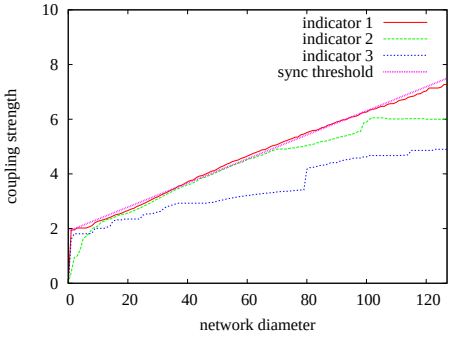
(a)



(b)

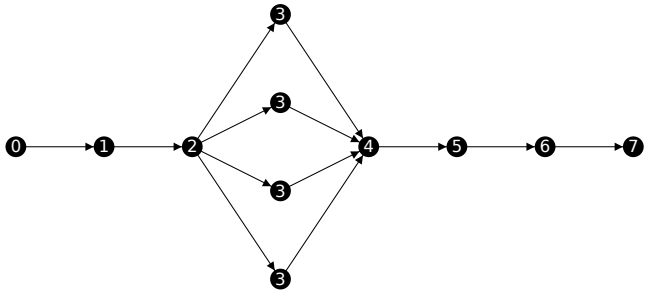


(c)

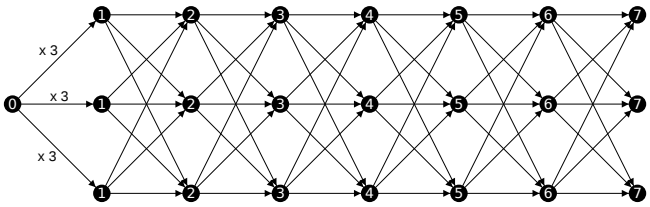


(d)

Figure 5: Burst synchronization indicators for a chain network as a function of the coupling strength and the network diameter. (a) Ratio between the shortest inter-burst and the longest inter-spike distances. (b) Proportion of matching bursts. (c) Average span of burst starts within matching groups. (d) Superposition of the three indicators and linear approximation of the synchronization threshold.



(a)



(b)

Figure 6: (a) Modified chain network. Level 3 contains four nodes. (b) Regular level network. Each level contains 3 nodes connected to all the nodes from the previous level.

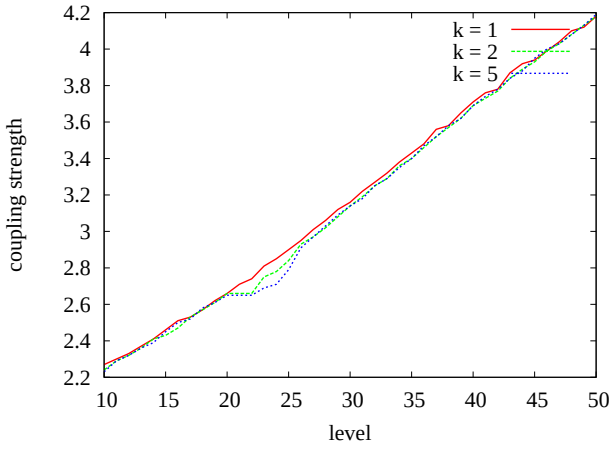


Figure 7: Coupling strength needed to synchronize modified chain networks. k nodes are introduced on level 20. Results for $k = 1, 2$ and 5.

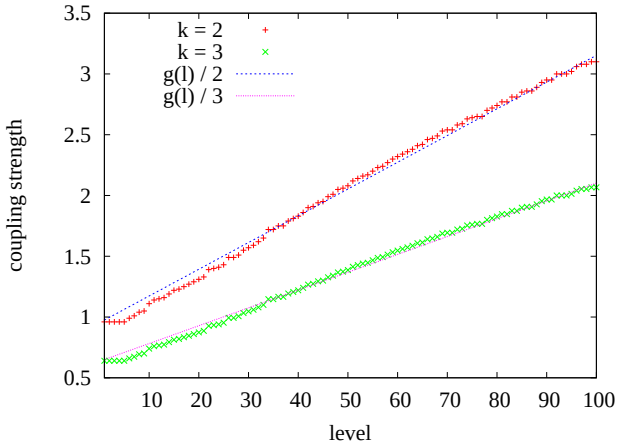


Figure 8: Coupling strength needed to synchronize regular network with k nodes per level. Results for $k = 2$ and 3.

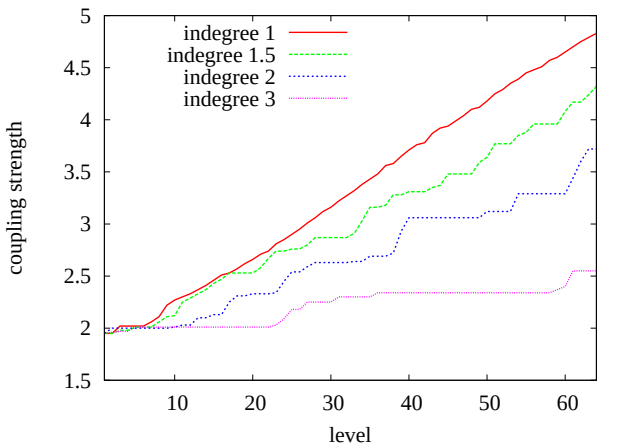


Figure 9: Coupling strength needed to synchronize random networks with different average in-degrees.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

TOPOLOGY IDENTIFICATION OF COMPLEX DYNAMICAL NETWORK WITH HINDMARSH-ROSE NEURONS

Junchan Zhao, Nathalie Corson, M. A. Aziz-Alaoui, and Cyrille Bertelle *†

Abstract. Complex networks, such as transportation and phone networks, Internet, wireless networks and the World Wide Web, play an important role in our life nowadays. Significant progress has been made in studying complex networks since the discovery of their small-world [1] and scale-free[2] characteristics.

Synchronization, as an interesting phenomenon of a population of dynamically interacting units, has received wide attention from different fields. Assuming the topology of complex network known, many researchers discuss the synchronization problem of complex network and obtain a lot of important results. For example, Pecora and Carroll studied the master stability functions for synchronized coupled systems[3]. Using the theory of inhomogeneous Markov chains, Wu proposed a synchronization criterion for nonautonomous discrete-time linear system in random directed network [4]. Belykh *et al.* proposed the synchronization of coupled chaotic systems via connection graph stability method [5, 6].

The topology identification, as an inverse problem, is a significant issue in the study of complex networks. For example, if a major malfunction occurs in a communication network, power network or the Internet, it is very important to quickly detect the location of the faulty line. Topology identification of complex dynamical networks has received more and more attention from the systems science community. Generally speaking, two methods are used to study the identification problem [7]-[14]. One is based on the classic adaptive control. Considering the complex dynamical network with unknown topology as a drive system, some researchers built a response system and adaptive controls to estimate the unknown topology. The other mainly used method consists on solving large scale linear equation. Driving the complex dynamical network to an equilibrium state by a control law, the identification problem can be transformed to solve the large linear equation. It is worth remarking that the earlier results neglected a crucial condition: persistent excitation (PE) condition [15]. The condition is so important that the topology cannot be identified successfully without it.

In 1952, two neurophysiologists proposed a mathematical model that describes neuron activity [16]. This model have been modified into different other models. In this study, we focus on one of them, the Hindmarsh-Rose model (HR), which can exhibit most of biological neuron behavior, such as spiking, bursting [17, 18]. Meanwhile, it is also important to investigate the group action when they are coupled each other.

Let us consider a network composed by n HR neurons. These neurons are coupled by their first variable x_i, which described by

$$\begin{cases} \dot{x}_i = ax_i^2 - x_i^3 + y_i - z_i - \sum_{j=1}^{n} c_{ij} h(x_i, x_j), \\ \dot{y}_i = (a+\alpha)x_i^2 - y_i, \\ \dot{z}_i = \epsilon(bx_i + c - z_i), \end{cases} \tag{1}$$

for $i = 1, ..., n$, where h is the coupling function and $C = c_{ij}$ is the network adjacency matrix. We assume that the elements of the matrix C are unknown. When the neurons are chemically coupled, the coupling function h is given by

$$h(x_i, x_j) = k\frac{x_i - V}{1 + exp(-\lambda(x_j - \Theta))}, \tag{2}$$

where k is the coupling strength, Θ is the threshold reached by every action potential for a neuron and V is the reversal potential.

Recently, Chenhui Jia *et al* designed a bridging network and adaptive law to estimate the topology of systems (1), and give many significant results. However, these results seem questionable, since they neglected the PE condition. In this paper, we estimate these unknown parameters in an asymptotic manner and discuss the important of PE condition on the process of identification. Pinning control is used to drive the response system synchronize with the drive system because it is difficult to control all the nodes of the large network. Additionally, we consider the topology identification of complex network with noise as many real systems will be subject to perturbation.

Keywords. Complex network, Topology identification, Synchronization, Persistent excitation.

*Junchan Zhao is with College of Mathematics and Computer, Wuhan Textile University, 430073, Wuhan, China and also with LMAH, University of Le Havre, 76058, Le Havre Cedex, France. E-mail: junchanzhao@gmail.com

†Nathalie Corson, M. A. Aziz-Alaouib are with LMAH and Cyrille Bertelle is with LITIS, University of Le Havre, 76058, Le Havre Cedex, France. E-mails: nathalie.corson@gmail.com, aziz.alaoui@univ-lehavre.fr, cyrille.bertelle@gmail.com.

References

[1] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature*, vol. 391, 1998, pp. 440-442.

[2] A. L. Barabasi and R. Albert, Emergence of scaling in random networks, *Science*, vol. 286, 1999, pp. 509-512.

[3] L. M. Pecora and T. L. Carroll, Master stability function for synchronized coupled systems, *Phys. Rev. Lett.*, vol. 80, 1998, pp. 2109-2112.

[4] C. W. Wu, Synchronization and convergence of linear dynamics in random directed networks, *IEEE Trans. Autom. Control*, vol. 51, 2006, pp. 1207-1210.

[5] V. N. Belykh, I. V. Belykh and M. Hasler, Connection graph stability method for synchronized coupled chaotic systems, *Physica D*, vol. 195, 2004, pp. 159-187.

[6] I. V. Belykh, V. N. Belykh and M. Hasler, Generalized connection graph method for synchronization in asymmetrical networks, *Physica D*, vol. 224, 2006, pp. 42-51.

[7] J. Lu and J. Cao, Synchronization-based approach for parameters identification in delayed chaotic neural networks, Physica A, vol. 382, 2007, pp. 672-682.

[8] X. Q. Wu, Synchronization-based topology identification of weighted general complex dynamical networks with time-varying coupling delay, Physica A, vol. 387, 2008, pp. 997-1008 .

[9] W. Yu, G. Chen, J. Cao, J. Lü, and U. Parlitz, Parameter identification of dynamical systems from time series, Phys. Rev. E. 75, 2007, 067201.

[10] L. Chen, J. Lu and Chi K. Tse, Synchronization: an obstacle to identification of network topology, IEEE Trans. Circ. Syst-II, vol. 56, 2009, pp. 310-314.

[11] J. Zhao, Q. Li, J. A. Lu, and Z. P. Jiang, Topology identification of complex dynamical networks, Chaos, 20, 2010, 023119.

[12] C. Jia, J. Wang, B. Deng, X. Wei, and Y. Che, Estimating and adjusting abnormal networks with unknown parameters and topology Chaos, 21, 2011, 013109.

[13] Jie Ren, Wen-Xu Wang, Baowen Li and Ying-Cheng Lai, Noise bridges dynamical correlation and topology in coupled oscillator networks, Phys. Rev. Lett, 104, 2010, 058701.

[14] C. Zhan, B. Li and L. Yeung, Identifying topology of general dynamical networks from noisy data, J. Stat. Mech. 2010, P12034.

[15] H. Khalil, Nonlinear systems, 2nd ed, Englewood Cliffs, NJ: Prentice Hall, 1996.

[16] A.L. Hodgkin, A.F. Huxley, A quantitative description of mem- brane current and its application to conduction and excitation in nerve, J. Physiol, vol. 117, 1952, pp. 500-544.

[17] J.L. Hindmarsh, R.M. Rose, A model of the nerve impulse using two first-order differential equations, Nature, vol. 296, 1982, pp. 162-164.

[18] J.L. Hindmarsh, R.M. Rose, A model of neuronal bursting us- ing three coupled first order differential equations, Proc. R. Sc. Lond. B, vol. 221, 1984, pp. 87-102.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

NEW ALTERNATE LOZI FUNCTION FOR RANDOM NUMBER GENERATION

Andrea Espinel, Ina Taralova and René Lozi *†

Abstract. An improved Lozi function with alternate coefficients has been proposed. The modifications in the model allow to remove the holes in the attractor which are not desirable, but appeared in the previous Lozi function; in this way, an everywhere dense attractor can be obtained. Moreover, the strong sensitivity to the type of binarisation (conversion of real values to 0 and 1) has been demonstrated; this conversion to binary numbers is instrumental to apply the NIST tests for randomness. The results have been validated and compared via NIST tests, for the different methods of quantization. Finally, is has been verified that the random properties of the output signal have been improved thanks to the following strategies : under-sampling of the output signal, and the system order increasing.

Keywords. Nonlinear dynamical system, Lozi function, NIST test, discrete-time map, dense chaotic attractor, pseudo random number generator

1 Introduction

The accelerated development of modern data transaction applications such as telecommunications requires encoding techniques with higher standards of security. Classically, these encoding sequences are obtained using Pseudo Random Number Generators (PRNG). As an efficient alternative, the chaotic-based generators are used to achieve even higher demanding encryption standards. Indeed, the chaotic systems exhibit a plethora of properties which make them suitable to meet the above requirements. The advantage to use chaotic systems lies in their extreme sensitivity to small parameter and initial conditions variations: in this way, as many different chaotic carriers as wanted can be generated.

However, the appropriate selection of a chaotic map that satisfies cryptographic applications requirements is a huge problem. It has to be emphasized that all chaotic maps are not applicable, because the chaotic generator - which is deterministic - has to satisfy at the same time the criteria for closeness to random signals. Therefore many practical problems arise, from the choice of the chaotic generator and its parameters, to the chaotic properties verification after the quantisation. Ideally, for cryptographic applications and higher security, an everywhere dense chaotic attractor is required, so all chaotic signal samples shall appear with the same probability (indeed, if there are holes in the chaotic attractor, the values of the state vector corresponding to the holes will never take place). Lozi had demonstrated that highly efficient discrete-time chaotic generators can be obtained from quite simple models such as the piece-wise linear ones, under some conditions [3]. To evaluate the random properties of these generators, a set of statistical based test known as NIST test developed by the National Institute of Standards and Technology have been used. A first coupled chaotic map confined to the 2D torus has already been proposed as a PRNG in [1], which random characteristics have been validated using the NIST tests. Nevertheless, since the state variables were not equidistributed, it has been demonstrated that the chaotic attractor exhibited holes delimited by the discontinuity lines and their forward iterates [1]. Therefore, there have been regions in the state space which the system orbits never visited, thus deteriorating the randomness.

2 System Definition

To improve the latter results, in this paper we deal with the new Lozi system with alternate coupled maps, confined to the p-dimensional torus $T^p = [-1, 1]^p$ by the map $M_p : T^p \Rightarrow T^p$

$$\begin{aligned} x^1_{n+1} &= 1 - 2\left|x^1_n\right| + k^1 \times x^2_n \\ M_p : x^2_{n+1} &= 1 - 2\left|x^2_n\right| + k^2 \times x^3_n \\ &\vdots \\ x^p_{n+1} &= 1 - 2\left|x^p_n\right| + k^p \times x^1_n \end{aligned} \tag{1}$$

where the parameters $k^i = (-1)^{i+1}$. A previous model with non alternate coefficients has been proposed in [1], with $k^i = 1$. The state variables are contained on the

*Andrea Espinel and Ina Taralova are with IRCCyN, UMR CNRS 6597, Ecole Centrale de Nantes, France. E-mails: andrea.espinel-rojas, ina.taralova@irccyn.ec-nantes.fr

†René Lozi is with Laboratoire J.A. Dieudonné, UMR CNRS 6621, Universit de Nice Sophia-Antipolis, France. E-mail:lozi@unice.fr

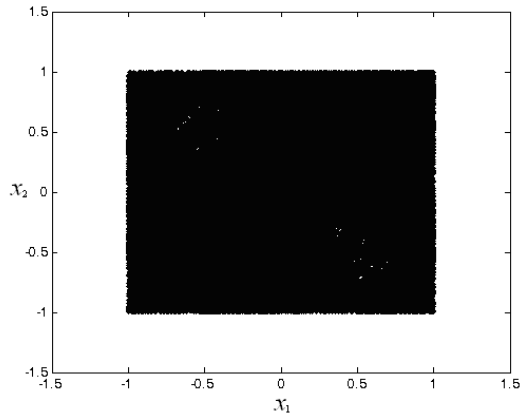


Figure 1: Map M_2 (1) on the torus $T^2 = [-1, 1]^2$

torus:

$$\begin{aligned} &if\ x_{n+1}^j = 1 - 2\left|x_n^j\right| + k^j \times x_n^{j+1} < -1 \\ &\qquad add\ 2 \\ &if\ x_{n+1}^j = 1 - 2\left|x_n^j\right| + k^j \times x_n^{j+1} < -1 \\ &\qquad substract\ 2 \end{aligned} \tag{2}$$

where $|x_n|$ denotes the absolute value of x_n. The alternate sign modification proposed in this paper eliminates the holes from the previous model (with only positive signs), and therefore the resulting basin of attraction is everywhere dense, which is very satisfactory for the RNG applications, see Figure 1. (The transient of 10.000 points has been cut off).

3 Results and Discussion

The random properties validation of a 4- dimensional system has been carried out. Additionally, the chaotic carrier output needs to be quantized and binarised (0 and 1) in order to be validated as being random using NIST tests [5]. First, the tests validate by default a sequence as being random (called the null hypothesis H0), and the idea is to show that there is no enough evidence to reject this proposition. For our application, different methods of binarisation(converting real signals to binary ones) have been implemented and compared:

A first 1-bit binarisation has been applied to the system (1) output:

$$\begin{aligned} &if\ y_n \geq 0 && b = 1 \\ &else && b = 0 \end{aligned} \tag{3}$$

The results for the 4-dimentional system have shown to be highly sensitive to the type of binarisation. Standart single (32 bits) and double (64 bits) precision types of binarisation as the IEEE754 form [6] have been compared. Therefore, after testing different methods, a 32-bit binarisation has been chosen as being the most suitable solution: Since the system is confined to the p-dimensional torus $T^p = [-1, 1]^p$, 31 bits are assigned to represent the decimal part, and 1 bit to the sign.

To illustrate the results, the NIST test for the four dimensional Lozi system M_4 (1) with parameters $\left[k^1, k^3\right] = 1$, $\left[k^2, k^4\right] = -1$ are shown in Table 1. For comparison, the same conditions as in [1] have been chosen:

Length of the original sequence: 10^8 bits
Length of *bit string*: 1M
Quantity of *bit strings*: 100

The output of the system has been arbitrary chosen as being: $y = x_4$.

P-VALUE	PROPORTION	STATISTICAL TEST
0.419021	100/100	Frequency
0.213309	100/100	BlockFrequency
0.978072	99/100	CumulativeSums
0.964295	99/100	CumulativeSums
0.075719	100/100	Runs
0.867692	99/100	LongestRun
0.494392	99/100	Rank
0.334538	99/100	FFT
0.213309	99/100	NonOverlappingTemplate
0.616305	99/100	OverlappingTemplate
0.779188	100/100	Universal
0.474986	99/100	ApproximateEntropy
0.452799	68/69	RandomExcursions
0.063482	69/69	RandomExcursionsVariant
0.437274	98/100	Serial
0.739918	99/100	LinearComplexity

Table 1: NIST test for the 4^{th} order Lozi function (1)

Furthermore, as the results show their independence from the initial conditions, every *bit string* in this first test is the resulting sequence of a different randomly chosen initial condition. The criterion for a successful test is that the p-value has to be superior to the significance level (0.01 for this case). This quantifier evaluates the uniformity of the zeros and the ones distribution in the sequence. For the present model (1), all tests have been successful thus the sequence can be accepted as being random. Thus, the results demonstrate that the new system has better statistical performances than the initial system without alternate coefficients presented in [1].

Finally, to improve the random properties of the signal, two possible strategies are suggested: undersampling of the output signal, or increasing the system order.

Different under-samplings have been tested from which the "1 out of 10" showed to be particularly successful. The "1 out of 10" under-sampling strategy results are shown in Table 2.

For the second method, the random properties validation of a 10-dimensional system has been carried out and the results are shown in Table 3. The conditions for the NIST test are identical to the NIST test for the

4-dimentional Lozi system. In addition, the initial condition has been randomly chosen:

$$x_0 = [-0.3365, 0.9501, 0.8913, -0.7764, 0.0185, 0.4447, 0.7919, -0.9218, -0.9355, 0.0579]$$

The output of the system has been arbitrary chosen as being: $y = x_{10}$.

P-VALUE	PROPORTION	STATISTICAL TEST
0.911413	99/100	Frequency
0.759756	97/100	BlockFrequency
0.897763	100/100	CumulativeSums
0.955835	99/100	CumulativeSums
0.122325	99/100	Runs
0.474986	99/100	LongestRun
0.911413	97/100	Rank
0.366918	99/100	FFT
0.419021	97/100	NonOverlappingTemplate
0.334538	99/100	OverlappingTemplate
0.935716	100/100	Universal
0.816537	98/100	ApproximateEntropy
0.128379	63/63	RandomExcursions
0.654467	61/63	RandomExcursionsVariant
0.554420	98/100	Serial
0.678686	99/100	LinearComplexity

Table 2: NIST test for the 4^{th} order system (1), "1 out of 10" under-sampling

P-VALUE	PROPORTION	STATISTICAL TEST
0.213309	100/100	Frequency
0.108791	98/100	BlockFrequency
0.075719	100/100	CumulativeSums
0.719747	100/100	CumulativeSums
0.719747	100/100	Runs
0.108791	100/100	LongestRun
0.816537	98/100	Rank
0.946308	98/100	FFT
0.115387	99/100	NonOverlappingTemplate
0.798139	98/100	OverlappingTemplate
0.058984	100/100	Universal
0.616305	98/100	ApproximateEntropy
0.054199	60/60	RandomExcursions
0.232760	59/60	RandomExcursionsVariant
0.437274	99/100	Serial
0.401199	100/100	LinearComplexity

Table 3: NIST test for the 10^{th} order Lozi function (1)

The improvement of random properties of both strategies has been corroborated by the experimental results.

4 Conclusion

A new new alternate Lozi function confined to the torus has been proposed as a pseudo random number generator. Unlike the previous model, here the chaotic attractor is everywhere dense and there are no holes inside, in this way all output values are supposed to appear with the same probalility. Therefore, the new alternate Lozi function proposed in this paper has proved to be more efficient than the first one (without alternation of the coefficients). Moreover, the fourth order system has been analyzed and all the NIST tests for randomness have been successfull for the representative test sequences. Finally, a higher order system and an under-sampling of the output signal has been added and the results have shown to be very satisfactory.

By consequence, the proposed PRNG could be used for encryption and many other applications wherever a random number generator is required.

References

[1] A. Espinel, I. Taralova, R. Lozi, "Dynamical and Statistical Analysis of a New Lozi Function for Random Numbers Generation," *PHYSCON 2011*, León, Spain, 5 − 8 September, 2011.

[2] R. Lozi, "Random properties of ring-coupled tent maps on the torus", submitted to *Discrete and continuous Dynamical Systems Series-B*.

[3] R. Lozi, "New enhanced chaotic number generators", *Indian Journal of Industrial and Applied Mathematics*, vol.1, pp. 1-23, 2008.

[4] S. Hénaff, I. Taralova, R. Lozi, "Statistical and spectral analysis of a new weakly coupled maps system", *Indian Journal of Industrial and Applied Mathemaics*, vol 2. N2, pp. 1-18 (to appear).

[5] A. Rukhin, et al, "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", *NIST (2001)*, http://csrc.nist.gov/rng/.

[6] W. Kahan, "IEEE Standard 754 for Binary Floating Point Arithmetic, *Lecture Notes on the Status of IEEE 754*, Elect. Eng. & Computer Science University of California, Berkeley CA 94720-1776, May 1996.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

IMPROVING CHAOTIC OPTIMIZATION ALGORITHM USING A NEW GLOBAL LOCALLY AVERAGED STRATEGY

Tayeb Hamaizia, René Lozi *†

Abstract. Recently chaotic optimization algorithms as an emergent method of global optimization have attracted much attention in engineering applications. Their good performances have been emphasized [1, 2, 3]. In the frame of evolutionary algorithms, the use of chaotic sequences instead of random ones has been introduced by Caponetto and al. [4]. Since their original work, the literature on chaotic optimisation is flourishing. They are used in the scope of tuning method for determining the parameters of PID control for an automatic regulator voltage, or in order to solve economic dispatch problems, or also for engineering design optimization and in many others physical, economical and biological problems. Different chaotic mapping have been considered, combined with several working strategies. The assessments of the algorithms have been done with respect to numerous objective functions in 1, 2 or 3-dimension. In this paper we present an improvement of the COLM (Chaotic Optimization based on Lozi Map) presented in [1], which is based on a new global locally averaged strategy. The simulation results are done with a 2-D objective function possessing hundreds of local minima, in order to test this new method vs. the previous one in very tough conditions. We emphasize an improvement of the optimisation.

1 Improved COLM Method

Chaos theory (the term chaos was coined par Li and Yorke [5]) is recognized as very useful in many engineering applications. An essential feature of chaotic systems is sensitive dependence on initial condition, (i.e. small changes in the parameters or the starting values for the data lead to drastically different future behaviours). Details about analysis of chaotic behavior can be found in [5, 6, 7, 8, 9]. The application of chaotic sequences can be an interesting alternative to provide the search diversity in an optimization procedure. Due to the non-repetition of chaos, it can carry out overall searches at higher speeds than stochastic ergodic searches that depend on probabilities. A novel chaotic approach is proposed in [1] based on Lozi map [6] which is piecewise linear simplification of the Hénon map [10] and it admits strange attractors. It is given by

$$\begin{cases} y_1(k) = 1 - a|y_1(k-1)| + by(k-1) \\ y(k) = y_1(k-1) \end{cases} \quad (1)$$

where k is the iteration number. In this work, the values of y are normalized in the range [0,1] to each decision variable in 2-dimensional space of optimization problem. This transformation is given by

$$z(k) = \frac{(y(k) - \alpha)}{\beta - \alpha}. \quad (2)$$

where $y \in [-0.6418, 0.6716]$ and $[\alpha, \beta] = [-0.6418, 0.6716]$. The parameters used in this work are $a = 1.7$ and $b = 0.5$. Numerical computation leads to the density $d(s)$ of iterated values of $y(k)$ displayed on Fig. 1. In this figure, the density is normalized to 1 over the whole interval $[0, 1]$ i.e.

$$\int_0^1 d(s)ds = 1.$$

The COLM method introduced in [1] is improved by locally averaging the global search, doing few steps of chaotic local search around every point obtained by the chaotic series.

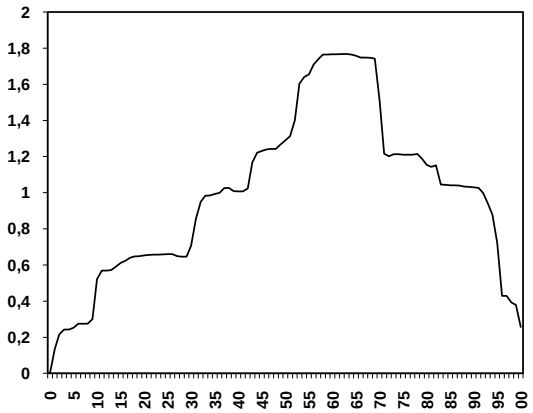


Figure 1: density of iterated values of $y(k)$ of equation (1) over the interval $[0, 1]$ splitted in 100 boxes for 10,000,000,000 iterated values.

Heuristics: the global locally averaged strategy of Improved COLM leads to better results than COLM as shown on Fig. 2. In this figure only three global search results are displayed $x_1(k)$, $x_2(k)$, $x_3(k)$ with

$$f(x_2(k)) < f(x_3(k)) < f(x_1(k)). \quad (3)$$

The local search following global one starts from the best global result $x_2(k)$ (from (3)) and gives $x_2(k+1)$. Instead the local-global search around $x_1(k)$, $x_2(k)$ and $x_3(k)$, leads to $x_1(k+1)$, $x_2(k+1)$, $x_3(k+1)$ which verify

$$f(x_1(k+1)) < f(x_3(k+1)) < f(x_2(k+1)). \quad (4)$$

The local search following the local-global one starts now from the best globally averaged result $x_1(k+1)$ (from(4)) and leads to $\bar{x}$

$$f(\bar{x}) < f(x_1(k+1)). \quad (5)$$

During the chaotic local search, the step λ is an important parameter in convergence behavior of optimization. Hence, two different

*Department of Mathematics, University of Constantine, (25000) Algeria, and Laboratoire J.A. Dieudonné, UMR CNRS 6621 Université de Nice Sophia-Antipolis, Parc Valrose 06108 NICE Cedex 02, France.

†Laboratoire J.A. Dieudonné, UMR CNRS 6621 Université de Nice Sophia-Antipolis, Parc Valrose 06108 NICE Cedex 02, France. E-mails:rlozi@unice.fr.

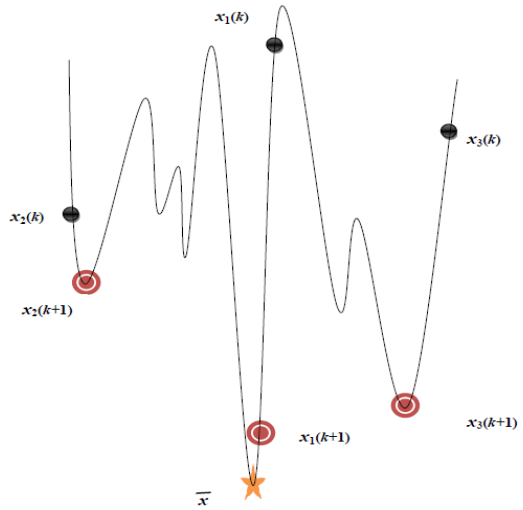


Figure 2: Heuristics of the global locally-averaged strategy.

values of λ are successively employed during the local search. We call this method ICOLM (improved COLM). Many unconstrained optimization problems with continuous variables can be formulated as the following functional optimization problem. Find X to minimize $f(X), X = [x_1, x_2, , x_n]$ Subject to $x_i \in [L_i, U_i]$. Where f is the objective function, and X is the decision solution vector consisting of n variables $x_i \in R^n$ bounded by lower (L_i) and upper limits (U_i).The ICOLM can be illustrated as follows:

Inputs:
M_g: max number of iterations of chaotic Global search.
M_l: max number of iterations of chaotic Local search.
Mgl_1: max number of iter of chaotic Local search in Global search.
Mgl_2: max number of iter of chaotic Local search in Global search.
$Mg \times (Mgl_1 + Mgl_2) + Mg$: stopping criterion of chaotic optimization method in iter.
λ_{gl1}: step size in first global-local search.
λ_{gl2}: step size in second global-local search.
λ: step size in chaotic local search.
Outputs:
$\bar{X}$: best solution from current run of chaotic search.
$\bar{f}$: best objective function (minimization problem).

Step 1 : Initialize the number M_g, M_{gl1},M_{gl2}, M_l of chaotic search and initialization of variables and initial conditions Set k=1, y (0), $y_1(0)$,$a = 1.7$ and $b = 0.5$ of Lozi map. Set the initial best objective function $\bar{f} = +\infty$

- **Step 2**: algorithm of chaotic global search:**while** $k \leq M_g$ **do**
$x_i(k) = L_i + z_i(k).(U_i - L_i)$
if $f(X(k)) < \bar{f}$ **then**
$\bar{X} = X(k); \bar{f} = f(x(k))$
end if
- **Step 2-1**: sub algorithm of chaotic local search:
while $j \leq M_{gl1}$ **do**
for $i = 0$ to n **do**
if $r \leq 0.5$ **then**
$x_i(j) = \bar{x_i} + \lambda_{gl1} z_i(j).|(U_i - L_i)|$
else
$x_i(j) = \bar{x_i} - \lambda_{gl1} z_i(j).|(U_i - L_i)|$
end if
end for
if $f(X(j)) < \bar{f}$ **then**
$\bar{X} = X(j); \bar{f} = f(x(j))$
end if
$j = j + 1$
end while
- **Step 2-2**: sub algorithm of chaotic local search:
while $s \leq M_{gl2}$ **do**
for $i = 0$ to n **do**
if $r \leq 0.5$ **then**
$x_i(s) = \bar{x_i} + \lambda_{gl2} z_i(s).|(U_i - L_i)|$
else
$x_i(s) = \bar{x_i} - \lambda_{gl2} z_i(s).|(U_i - L_i)|$
end if
end for
if $f(X(s)) < \bar{f}$ **then**
$\bar{X} = X(s); \bar{f} = f(x(s))$
end if
$s = s + 1$
end while
$k = k + 1$
end while
- **Step 3**: algorithm of chaotic local search:
while $k \leq M_g \times (M_{gl1} + M_{gl2}) + M_l$ **do**
for $i = 0$ to n **do**
if $r \leq 0.5$ **then**
$x_i(k) = \bar{x_i} + \lambda z_i(k).|(U_i - L_i)|$
else
$x_i(k) = \bar{x_i} - \lambda z_i(k).|(U_i - L_i)|$
end if
end for
if $f(X(k)) < \bar{f}$ **then**
$\bar{X} = X(k); \bar{f} = f(x(k))$
end if
$k = k + 1$
end while

2 A tough objective function

In order to test this new method vs. the previous one in very tough conditions the simulation results are done with the following 2-D objective function possessing hundreds of local minima: The function f which is very complex, has serval local maxima.

$$f = x_1^4 - 7x_1^2 - 3x_1 + x_2^4 - 9x_2^2 - 5x_2 + 11x_1^2x_2^2 + 99sin(71x_1) + 137sin(97x_1x_2) + 131sin(51x_2). \quad (6)$$

We test ICOLM on the search domain: $-10 \leq x_i \leq 10, i = 1, 2$.

The essential feature of this benchmark function is that location of minima is not symmetric. In a forthcoming paper we will extend our numerical analysis in higher dimension with an extended benchmark suite [11].

3 Numerical results

We display few of the results we have obtained showing the better optimization results obtained by this new methods.In each case study, 48 independent runs were made for each of both the COLM and ICOLM methods involving 48 different initial trial conditions $y_1(0)$, $y(0)$ (parameters of Lozi map). not far from the value of f on the global minimum.
For all studied cases, the four configurations, numbered from IC1 to IC4 and C1 ti C4, that are used are presented in tab. 1. The locally averaged strategy of ICOLM is illustrated on Fig. 3 on which the result of every step 2-2 is plotted.

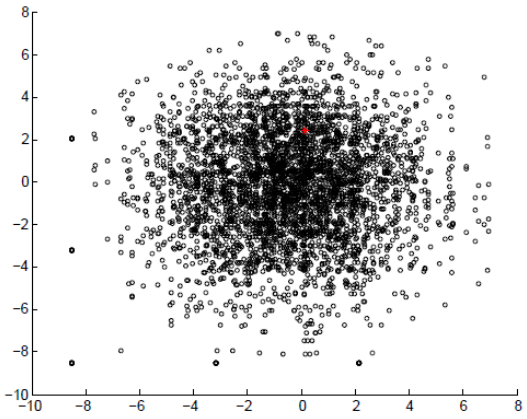


Figure 3: Locally-averaged strategy of chaotic search. Results of every Step 2-2 for f

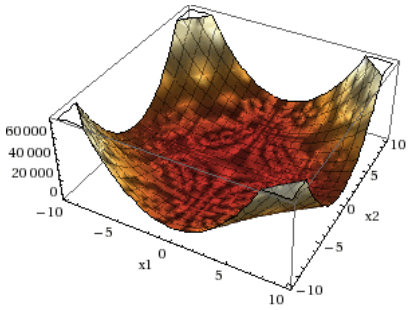


Figure 4: plot of test function used in this study

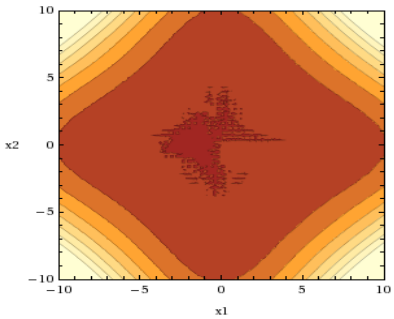


Figure 5: Position of the minima in the search domain

	λ	$\lambda_{M_{gl1}}$	$\lambda_{M_{gl2}}$	M_g	M_l	M_{gl1}	M_{gl2}
IC1	0.001	0.04	0.01	6	50	2	2
IC2	0.01	0.04	0.01	10	50	2	2
IC3	0.1	0.04	0.01	10	50	2	2
IC4	0.1	0.04	0.01	100	50	5	5
C1	0.001			24	50		
C2	0.01			40	50		
C3	0.1			40	50		
C4	0.1			1000	50		

Table 1: The set of parameters values for every run on the benchmark suite defined in Sec. 2. 2.

4 Conclusion

In every test, with the same computational cost, ICOLM gives better than COLM best values and Mean Best values but in one case. The presented study allows us to conclude that the proposed method is fast and converges to a good optimum. because we used a sampling mechanism to coordinate the research methods based on chaos theory, and we refined the final solution using a second method of local search. Further research is needed to gain more confidence and better understanding of the proposed methodology. The proposed algorithm has to be evaluated for a large number of test functions in higher dimension.

References

[1] L. S. Coelho. "Tuning of PID controller for an automatic regulator voltage system using chaotic optimization approach", Chaos, Solitons and Fractals, 39, 2009, 1504-1514.

[2] L. S. Coelho. "Reliability-redundancy optimization by means of a chaotic differential evoluytion approach", Chaos, Solitons and Fractals, 41, 2009, 594-602.

[3] D. Davendra, I. Zelinka, R. Senkerik, "Chaos driven evolutionary algorithms for the task of PID control", Computers and Mathematics with Applications, 60, 2010, 1088-1104.

[4] R. Caponetto, L. Fortuna, S. Fazzino and M. G. Xibilia, "Chaotic Sequences to Improve the Performance of Evolutionary Algorithms", IEEE Transactions on Evolutionary Computation, Vol. 7, n 3, June 2003, 289-304.

[5] T. Y. Li and J. A. Yorke, "Period three implies chaos", Amer. Math. Monthly, 82 , 1975, 985-992.

[6] R. Lozi, " Un attracteur étrange? du type attracteur de Hénon ", J. Phys., 39(C5), 1978, 9-10.

[7] S. H. Strogatz, Nonlinear dynamics and chaos, Massachussets: Perseus Publishing, 2000.

[8] T. S. Parker, L.O. Chua, Practical numerical algorithms for chaotic system. Berlin, Germany: Springer; 1989.

[9]] K. T. Alligood, T. D. Sauer, J. A. Yorke, Chaos: an introduction to dynamical systems, London, UK: Springer; 1996.

[10] M. Hénon, "A two dimensional mapping with a strange attractor" Commun. Math. Phys. 1976; 50(1):69-77.

[11] S. Cong, G.Li and X.Feng, "'An Improved Algorithm of Chaos Optimization", Proceedings 2010 8th IEEE International Conference on Control and Automation.

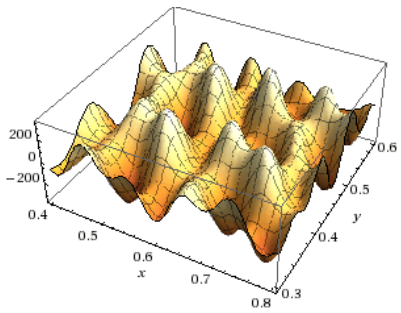


Figure 6: magnification of Fig. 4.

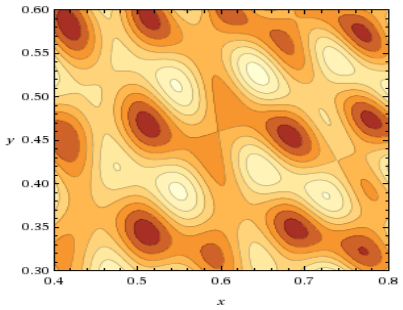


Figure 7: magnification of Fig. 6.

	Best value	Mean value	Std.Dev	$\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}$
IC1	-384.2891	-363.0185	11.5770	$\begin{pmatrix}-2.7686\\ -0.4045\end{pmatrix}$
IC2	-392.5400	-365.7837	14.0615	$\begin{pmatrix}-0.7327\\ 1.3203\end{pmatrix}$
IC3	-393.3134	-379.6872	8.8797	$\begin{pmatrix}-1.9677\\ -1.9982\end{pmatrix}$
IC4	-395.7338	-381.8734	9.6707	$\begin{pmatrix}-1.8815\\ -2.2501\end{pmatrix}$

Table 2: ICOLM

	Best value	Mean value	Std. Dev	$\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}$
C1	-371.0150	-368.5212	11.1135	$\begin{pmatrix}0.3105\\ 0.2442\end{pmatrix}$
C2	-358.4331	-352.9766	2.4424	$\begin{pmatrix}3.7283\\ 6.6158\end{pmatrix}$
C3	-377.9280	-368.7777	8.8169	$\begin{pmatrix}3.2738\\ 6.1685\end{pmatrix}$
C4	-382.7108	-379.7557	1.7817	$\begin{pmatrix}-5.8930\\ 2.9309\end{pmatrix}$

Table 3: COLM

Part II

Dynamical Networks and Graphs

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

CENTROIDS: A DECENTRALIZED APPROACH

Antoine Dutot*, Damien Olivier*and Guilhelm Savin*†

Abstract. The centroid of a graph is a structure composed of nodes closest from all others. This suggests the presence of center of mass average of all edges, weighted by the local density or specific weight. To compute this centroid in a *classic* way needs a global view of the graph environment. In this paper, we propose an algorithm using ant colony is proposed to compute an approximate solution of the centroid using a local view of the graph. This allows to study the centroids of complex networks such as protein-protein interaction networks and also those generated by social interactions or Internet, for example.

Keywords. complex system, self-organization, centroid, ant algorithm

1 Introduction

Complex systems are constituted by interactive numerous entities. These entities and their interactions define an interaction network which evolves in the time. We can use a dynamic graph [1, 2] to model this network and to describe its evolution. Through this graph, we can detect some properties emerging from interactions between entities.

A lot of real networks exhibit non-trivial properties. These complex networks present generally patterns of connection between their entities such as particular degree distribution, high clustering coefficient, hierarchical structure or communities. These features emerge from interactions and evolve in the time, notably, the self-organization defining a group of entities whose members maintain privileged interactions. The challenge is to detect them [3] and to define metrics. We assume that the notion of centers and centroids can help us in this approach.

It is impossible to make a reification of the whole graph because of numerous elements which consitute itself, however that is possible to distributed this graph over a set of machines. In such cases, each machine only has a *local view* of the graph. This local view raises problems when trying to use algorithms needing a *global view* of the graph.

Let us add another problem, the dynamic of the environment. In fact nodes and edges can appear or disappear, and attributes of elements can evolve, that makes obsolete the previously computed solution. Most algorithms are not able to use a previous solution to update it and they have to compute a new solution from scratch.

Ant System paradigm [4] allows to define metaheuristic where computation is distributed over a set of agents called *ants*. These ants explore the graph and make local action on crossed nodes and edges. The well-known *Traveling Salesman Problem* is given as practical example of problem that can be solved with an ant system.

We are interested here by finding a structure called *centroid* of a graph using an ant system. This paper is divided into two parts: section 2 is a state of art about centroid in literature while section 3 describes the ant heuristic and shows results on several graph categories.

2 Centers and Centroids

The term *centroid* is used in several scientific fields such as geometry or physics. Centroid is also used in human sciences to summarize, for example, a set of spatial points. The centroid of an object is its barycenter or its geometric center. In the case of a weighted object with uniform density, the centroid of this object is its *center of mass*. Centroid evokes the idea of a balancing point of the object. The following describes the concept of centroid applied to graphs. We not provide here a full referencing of properties or theorems about centroids but only an approach of what is a centroid in a graph. The concept of *center* being associated to the one of centroid, it is described to.

Centers and *centroids* are in literature the object of many works. From C. Jordan giving a proof of their existence in every tree [5] to P. J. Slater generalizing their definitions [6], through parts of graph theory books [7, 8].

In the following, some notations are used and have to be defined first. G designates a connected graph with $V(G)$ its node set and $E(G)$ its edge set. T designates a tree. $op(e, u)$, with $e = (u, v)$ an edge, designates the vertex v, the opposite of the vertex u according to e. $d(a, b)$ is the distance between two nodes a and b, *i.e* the number of edges in the shortest path between these two nodes. $e(v)$ is the *eccentricity* [8] of a node v and it is

*L.I.T.I.S., University of Le Havre, France,
authors are sorted alphabetically
†corresponding author: guilhelm.savin@litislab.fr

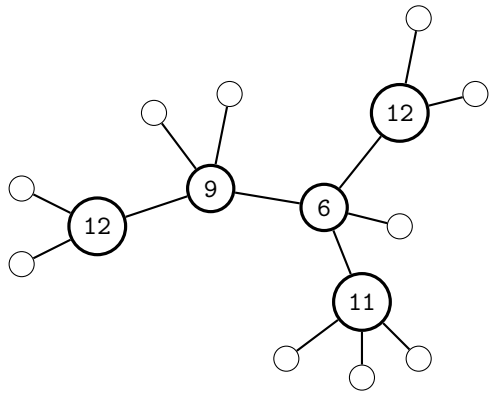


Figure 1: *Weights of non-leaf nodes in a tree as defined in [8].*

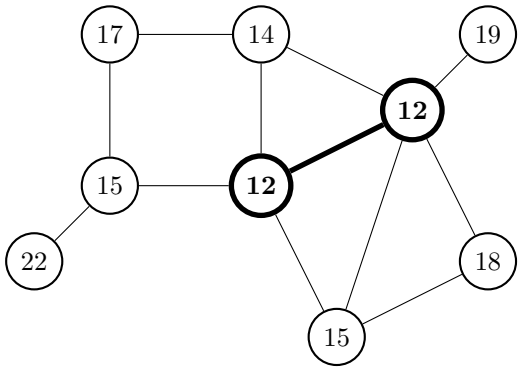


Figure 2: *Distances of nodes. Bold structure is the centroid of the graph.*

the maximal distance between v and any other node as defined in equation 1. The distance $d(v)$ of a node v is the sum of distance between v and the other nodes as defined in equation 2. The weight of a node v in a tree T is the maximum number of nodes in any branch of v. The figure 1 shows the weight of non-leaf nodes.

$$e(v) = \max_{x \in V(G)} (d(v, x)) \tag{1}$$

$$d(v) = \sum_{x \in V(G)} d(v, x) \tag{2}$$

The concept of *center* and *centroid* have been first defined for trees [5, 7, 8] and then extended to graphs.

A node c_0 of a graph G is considered as a *central point* of G [8] if its eccentricity $e(c_0)$ is minimal for G as defined in equation 3. The set of these central points define the *center* of G.

$$e(c_0) = \min_{v \in V(G)} (d(v)) \tag{3}$$

There are usually qualified centers. *Center of gravity* of a graph G is the set of nodes that minimize the function $m()$ defined in equation 4. A *mass center* of a tree T is a node with minimal weight.

$$m(v) = \frac{1}{|V(G)|} \sum_{x \in V(G)} d(v, x)^2 \tag{4}$$

The *centroid* of a graph G is a set of nodes called *centroid points* [8]. One defines the centroid as the subgraph induced by these centroid points [9], *i.e* the subgraph H of G with $V(H)$ is the set of centroid points and $E(H) = \{(u, v) | (u, v) \in E(G); u, v \in V(H)\}$. There are several functions in literature to characterize centroid points. In [8], the author defines centroid points of a tree T as the nodes with minimal weight, *i.e* the mass centers of the tree. One defines centroid points of a finite, connected, undirected graph G [6] using distance of a node (equation 2): centroid points are nodes with minimal distance. The figure 2 shows distance value of nodes. The subgraph induced by nodes with minimal distance is the centroid of G.

Slater defines the *k-centrum* of G [6] allowing to link definitions of center and centroid. "The sum of the k-largest vertex distances to" a node v is denoted $r_k(v)$ and is defined in equation 5. The set of nodes of a graph G that minimize is called the *k-centrum* of G and is denoted $C(G; k)$. When $k = 1$, $r_1(v)$ is the eccentricity of v so $C(G; 1)$ is the center of G. When $k = n$ with $n = |V(G)|$, $r_n(v)$ is the distance of v so $C(G; n)$ is the centroid of G.

$$r_k(u) = \max_{S \subseteq V(G); |S| = k} \left\{ \sum_{s \in S} d(u, s) \right\} \tag{5}$$

3 Decentralized algorithm

3.1 Definition

We are interested here by describing an *heuristic* that aims to find centroid of a graph using only local informations. This is an *ant-based* algorithm [4] where an ants colony explores a graph, dropping informations on elements, nodes and edges. These dropped informations are a form of indirect communication, called *stigmergy*, which allows ants to collaborate to find a solution. In such algorithms, one defines a behavior for ants that we denoted here the *step* of ants. The algorithm consists in a loop in which the step method of each ant is triggered.

The goal of this approach is to provide an algorithm that does not need a global reification of the environment (the graph) and that can handle dynamic of the environment by updating the solution. This algorithm computes an approximate centroid.

We consider that each node has a mass, which is initially the same for all nodes and ants are able to take a part of node mass and to distribute the taken mass on other nodes. The final aim is to increase mass of centroid

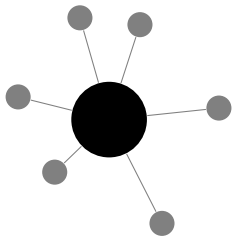


Figure 3: *The local view of the environment of an ant: the black node is the current position and gray nodes are its neighborhood.*

nodes and decrease mass of outlying nodes, the most massive nodes being parts of the centroid.

A ant a is composed of a few attributes:

- pos_a, the node position of the ant;
- $phn_{max}(a)$ and $mass_{max}(a)$ that are described in the following;
- $tmass_a$, the mass transported by a;
- a tabu list used to avoid to back track.

While ants are exploring the graph, they drop pheromones on edges they crossed. For an edge e, the measure $phe(e)$ is the load of pheromones of e and for a node n, $phn(n)$ is the sum of $phe(e_i)$ where e_i is an edge adjacent to n. These pheromones *evaporate* through time according to a factor ρ, so we have:

$$phe_{i+1}(e) = phe_i(e) \times \rho, \forall e \in E(G)$$

where i is the iteration of the algorithm. Each ant a remembers maximum pheromone load $phn_{max}(a)$ and maximum node mass $mass_{max}(a)$ that it has been found. Like pheromones, these two values evaporate through time to keep an up-to-date value reflecting the current state of the environment. The choice of the next edge to cross e_i adjacent to the ant position has a probability distribution $P(e_i)$ based on pheromone load and described in equation 6. The constant α helps to control the importance of pheromone load in choosing the next edge to cross. The function $\eta_a(e)$ defined in equation 7 is used to influence ants choice in the aim of leading them to massive nodes according to a parameter $\lambda \in [0;1]$: if λ is close to zero, then the probability of edges to be choosen is all the higher as the mass of the opposite node is high, else if λ is close to one, mass of node has no effect in the choice of the next edge.

$$P(e) = phe(e)^{\alpha} \times \eta_a(e) \tag{6}$$

$$\eta_a(e) = \lambda + (1 - \lambda) \times \frac{mass(op(e, pos_a))}{mass_{max}(a)} \tag{7}$$

The second part of the ant behavior is the ability to distribute node mass on other nodes. Each ant a has an attribute $tmass_a$ corresponding to the mass being transported. If $tmass_a$ is null, the ant takes a quantity q of the mass of its current position bounded between parameters $tmass_{min}$ and $tmass_{max}$, else the ant a computes a quantity $drop(a, n)$, defined in equation 8 to drop on its current position n.

$$drop(a, n) = tmass_a \times \frac{phn(n)}{phn_{max}} \tag{8}$$

The algorithm is a loop where evaporation of pheromones is made and where the behavior of each ant is triggered as described in algorithm 1. This behavior consists in choosing an edge to cross, cross it and then to transfer some mass. The algorithm describing this behavior is algorithm 2.

Algorithm 1: Global ant algorithm

```
Data:
    - G, the graph, E(G) is the edge set
    - A, set of ants
begin
    repeat
        forall edge e ∈ E(G) do
            phe(e) ← phe(e)×ρ
        forall ant a ∈ A do
            execute step of a
    until stop condition is reached
end
```

Algorithm 2: Step of an ant

```
Data:
    - a, the ant
    - t, mass transported by a
    - p, position of a
    - m, mass of p
begin
    forall edge i adjacent to p do
        compute P[i]
    randomly choose edge according to P
    p = ope(e, p) /* cross chosen edge */
    if t == 0 then
        alea ← random(tmass_min, tmass_max)
        q ← min(alea, m)
        t ← t + q
        m ← m- q
    else
        q = drop(a, p)
        t ← t- q
        m ← m + q
end
```

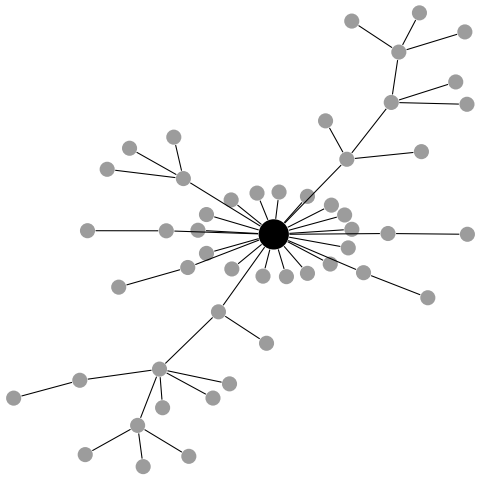


Figure 4: *In this tree of 50 nodes, ants have detected the optimal centroid (the biggest black node), i.e the centroid computed by the classic algorithm.*

3.2 Results

We compare here results of our ant algorithm with result of a *classic* algorithm consisting in computing all shortest paths of the graph using a Floyd-Warshall algorithm [10] and then to use the previous results to compute the distance of each node as defined previously in equation 2. The complexity of this algorithm is $O(n^3)$ with n is the amount of nodes. Values of parameters used in tests are available on table 5.

To compare results, one has to define a condition to stop the loop of the ant algorithm. For each node, the *stabilization* of the node is a measure described in equation 10 and computed from previous and current masses and bounded between 0 and 1 that indicates the evolution of the node mass : 0 means that the mass is strongly evolving, 1 means that the mass is stable. This allows the definition of the stabilization of the graph described in equation 11 which is the average value of node stabilization. The algorithm can stop when stabilization of the graph reaches a given threshold ϵ. For tests we used $\epsilon = 0.99$.

$$diff_t(n) = \frac{|mass_{t-1}(n) - mass_t(n)|}{mass_t(n)} \quad (9)$$

$$stab_t(n \in V(G)) = min(1, 1 - diff_t(n)) \quad (10)$$

$$stab_t(G) = \frac{\sum_{n \in V(G)} stab_t(n)}{|V(G)|} \quad (11)$$

Several kinds of graph have been considered for tests :

1. Trees built using a *preferential attachment* algorithm [11] as shown on figure 4;
2. Graphs built using a *Dorogotvsev Mendes* algorithm [12] that produces graphs with a power-low degree distribution ;
3. Two-dimensional grids ;
4. Two-dimensional grids with random holes.

Table 1: *Results for preferential attachment trees*

nodes	51	101	501	1001
edges	50	100	500	1000
$\overline{t_\Delta}$	-37.1%	-84.2%	-98.6%	-99.0%
$\sigma(t_\Delta)$	38.03%	5.07%	0.51%	0.00%
$\overline{s_\Delta}$	+0.5	+0.4	+0.5	+0.5
$\sigma(s_\Delta)$	1.13	0.77	0.77	0.87
$\overline{\Gamma}$	7.7%	6.7%	5.0%	4.0%
$\sigma(\Gamma)$	8.6%	7.3%	5.2%	4.1%

Table 2: *Results for dorogovtsev-mendes graphs*

nodes	53	103	503	1003
edges	103	203	1003	2003
$\overline{t_\Delta}$	-34.2%	-79.1%	-97.9%	-99.0%
$\sigma(t_\Delta)$	19.84%	5.28%	0.30%	0.00%
$\overline{s_\Delta}$	+1.5	+0.7	+0.6	+0.6
$\sigma(s_\Delta)$	2.67	1.13	0.83	0.68
$\overline{\Gamma}$	10.4%	6.2%	4.3%	3.7%
$\sigma(\Gamma)$	9.4%	5.8%	3.8%	3.4%

Comparison of results focuses on differences between computation time and between computed centroid. If $time_A$ is the computation time of the ant algorithm and $time_C$ the one of the classic algorithm, then

$$t_\Delta = \frac{time_A - time_C}{time_C}$$

express the gain or loss of time of the ant algorithm compared to the classic one. If C_A is the centroid computed by ants and C_C the one computed by the classic algorithm, then

$$s_\Delta = |C_A| - |C_C|$$

express the size difference between both centroids.

When a node v of C_A is not in C_C, we can define the shortest distance $\gamma(v)$ between v and any node of C_C normalized by the diameter of the graph. The measure Γ is the average value of $\gamma(v)$: if equals to 0 then both centroids are confused, else higher is the value, farther are the centroids.

3.3 Conclusion on results

Two points have to be considered in studying results. The first one is the computation time needed by the algorithm, the second one is the quality of the solution.

Table 3: *Results for grids*

nodes	36	121	256
edges	60	220	480
$\overline{t_\Delta}$	-15.9%	-89.3%	-96.7%
$\sigma(t_\Delta)$	31.99%	2.48%	0.78%
$\overline{s_\Delta}$	+2.9	+11.4	+13.3
$\sigma(s\Delta)$	3.51	6.83	9.69
$\overline{\Gamma}$	17.4%	17.8%	13.8%
$\sigma(\Gamma)$	4.7%	2.4%	1.8%

Table 4: *Results for incomplete grids*

nodes	22	93	214
edges	53	292	730
$\overline{t_\Delta}$	+197.8%	-77.8%	-94.4%
$\sigma(t_\Delta)$	122.77%	3.52%	0.88%
$\overline{s_\Delta}$	+4.1	+9.2	+13.7
$\sigma(s_\Delta)$	2.81	5.87	8.87
$\overline{\Gamma}$	25.3%	19.4%	15.1%
$\sigma(\Gamma)$	9.1%	5.2%	2.0%

Table 5: *Parameters value*

ϵ	0.99
λ	0.50
α	1.00
pheromon drop	0.03
pheromon evaporation	0.86

From results we can observe that the ant-algorithm is significantly faster than the classic algorithm. In addition, ant-algorithm is able to handle dynamic of graph by updating its current solution, whereas the classic algorithm needs to start from scratch. So, in these tests that use static graphs, benefits of the ant algorithm about the robustness of its solution within the ambit of a dynamic evolution of the graph.

Quality of the solution is harder to assess because it can be dependent on how is used this solution. We focus on how the ant-algorithm solution S_a matches the classic one S_c by observing size difference between these two solutions and the average of shortest distance from nodes in S_a to any node of S_c, denoted Γ. Unlike time difference, size difference and Γ are more heterogeneous. Best matches is for trees (table 1) followed by graphs generated with Dorogovtsev-Mendes algorithm (table 2): size difference is less than one and Γ is low. For grids (table 3) and incomplete grids (table 4), S_a is far from S_c : ants find more centroid points which are not close to centroid points of S_c.

4 Conclusion

We have proposed in this paper a new heuristic to compute an approximate centroid of a graph. This algorithm uses an ant system with ants that explore the graph dropping pheromone on crossed edges and moving the mass of nodes.

The algorithm has been applied to several static graph categories and results compared with the one of a greedy deterministic algorithm. This comparison shows that the ant algorithm is significantly faster than the classic one and the quality of its solution depends on the graph category: the centroid of ants algorithm is really close of the classic one for trees but for grid solution is far.

However, we have made these tests on static graphs whereas this new ant algorithm aims to be applied on dynamic graphs, providing a robust solution. More, the ant algorithm provides a distributed and decentralized computation of the solution allowing to use it on massive graphs.

4.1 Further work

A first future work is to compute centroid of dynamic graphs using this algorithm to test the robustness of the solution.

Then we have to study the impact of parameter values on the algorithm solution and to find a way the determine an optimal value for these parameter according to the graph.

References

[1] A. Dutot, F. Guinand, D. Olivier, and Y. Pigné, "Graphstream: A tool for bridging the gap between complex systems and dynamic graphs," in *EPNACS: Emergent Properties in Natural and Artificial Complex Systems*, 2007.

[2] Y. Pigné, "Modélisation et traitement décentralisé des graphes dynamiques : Application aux réseaux mobiles ad hoc," Ph.D. dissertation, Université du Havre, 12 2008. [Online]. Available: http://tel.archives-ouvertes.fr/tel-00371962/PDF/these.pdf

[3] C. Bertelle, A. Dutot, F. Guinand, and D. Olivier, "Organization detection using emergent computing," *International Transactions on Systems Science and Applications*, vol. 2, no. 1, pp. 61–70, 2006. [Online]. Available: http://litis.univ-lehavre.fr/~dutot/biblio/ITSSA2006.pdf

[4] M. Dorigo, V. Maniezzo, and A. Colorni, "The ant system: Optimization by a colony of cooperating agents," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART B*, vol. 26, no. 1, pp. 29–41, 1996.

[5] C. Jordan, "Sur les assemblages des lignes," *J. Reine Angew. Math*, vol. 70, pp. 185–190, 1869.

[6] P. J. Slater, "Centers to centroids in graphs," *Journal of Graph Theory*, vol. 2, pp. 209–222, 1978.

[7] O. Ore, *Theory of graphs*. American Mathematical Society, Providence,, 1962.

[8] F. Harary, *Graph Theory*. Westview Press, Oct. 1969. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20\&path=ASIN/0201410338

[9] M. Truszczynski, "Centers and centroids of unicyclic graphs," *Mathematica Slovaca*, vol. 35, pp. 223–228, 1985.

[10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press, 2009, ch. The Floyd-Warshall algorithm, pp. 558–565.

[11] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: http://www.sciencemag.org/content/286/5439/509.abstract

[12] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," in *Adv. Phys*, 2002, pp. 1079–1187.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

EFFECTS OF TIME-DEPENDENT EDGE DYNAMICS ON PROPERTIES OF CUMULATIVE NETWORKS

Richard O. Legendi and Laszlo Gulyas *†‡

Abstract. Inspecting the dynamics of networks opens a new dimension in understanding the interactions among the components of complex systems. Our goal is to understand the baseline properties to be expected from elementary random changes over time, in order to be able to assess the effects found in longitudinal data.

In our earlier work, we created elementary dynamic models from classic random and preferential networks. Focusing on edge dynamics, we defined several processes changing networks of fixed size. We applied simple rules, including random, preferential or assortative modification of existing edges - or a combination of these. Starting from initial Erdos-Renyi or Barabasi-Albert networks, we examined various basic network properties (e.g., density, clustering, average path length, number of components, degree distribution) of both snapshot and cumulative networks (of various lengths of aggregation time windows). In the current paper, we extend this line of research by applying time-dependent edge creation and deletion algorithms. I.e., we model processes where edge dynamics is defined as a function of time.

Our results provide a baseline for changes to be expected in dynamic networks. Also, they suggest that certain network properties have a strong, non-trivial dependence on the length of the sampling window.

Keywords. dynamic network analysis; cumulative network properties; time-dependent edge dynamics

1 Introduction

Social Network Analysis (SNA) is a field of research of growing importance. However, traditionally, it works with static networks (i.e., with systems where interaction topology is constant over time). This means, that the starting point of traditional SNA is either a snapshot of a network in a specific point of time, or some kind of aggregation (e.g., a collection of longitudinal samples of networks). While there are networks that usually do not change their topology over time (e.g., metabolic networks or blood vessels) and are fundamentally static, many of them change substantially over time, thus both of the above approaches may miss important facts and tendencies of these networks.

Recent papers seem to broaden the analysis of networks changing in time (i.e., dynamic networks) and studies in the subject getting more and more attention.

In this paper we extended our investigation with further elementary dynamic network models as an attempt to study the baseline properties of dynamic networks, with the intention to analyze the structural property changes induced by the effects of time-dependent edge dynamics. Our goal is to understand the inherent properties brought about by the changes to elementary idealized static networks through dynamic processes.

Focusing on edge dynamics, in earlier works [12], we defined three dynamic versions of Erdős-Rényi [4] models, one of which can also interpreted as a dynamic Watts-Strogatz [16] network, and a variety of dynamic Preferential Attachment [3] versions. In the present paper, we extend this line of research by introducing new elementary models that apply time-dependent edge dynamics.

Working with dynamic networks (either collecting longitudinal samples of networks or trying to model the evolution of networks in time), one realises that sampling always involves the act of aggregation (cumulating interactions that happened in a certain time window into one single static network instance). This problem is unavoidable, but poorly studied (See Figure 1.). Therefore, we pay special attention to the length of the cumulation time window and its effect to the aggregated network's properties like density, degree distribution, or clustering.

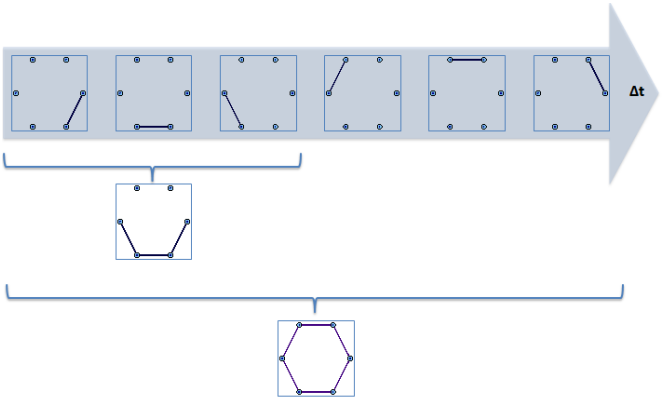


Figure 1: The importance of the sampling window's length

*Richard O. Legendi and Laszlo Gulyas are with AITIA International, Inc. E-mails: rlegendi@aitia.ai, lgulyas@aitia.ai

†Richard O. Legendi is with Eötvös Loránd University, Department of Programming Languages and Compilers, Faculty of Informatics, Hungary. E-mail: legendi@inf.elte.hu

‡Laszlo Gulyas is with Collegium Budapest, Hungary. E-mail: lgulyas@colbud.hu

Our work has a direct implication to the interpretation of SNA, because dissemination of information, innovation, or even epidemic processes may unfold inherently differently if we use different time windows of investigation.

The paper is structured as follows. The next section discusses basic considerations regarding dynamic networks and our research questions relating to them. Section 3 formalizes this by providing basic definitions and introduces the Elementary Dynamic Network models of our studies, and reports our results, discusses related work. The last section outlines directions for future work and concludes the paper.

2 Preliminaries

Dynamic processes in connection with social networks is a popular subject of investigation. This usually means two kind of studies. One class studies dynamics on networks, i.e., the cascading state changes of network nodes (percolation or contagion processes [14], information cascades [15], [1], etc.). The other class is concerned with the dynamics of the network: the structural, topological changes among the nodes and their links that emerge in time [10], [9], [2], [13]. This is the subject of our work in this paper. A third, smaller, class of works also exists that assume that these two processes are not independent: structural changes (i.e., new links) induce state changes and vice versa [6].

2.1 Taxonomy of Network Dynamics

Let us start with making a few elementary considerations. Given a network of $G(V, E)$, structural change means changes in the node set V and/or changes in the links E. That is, the changes in the node set can be classified as:

A. the node set does not change,

B. a period when the node set remains constant,

C. the size of the node set increases, and

D. the size of the node set decreases

In real-life examples, the timescale on which new edges are added to the network is far shorter than how often new nodes are appended to the network (e.g., in case of social networks people tend to form and break up connections, which can be measured in hours or days, while it may take several years for new members to join or leave the network). For this reason, in our first part of our research we focus on constant node set as described by scenario A. We expect that the modification of the node set will not be a major factor determining the properties of social networks.

On the other hand, the number of links is a key property of networks, often expressed as the density (ratio of links per possible links), or as the average number of links per node. In this paper we will work with the former and classify link changing processes as those of

1. constant density,
2. increasing density, and
3. decreasing density.

Naturally, a dynamic process can show a mixture of the above basic scenarios. For example, for a period, it may display an activity that increases the number of nodes, while keeping density constant. Then it may continue by switching to a period when the nodeset remains constant but link density falls, and so forth. In the first part of our studies reported in this paper we will consider only *homogeneous* dynamics: working under the assumption that the different sections can be investigated in separation. As we will see later, certain processes induce dynamics where the measured property (i.e., density) is constant *on average*, while often diverging from the exact value dictated by this value. Therefore, we will allow for minor oscillations.

2.2 What to Study?

Many structural and statistical properties of networks are well known from traditional network analysis. In a dynamic setting, it is but natural to study their temporal evolution. Therefore, in this first analysis of elementary dynamic network models, we focus on the evolution of the number of components (CN), size of the largest component(S_{max}), clustering (C), average degree (D_{avg}), degree distribution and most importantly, average path length (l) and density (d). The reason why l and d might be of special interest is the fact that there is a direct connection between them and the *average* of various centrality measures [7], [8].

- Average degree centrality is directly and trivially connected to density.
- The connection between the sum of betweenness centralities and the average path length: the average betweenness in a network will be about equal to the average path length in the network
- The connection between the sum of closeness centralities and the average path length

Betweenness centrality may also be relevant for several reasons, so we also included the average and maximum betweenness centrality to the list of evaluated metrics (BC_{avg} and BC_{max}, respectively). The results then may be compared with the expectations based on the conclusions enumerated in the listing above (i.e., connection between the centralities and the average path length).

To compare the difference of degree distributions we used the Gini-coefficient [5], which shows the inequality of the distribution ranging from the value of 0 (*expressing total equality*) and the value of 1 (*representing maximal inequality*).

Another interesting aspect is to consider different aggregations of the individual network instances. Many properties of dynamic networks depend on the length of the time window within which network events (e.g., the appearance and disappearance of links) are aggregated. Therefore, we analyze the above measures both for the individual networks generated by the events and for the networks yielded by various time windows of cumulation. In fact, in many cases (when the observation time unit is fine enough) individual *snapshot* networks are not even interesting (i.e., they are very sparse, having almost no links, so it is not even connected, just a composed of a bunch of unconnected components). However, they become interesting when a number of them are *cumulated* together. Naturally, as the cumulation window length grows further, *cumulative networks* become less interesting as they will almost surely eventually approach the complete network (with all possible edges present). One of our main purpose with this research is to identify the time window ranges to use, where the resutlting network features meaningful properties.

The properties of the snapshot and cumulated networks are inherently different: the former is usually an unconnected graph with several components and isolated vertices; the latter is a growing graph that converges to the full graph as $t \to \infty$.

In the following section we describe the natural expectations one can think of about the models, and discuss the observed results.

2.3 Expectations

Based on the fact that the cumulative network accumulates all created edges, but never looses them, the following observations can be made:

1. density (D) grows with cumulation window length (T), and it has direct connection to other properties (e.g. centralities [7], [8])
2. Transitivity (measured by clustering, C) grows with density (D)
3. Average path length (l) drops with density (D)
4. Number of components (CN) drops with density (D)
5. Size of the largest component (S_{max}) grows with density (D)

The aforementioned simple expectations are obvious, however, the transition rate how the properties change is still a question that may show unexpected trends.

3 Formulation

Formally speaking, a dynamic network is a series of graphs. That is, $DN = G_t(V_t, E_t)$, where $E_t \subseteq V_t \times V_t$ ($\forall t \geq 0$). The initial network, G_0, is considered as a parameter of the process. In the current discussion, we assume that the node set is fixed. That is, $V_t = V_0$ for all $t \geq 0$. In addition, we assume that the evolution of the network (i.e., the changing composition of links) can be described as the result of the workings of two processes: i) an edge creation process and ii) an edge deletion process, that both identify a set of edges, $CE_t \subseteq (V_t \times V_t) \setminus E_t$ and $DE_t \subseteq E_t$, respectively. Given these, the edge dynamics of the network can be conceptualized as:

$$E_{t+1} = (E_t \setminus DE_t) \cup CE_t$$

Let us introduce the following notations. The number of nodes in the network will be denoted by $N = N_t$. The density of the network is the ratio of the number of existing edges over the maximum possible number of edges in the network of a given size. Assuming an undirected network where loops (self-links) and hyperlinks are not permitted, the density is given as

$$d_t = \frac{2\,|E_t|}{N(N-1)} \approx \frac{|E_t|}{N^2}$$

As we stated in section 2, we follow a *constant density assumption* during our current research, formally:

$$\exists \varepsilon > 0 : \forall t \in [0, T] : |d_t - d_{t-1}| \leq \varepsilon$$

3.1 Envelopes of Dynamic Networks

For any model of dynamic networks, there is a number of various networks of interest. The most immediate ones to consider are the momentary *snapshots* of the dynamic process, i.e., the network G_t that exists at any given time t. (Conceptually, these may also be aggregations of individual network events, like changes in the relationship between a pair of nodes, but these events occur so close to each other in time that we cannot distinguish the time of their occurence.) In practice, however, we often deal with the aggregation of these networks. In the following, we study the properties of networks that are generated by the *cumulation* of a number of G_t's. For reasons of simplicity, in the following we only deal with cumulative series starting at $t = 0$ (this can be done without the loss of generality). Thus, for the purposes of this paper, we define a *cumulative network* as

$$G_T = (\bigcup_{t=0}^{T} V_t, \bigcup_{t=0}^{T} E_t) \text{ for } T \geq 0.$$

3.2 Elementary Models of Dynamic Networks

In our previous studies [12] we created elementary models of dynamic networks based on the classic network models for static networks. The three classic models we had our eyes on were the Erdős-Rényi (ER) model of *random* density networks, the *small–world* network of Watts and Strogatz (WS) and the *preferential attachment* model of Albert and Barabási for *scale-free netwroks* (PA).

In our current study, we extended these models with two new ER models, where the presence of edges is directly connected to the simulation time. The following sections describe the evaluated models with the analysis and conjectures for the cumulative networks.

3.3 ER4

We start from an initial graph G_0 created by the static Erdős-Rényi model with density of p_0. Our previous approaches to edge addition and deletion processes was independent (e.g., random, preferentioal or based on a probability). In our current study, we extend we extend our investigations by time-based modification processes: we assing the *time of presence* of an edge by choosing a random number from the interval $[0, T]$ uniformly, where T is the simulation time. The edge is present in the snapshot network only in the timestep of t (cf. Lee et al. [11])). So basically it is a model where random links are created with a random entry time and lifetime.

The basic idea behind the model to get a general idea about what properties does the snapshot and cumulative network has when we know that we get an Erdős-Rényi model as an aggregated network.

One thing to note is that in our previous models, the density of the *snapshot* network was adjustable. For the ER4 model, the density means the density of the *prototype* network (which is equal to the cumulative network at the end of the simulation time, T). However, the expected density of the snapshot network can be derived by scaling it to the different G_t instances: $d_{G_t} = d_0/T$.

3.3.1 Experiments and Results

We performed the simulation runs with the parameter settings for 100 nodes, an initial density for the prototype network $d = 0.02$. We set the edge lifetime to $s = 5, 10, 15, 20, 25$, setting the final simulation time to $T = 100$. To validate the results, we performed 125 simulation steps in practice. We have two seed parameters for the random generators (initialization and addition seeds). We performed the whole simulation with 3 different values for each seed parameter (9 runs in total), the results here are the averaged values of the results we obtained.

Density Density shows monolithic increase with the same for the cumulative networks reaching very similar values in the end. That is no surprise, since the edges are distributed uniformly in the interval $[0, T]$. In the beginning, this trend is followed by the snapshot networks to the point when time reaches the lifetime of the edges (it is shown for the lifetime is 20 and 25 cases). The edge lifetime has a direct influence on the density level, as expected: as we increase the lifetime the density increases linearly for the snapshot networks. At the end all of the snapshot networks become empty networks. The average degree shows exactly the same trend (since its value can be derived from density directly).

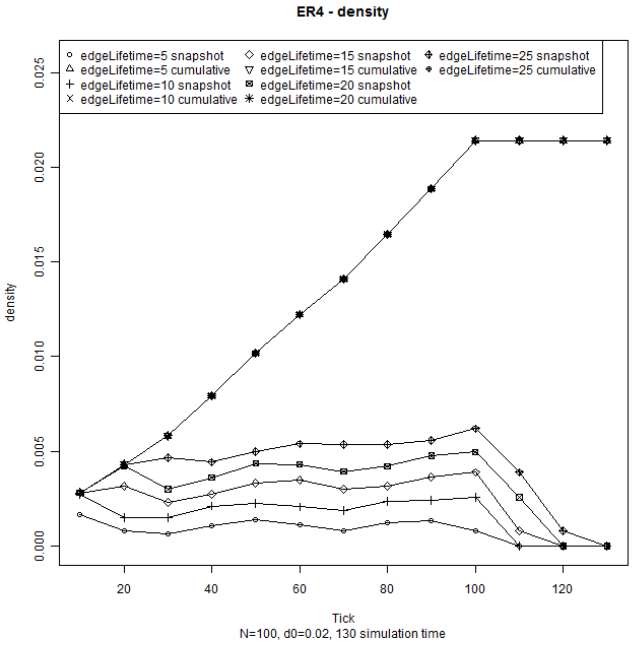


Figure 2: ER4 Density

Betweenness The betweenness values for the snapshot and cumulative network show significant differences. The average and maximum betwennes values show the same trends (but of course, on a different scale).

The betweenness value of the cumulative network shows some interesting phenomena to note. First, there is a very sharp transition between the 70th and 80th tick, the betweenness centrality grows by more than three times. Second, it is not a monotonic increasement, it continues increasing up to a point but then the average betweenness value falls back to lover rate (which then remains constant since there is no new edge added in the last 25 turns). That may happen in cases when new edges introduce several alternate paths between two components of the network.

The snapshot network shows only minimal differences, since the density of the network is very low as shown before. The maximum betweenness shows exactly the same trends as the average betweenness.

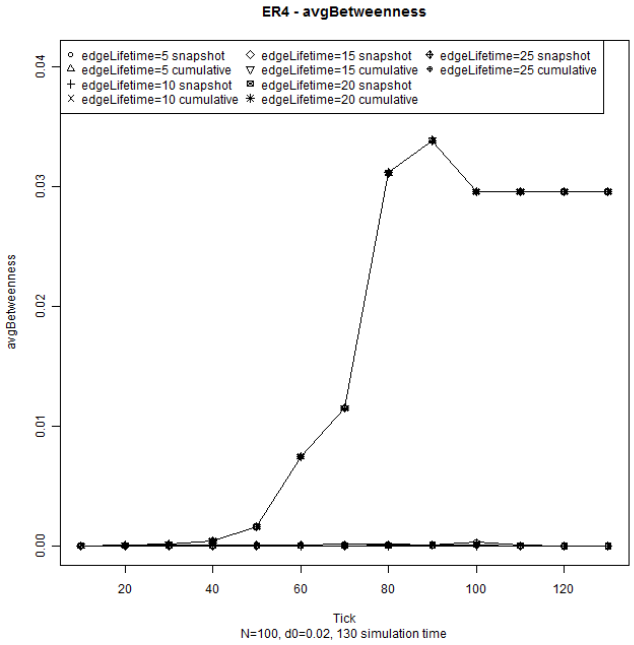


Figure 3: ER4 Average Betweenness Centrality

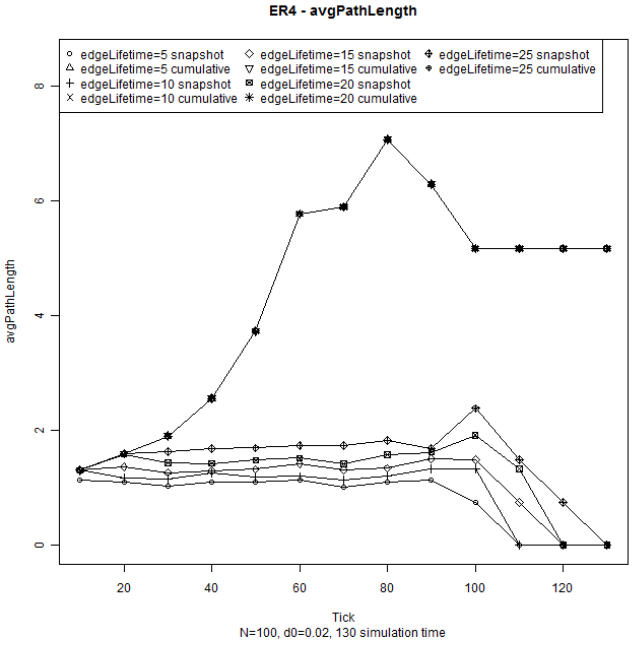


Figure 4: ER4 Average Path Length

Average Path Length The average path length reflects the same sort of change: the cumulative network has an outstanding value at the 80th tick, and starts decreasing.

Clustering The clustering shows intresting values as well. For the cumulative network, the values shows similiar trends to the average betweenness: a sharp transition may be observed a bit sooner in the clustering (between the 50th and 60th time step), and the fall back of the value may also be seen after 70th tick. Another interesting phenomenon here is that the clustering value of the snapshot network may exceed the clustering of the cumulative network by several times the in some of the cases. This situation have occured with even two of the models around the same time steps, these were the ones with the higher edge lifetime values (20 and 25).

Components The number of components show expected results. With the given number of nodes and initial density value, the fractional size of the maximum component of the cumulative network approaches a high value at the end of a simulation. The number of components has a linear connection to the edge lifetime, similarly to the density values.

As a general note, we can conclude that the results for the cumulative networks are almost identical.

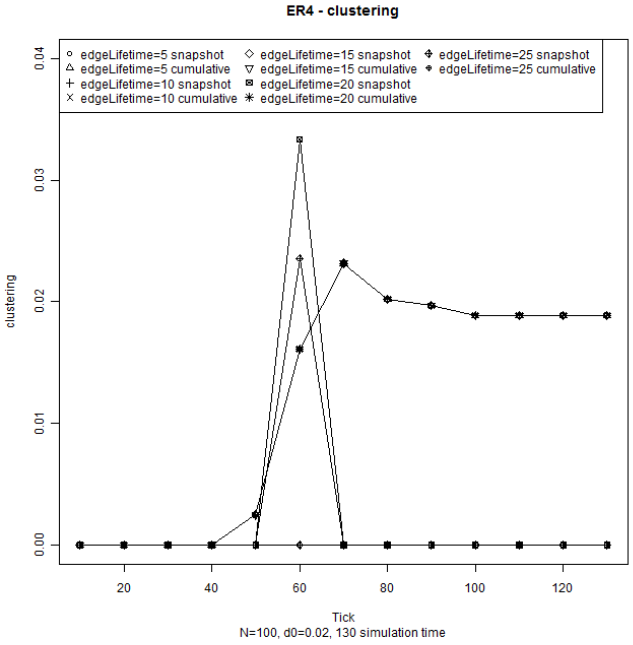


Figure 5: ER4 Clustering

3.4 ER5

This model is based on ER4, but with a minor modification. Edges in this model appear periodically in each

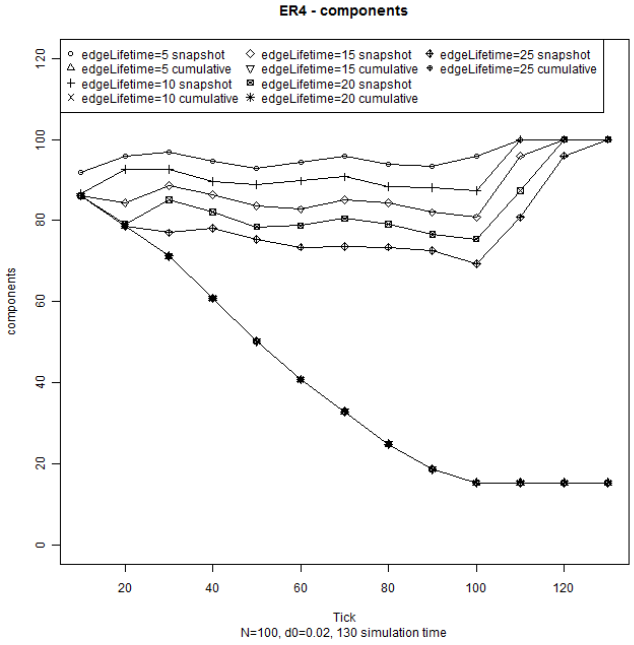


Figure 6: ER4 Components

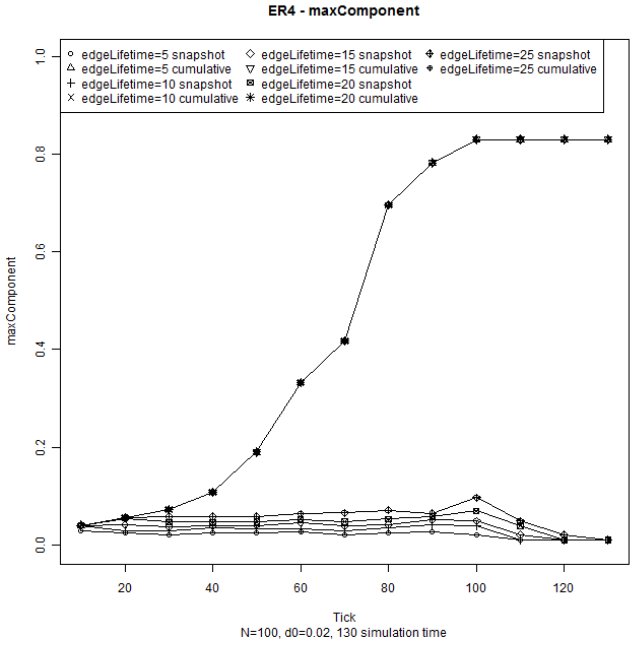


Figure 7: ER4 Fractional Size of the Maximal Component

$k * s$ timestep: random links are created with a random appearance frequency and lifetime (cf. Lee et al. [11]).

3.4.1 Experiments and Results

We performed the simulation runs with the parameter settings for 100 nodes, an initial density for the prototype network $d = 0.02$, setting the final simulation time to $T = 100$. We have two seed parameters for the random generators (for the initialization and addition). We performed the whole simulation with 3 different values for each seed parameter (9 runs in total), the summarized averaged values of the results we obtained may be found in Table 3.4.1.

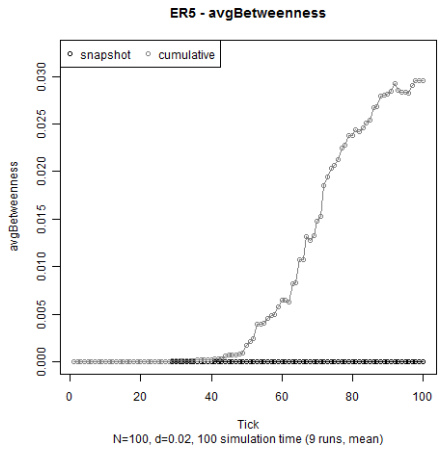
An interesting phenomena is that the cumulative network properties show smaller increasements up to the half of the simulation around the 50th tick, and shows significant increasements afterwards. The average and maximum betweenness centrality remains a very small value but starts growing in a very quick way. Density and the average degree grows in a linearly, it is because edges and their frequency is distributed uniformly between 0 and T. This is also the reason why the number of components decrease in a liear way as well. The average path length reflects a very sharp transition in the first few simulation steps, and starts growing to a level until the giant component appears in the network, and starts decreasing with new and new edges. The number of components and the fractional size of the maximal component show us the current size of the giant component.

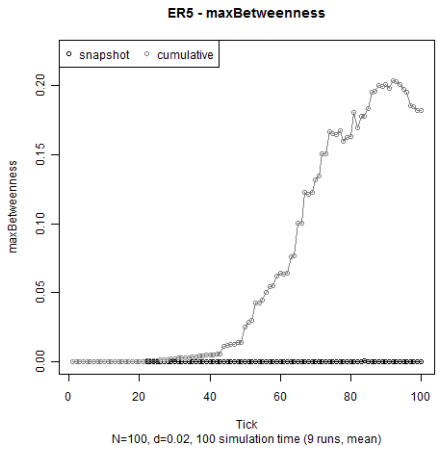
An extremely volatile and sensitive property for the cumulative network is the clustering which is also oscillating wildly. It has absolutely no meaning until the 50th simulation step, but increases within a few steps. Its value is not even stable, it may change within in every few steps.

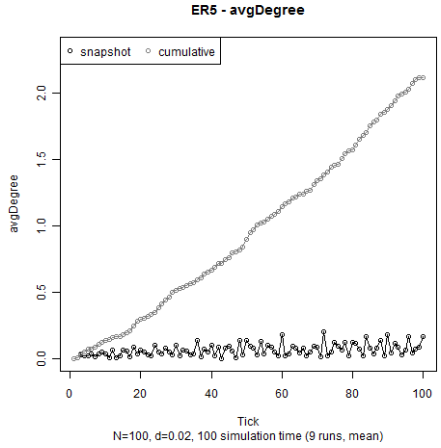
4 Conclusion and Future Work

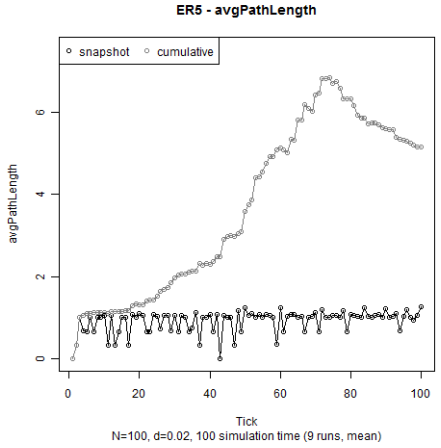
In this work we defined elementary dynamic network models as an attempt to study the baseline properties (number and size of components, clustering, degree distributions, average path length and density) of networks changing in time. Our goal was to understand the inherent properties brought about by the changes themselves to elementary idealized static networks through dynamic processes. Focusing on edge dynamics, we defined several versions Erdős-Rényi, Watts-Strogatz and a variety of dynamic Preferential Attachment versions.

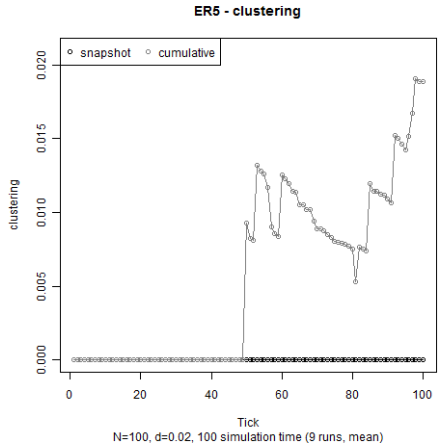
We found that, in case of these elementary models of dynamic networks, several network statistics are very sensitive to the length of the cumulation window. In particular, the density of the cumulative network changes superlinearly, especially in the relevant region of small to medium values. This is very important, since density has a great impact on the value of various other network measures.

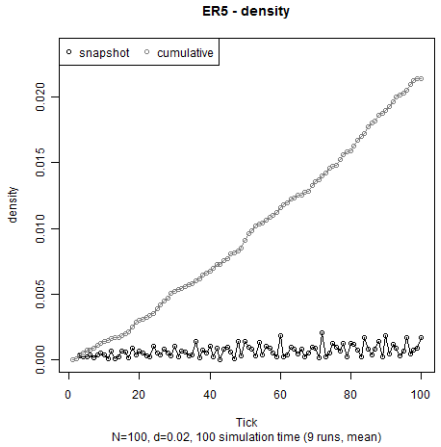

ER5 - avgBetweenness
snapshot
cumulative
avgBetweenness
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

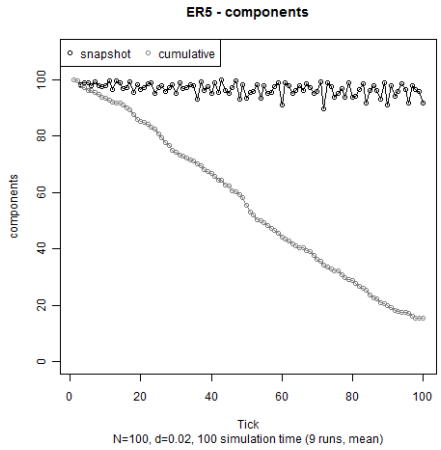

ER5 - maxBetweenness
snapshot
cumulative
maxBetweenness
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

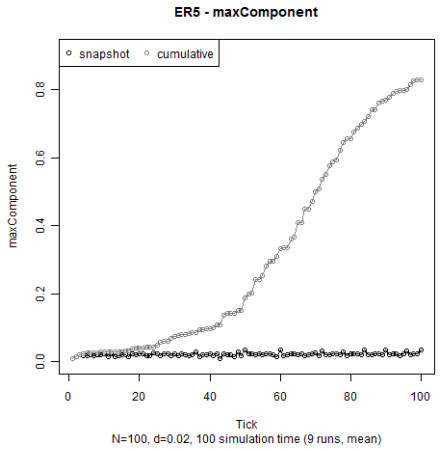

ER5 - avgDegree
snapshot
cumulative
avgDegree
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - avgPathLength
snapshot
cumulative
avgPathLength
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - clustering
snapshot
cumulative
clustering
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - density
snapshot
cumulative
density
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - components
snapshot
cumulative
components
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)


ER5 - maxComponent
snapshot
cumulative
maxComponent
Tick
N=100, d=0.02, 100 simulation time (9 runs, mean)

The result that the length of the accumulation window makes a difference, seem to be trivial. On the other hand, this problem is not even mentioned in most of the works.

Our results also shed light on the behavior of the various models, with regard to how much history (i.e., initial network structure) matters. In some cases the network dynamics generated by these elementary models eradicate history instantly. In some other cases, history had an impact on both the snapshot and the cumulative network for an extended period of time.

4.1 Future Work

In future works we plan to extend our studies to include experiments with models that have richer and more realistic edge dynamics. These will include the following, still simple, models of elementary dynamic networks.

Double Preferential Choosing both ends of new links preferenitally.

Assortative Mixing Add links between a random node and another selected randomly with probability to the end node degree's similarity to that of the starting node.

Lonely Node Preferentials Only disconnected nodes create new links (approximation of newcomers in original PA).

To derive the formulas for the probability distributions of the metrics (or for the expected value of the metrics) along time would be another research direction, since this would allow for a precise comparison between the different models and networks, and for a precise comparison of the usage of different window lengths within the same dynamic model.

Acknowledgements

This research was partially supported by the Hungarian Government (KMOP-1.1.2-08/1-2008-0002) and the European Union's Seventh Framework Programme: DynaNets, FET-Open project no. FET-233847 (`http://www.dynanets.org`). The supports are gratefully acknowledged.

References

[1] Andrea Apolloni and Floriana Gargiulo. Diffusion processes through social groups' dynamics. *1103.0973*, March 2011.

[2] Andrea Apolloni, Achla Marathe, and Zhengzheng Pan. A longitudinal view of the relationship between social marginalization and obesity. In John Salerno, Shanchieh Jay Yang, Dana Nau, and Sun-Ki Chai, editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 6589, pages 61–68. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509 –512, October 1999.

[4] P. Erdős and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[5] C. W. Gini. Variability and mutability, contribution to the study of statistical distributions and relations. *Studi Economico-Giuridici della R. Universita de Cagliari*, 1912. Reviewed in: Light, R.J., Margolin, B.H.: An Analysis of Variance for Categorical Data - J. American Statistical Association, Vol. 66 pp. 534-544 (1971).

[6] Laszlo Gulyas and Elenna R Dugundji. Emergent opinion dynamics on endogenous networks. http://adsabs.harvard.edu/abs/2006physics..10125G, October 2006.

[7] Laszlo et al. Gulyas. Betweenness centrality as a function of network density and size. *(under publication)*.

[8] László Gulyás, Gábor Horváth, Tamás Cséri, Zalán Szakolczi, and George Kampis. Betweenness centrality dynamics in networks of changing density. *Presented at the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010)*, 2010.

[9] L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Rava , C. Rizzo, and A. E Tozzi. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *1104.2515*, April 2011. PLoS ONE 6(2): e17144 (2011).

[10] Lorenzo Isella, Juliette Stehle, Alain Barrat, Ciro Cattuto, Jean-Francois Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *1006.1260*, June 2010. J. Theor. Biol. 271 (2011) 166-180.

[11] Sungmin Lee, Luis E. C Rocha, Fredrik Liljeros, and Petter Holme. Exploiting temporal network structures of human interaction to effectively immunize populations. *1011.3928*, November 2010.

[12] Richard O. Legendi and Laszlo Gulyas. Cumulative properties of dynamic social networks. 2011. Under publication.

[13] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1, March 2007. ACM ID: 1217301.

[14] Juliette Stehle, Alain Barrat, and Ginestra Bianconi. Dynamical and bursty interactions in social networks. *1002.4109*, February 2010. Phys. Rev. E 81, 035101(R) (2010).

[15] Duncan J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766 –5771, April 2002.

[16] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, June 1998.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

AN ESTIMATION OF THE SHORTEST AND LARGEST AVERAGE PATH LENGTH IN GRAPHS OF GIVEN DENSITY

László Gulyás, Gábor Horváth, Tamás Cséri and George Kampis *†‡

Abstract. Many real world networks (graphs) are observed to be 'small worlds', i.e., the average path length among nodes is small. On the other hand, it is somewhat unclear what other average path length values networks can produce. In particular, it is not known what the *maximum* and the *minimum* average path length values are. In this paper we provide a lower estimation for the shortest average path length (ℓ) values in connected networks, and the largest possible average path length values in networks with given size and density. To the latter end, we construct a special family of graphs and calculate their average path lengths. We also demonstrate the correctness of our estimation by simulations.

Keywords. Complex networks, average path length, density dependence, lower and upper bound, simulation.

1 Introduction

Many real world networks (graphs) are observed to be 'small worlds', i.e., the average path length (ℓ) among nodes is small. While there exists a widely accepted formulation of this property (i.e., ℓ scales with the logarithm of the number of nodes (N)), the intuitive meaning of the small world property is that the average path length is shorter "than *expected*".

On the other hand, it is somewhat unclear what average path length to expect. Observed networks tend to have short ℓ and many algorithmically constructed network models also share this property. In particular, Erdős-Rényi (ER) random networks almost always have short path lengths. ER networks are interesting for several reasons. First, they are constructed with a minimalistic set of assumptions: given N nodes, each possible link is present with a certain uniform p probability. Second, their minimalistic random construction makes them an approximation of a random sample from the set of all possible graphs with the given number of nodes and the given density.

However, if random or randomly sampled graphs have short ℓ, does this mean that all graphs share this property? No. It is easy to create counter examples, i.e., graphs with large average path lengths. Typical examples include chains or regular lattices. The question is, however, how general these counter examples are? How common they are among the set of all possible graphs? Also, chains or regular lattices assume a certain number of links for a given N, so they may not be constructed for a given N and density (d, the ratio of links present among all the possible links), while the average path length naturally depends on both of parameters. This makes the interpretation of the counter examples somewhat problematic.

Unfortunately, the simple enumeration of all possible graphs with a given N and d is prevented by their astronomical numbers. The following formula gives this number for undirected graphs (counting all isomorphs equally).

$$|G(N,d)| = \left(\begin{array}{c} \frac{N(N-1)}{2} \\ \\ \frac{dN(N-1)}{2} \end{array} \right) \quad (1)$$

This simple formula worths a moment of further consideration. While networks with $N = 100$ nodes are typically considered unrealistically small and a density of $d = 0.0001$ does not describe an over-populated graph, the number of possible networks with these parameters is 1.75876×10^{16}.

Given these astronomical numbers, correct sampling would be essential to determine how average path lengths are distributed among the possible graphs. While one of the definitions of the ER networks is exactly this ([10]), it is also known that the above described mechanism to generate ER graphs yields a uniform sample if and only if the network is unconnected [10].

Following these considerations, in this paper we provide a lower estimation for the shortest and the largest possible average path length (ℓ) in networks of given N and d. To the latter, we construct a special family of graphs and calculate their average path lengths.

The paper is structured as follows. The next section provides the estimation of shortest average path length in connected networks. Section 3 introduces the family of graphs constructed to estimate the maximum average path length, while Section 4 provides analytical results

*The authors are with Eötvös University, Budapest, Hungary
†Gulyás, Horváth and Kampis are also with Collegium Budapest, Institute for Advanced Study, Budapest, Hungary
‡Correspondance to {lgulyas, gkampis}@colbud.hu

for the average path length of these graphs. This is followed by a comparison of these results to the average path lengths of real-world graphs and various classic network models. The last section outlines future works and concludes the paper.

2 Estimation of the Shortest Average Path Length in Connected Graphs

The average path length of full graphs is obviously 1. This is because every node is connected by every other one, so every node can reach all of others on a 1-length path. If we remove only one edge from the graph we decrease the density of the graph by $\frac{2}{N(N-1)}$. The path lengths change in the following way: $\frac{N(N-1)}{2}-1$ path lengths will remain 1 and a path between a pair of nodes increases to 2. So ℓ is increased to:

$$\frac{\frac{N(N-1)}{2}-1+2}{\frac{N(N-1)}{2}} = \frac{\frac{N(N-1)}{2}+1}{\frac{N(N-1)}{2}} = 1+\frac{2}{N(N-1)} \quad (2)$$

This means that ℓ is increased by $1-d$. Repeating the above argument until every node has only $N-1$ edge, the increase of ℓ in this range will be linear.

Suppose that the network has only L edges. This means that there are L 1-length paths. So $d\frac{N(N-1)}{2}$ paths has a length of 1 and we do not know the length of the remaining paths. But we know that their length must be at least 2, because they cannot be 0 or 1. This yields an estimation of the minimum ℓ of a graph with given size and density:

$$\frac{\left[d\frac{N(N-1)}{2}\right]\cdot 1+\left[(1-d)\frac{N(N-1)}{2}\right]\cdot 2}{\frac{N(N-1)}{2}} \quad (3)$$

After some simplification this yields:

$$d+(1-d)\cdot 2 = 2-d \quad (4)$$

This is the shortest possible average path length a connected graph with given size and density may have.

3 A Family of Graphs with Large Average Path Length

Before constructing our graphs, let us make a few observations. It is common wisdom that large ℓ is to be expected in 'geographical' networks, i.e., where the nodes have same phisical or topological locations and links connect closeby nodes only. Thus, chains (i.e., networks where nodes are lined up along a line and links connect nearest neighbors only) would be an ideal candidate network. The problem is, chains can only accommodate $N-1$ of links. If density dictates more than this, we need another solution. Moreover, the addition of each new link decreases the the average path length, potentially at several points. Trivially, the new link creates a path of length 1 between its two endpoints. In addition, it potentially shortens all paths going to one of the link's nodes. Furthermore, if the new link created a new shortcut, paths among two different nodes may also be shortened.

In our construction, we want to preserve as long a chain as possible. Thus, our intuition is that it is best to add additional links in a single, densely connected 'blob' (cluster) at one end of the chain. Ideally, we create a fully connected blob (a clique) and preserve the rest of the links and nodes for the chain (or tail). (See Figure 1.)

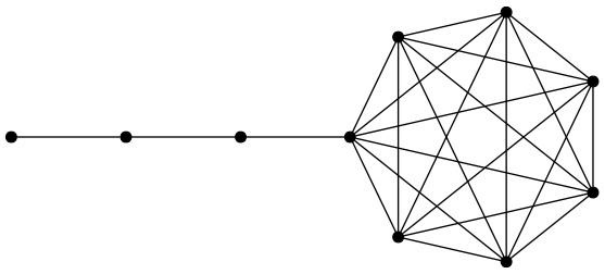


Figure 1: Example from the family of graphs constructed for maximum average path length

Naturally, the more links we need to place in the blob, the more nodes are needed. At the extreme, when all $N(N-1)/2$ links are present, no tail is possible. It is also clear that the length of the tail (z) is a non-continuous function of the link density. For example, if we have more than $(N-1)(N-2)/2+1$ edges then all nodes must belong to the blob. (For the sake of simplicity, in the following we will always assume that the blob is fully connected.)

The first step in creating a graph of this family for a given N and d is to determine the length of the tail (z). By definition, nodes in the tail may have at most 1 connection to nodes in the blob. (I.e., for clarity we define the 'last' node of the tail to be the one *not* in the blob.) Thus we have z nodes in the tail and $N-z$ in the blob, and since the blob is assumed to be fully connected, we get the following equation for the edges:

$$D = d\frac{N(N-1)}{2} = z+\frac{(N-z)(N-z-1)}{2}, \quad (5)$$

which yields the following formula for z

$$z = \frac{\pm\sqrt{8D-8N+9}+2N-3}{2}. \quad (6)$$

A necessary condition for real roots is $8D-8N+9 \geq 0 \Rightarrow D+8/9 \geq N$. This essentially corresponds to the requirement that the graph should have enough edges to be connected, without which our construction will not work.

In order for z to be meaningful, it is also necessary that $z \leq N$, which means that

$$\frac{\pm\sqrt{8D-8N+9}+2N-3}{2} \leq N$$

$$\frac{\pm\sqrt{8D-8N+9}}{2}+N-3/2 \leq N$$

$$\frac{\pm\sqrt{8D-8N+9}}{2} \leq 3/2$$

Therefore, in the following we will always work with the smaller root.

4 Calculation of the Average Path Length

In this section we calculate the average path length in the above constructed family of graphs. First, we will calculate the sum of all shortest paths, then divide it by $N(N-1)$ (as, for the sake of simplicity, we will count all paths in both directions).

There are three kind of nodes in our graph: i) the ones in the tail, ii) the one bridging the tail to the rest of the blob, and iii) the ones in the blob. Let's consider the i^{th} node in the tail. The length of the paths connecting it to the $i-1$ nodes preceeding it in the tail is $i-j$, where j is the position of the other node. Similarly, the length of the paths to the $z-i$ nodes following it in the tail is $j-i$. The path to the bridge node is of length $z-i+1$, while the $N-z-1$ nodes in the blob are reachable in $z-i+2$ steps. On the other hand, the bridge node needs $z-i+1$ steps to the i^{th} node in the tail, while 1 step to each of the $N-z-1$ nodes in the blob. For these latter, the connection to the rest of the blob, plus the bridge ($N-z-1$ nodes altogether) is of length 1, while to the nodes in the tail it takes $z-i+2$ steps. These yield the following formula.

$$\mathrm{SUM}(N,z)=\sum_{i=1}^{z}\sum_{j=1}^{i-1}(i-j)+\sum_{i=1}^{z}\sum_{j=i+1}^{z}(j-i)+\sum_{i=1}^{z}(z-i+1)+$$

$$+\sum_{i=1}^{z}(N-z-1)(z-i+2)+\sum_{i=1}^{z}(z-i+1)+(N-z-1)+$$

$$+(N-z-1)\left(N-z-1+\sum_{i=1}^{z}(z-i+2)\right)=$$

$$=N^2+Nz^2+Nz-N-\frac{2z^3}{3}-2z^2-\frac{4z}{3}$$

Substituting z with the smaller value in (6) we get:

$$\begin{aligned}\mathrm{SUM}(N,D) &= \frac{-N^3-6\,N^2}{3}+\\ &\frac{\sqrt{-8\,N+8\,D+9}\,(2\,N-2\,D-2)}{3}+\\ &\frac{(6\,D+13)\,N-6\,D-6}{3}\end{aligned} \tag{7}$$

or

$$\begin{aligned}\mathrm{SUM}(N,d) &= \left(d\,N^2+(-d-2)\,N+2\right)\cdot\\ &\frac{\sqrt{4\,d\,N^2+(-4\,d-8)\,N+9}}{3}+\\ &\frac{(1-3\,d)\,N^3+(6\,d+6)\,N^2}{3}+\\ &\frac{(-3\,d-13)\,N+6}{3}\end{aligned} \tag{8}$$

.

Therefore the path length can be given as

$$\begin{aligned}l(N,D) &= \frac{-N^3-6\,N^2}{3\,N^2-3\,N}+\\ &\frac{\sqrt{-8\,N+8\,D+9}\,(2\,N-2\,D-2)}{3\,N^2-3\,N}+\\ &\frac{(6\,D+13)\,N-6\,D-6}{3\,N^2-3\,N}\end{aligned} \tag{9}$$

or

$$\begin{aligned}l(N,d) &= \left(d\,N^2+(-d-2)\,N+2\right)\cdot\\ &\frac{\sqrt{4\,d\,N^2+(-4\,d-8)\,N+9}}{3\,N^2-3\,N}+\\ &\frac{(1-3\,d)\,N^3+(6\,d+6)\,N^2}{3\,N^2-3\,N}+\\ &\frac{(-3\,d-13)\,N+6}{3\,N^2-3\,N}\end{aligned} \tag{10}$$

.

As an illustration Figure 2 plots this function for various values of N accross the density spectrum.

5 Numerical Tests of the Estimations

To demonstrate the correctness of our estimation, we created ER and BA networks in the entire density range for different sizes. We also enumerated all possible 10 node graphs with $d = 0.2$. As shown on Figure 3 both estimations are matching with the ℓ of generated graphs. Figure (4) places various empirical networks on the plot as well. It is worth noting that for most densities networks are very close to the theoretical minimum. This is especially

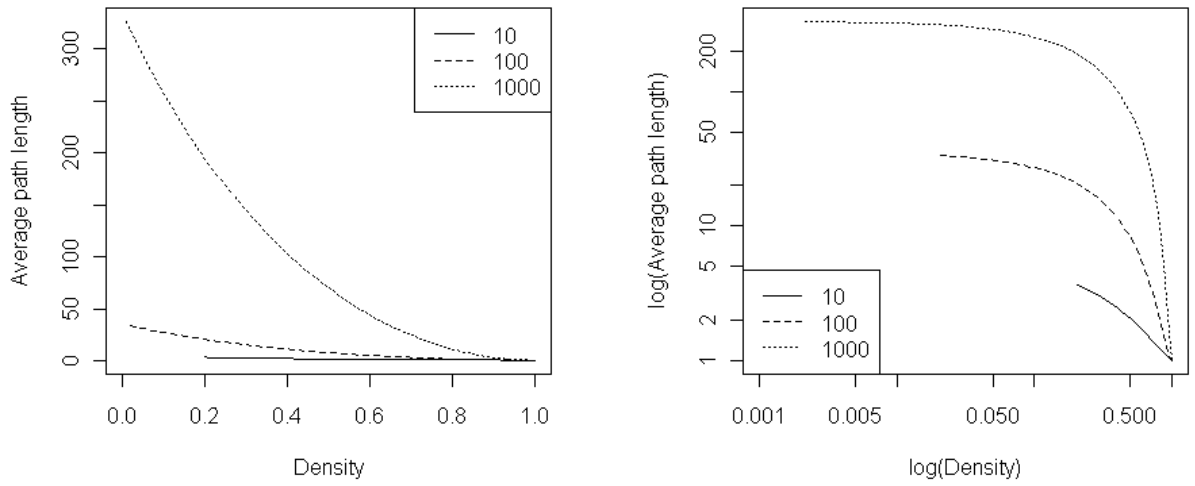


Figure 2: The lower estimation of the maximum average path length value for various network sizes across the density spectrum. Note that network size determines a minimum meaningful density that drops with N.

true for foodweb networks (that tend to be small that implies a relatively high density as well).

For completeness, we note that our estimation of shortest ℓ correct only for connected graphs. If a graph is unconnected it is possible to have shorter ℓ.

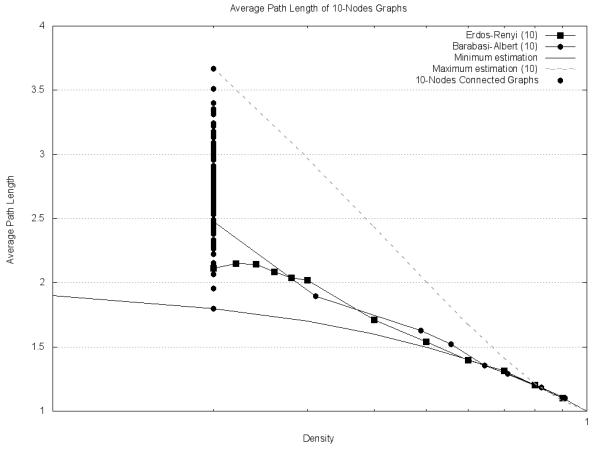


Figure 3: Average path length of 10-node graphs

We generated all the 10-node graphs with density 0.2. The following figure shows the simulation result of the 10-node connected graphs. (Fig. 5).

6 Conclusions

In this paper we provided minimum and maximum estimations for the average path length in complex networks of given size as a function of their density. First, we gave an exact estimation of shortest average path length of graphs. Then we created a special family of graphs in order to give a lower estimation for the largest ℓ of graphs with given size and density. This result is important because earlier works had valid results only for certain graph types and typically only in relatively short intervals of density. We demonstrated the correctness of our estimation by generating all possible graphs with given size and density.

We also provided a comparison of these estimations to the average path length values measured in empirical networks and in classic network models like the Erdos-Renyi exponential random graph and the Barabasi-Albert preferential attachment model. We observed that most of these measured values are close to the minimum theoretical estimation. In addition, we also show that, in contrast to the previous observation, a full enumeration of all possible graphs of a given size and density are spread 'evenly' between the theoretical minimum and maximum.

7 ACKNOWLEDGEMENTS

This research was partially supported by the Hungarian Government (Anyos Jedlik programme managed by the National Office for Research and Technology: TexTrend project (www.textrend.org), contract no. NKFP_07_A2 (2007) TEXTREND) and the European Union's Seventh Framework Programme: DynaNets, FET-Open project no. FET-233847 (www.dynanets.org). The supports are gratefully acknowledged.

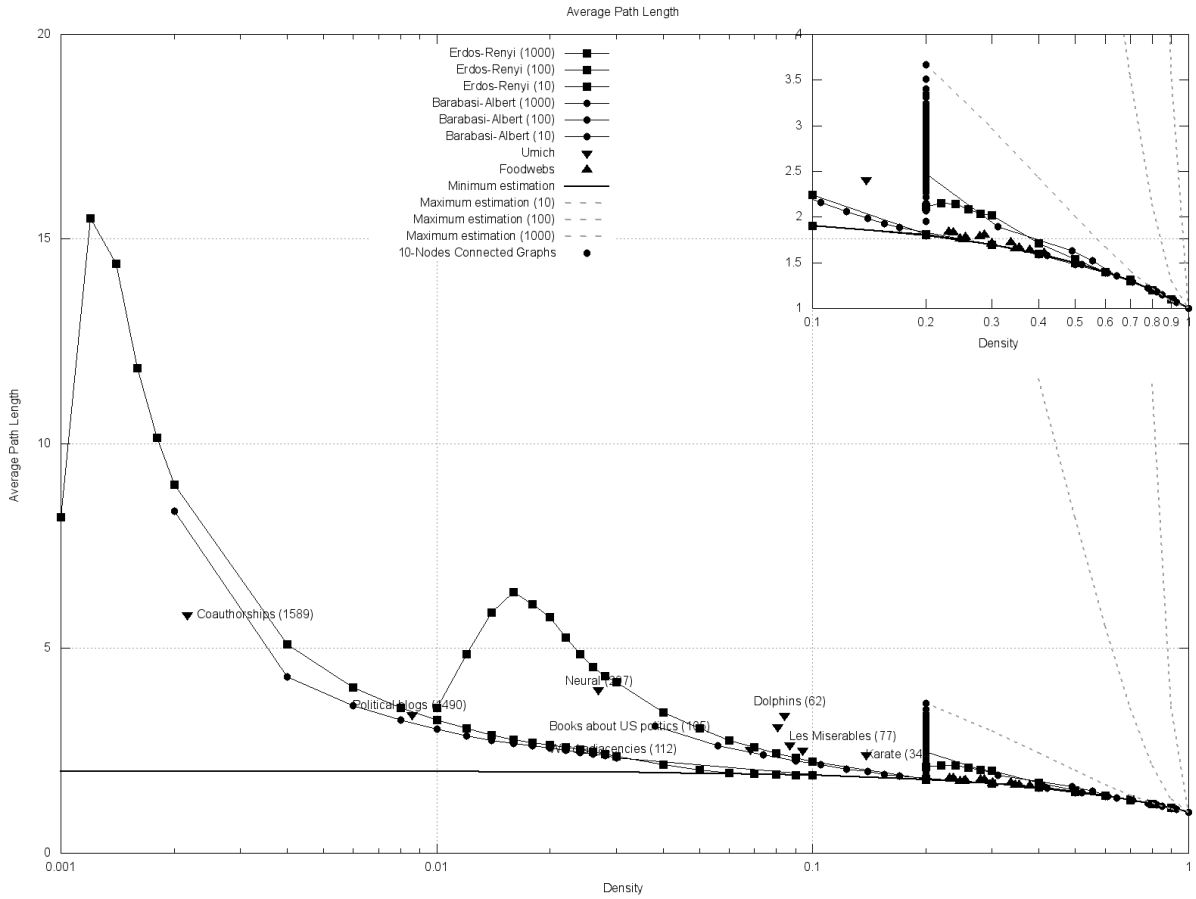


Figure 4: Average path lengths of various networks (lin-log plot)

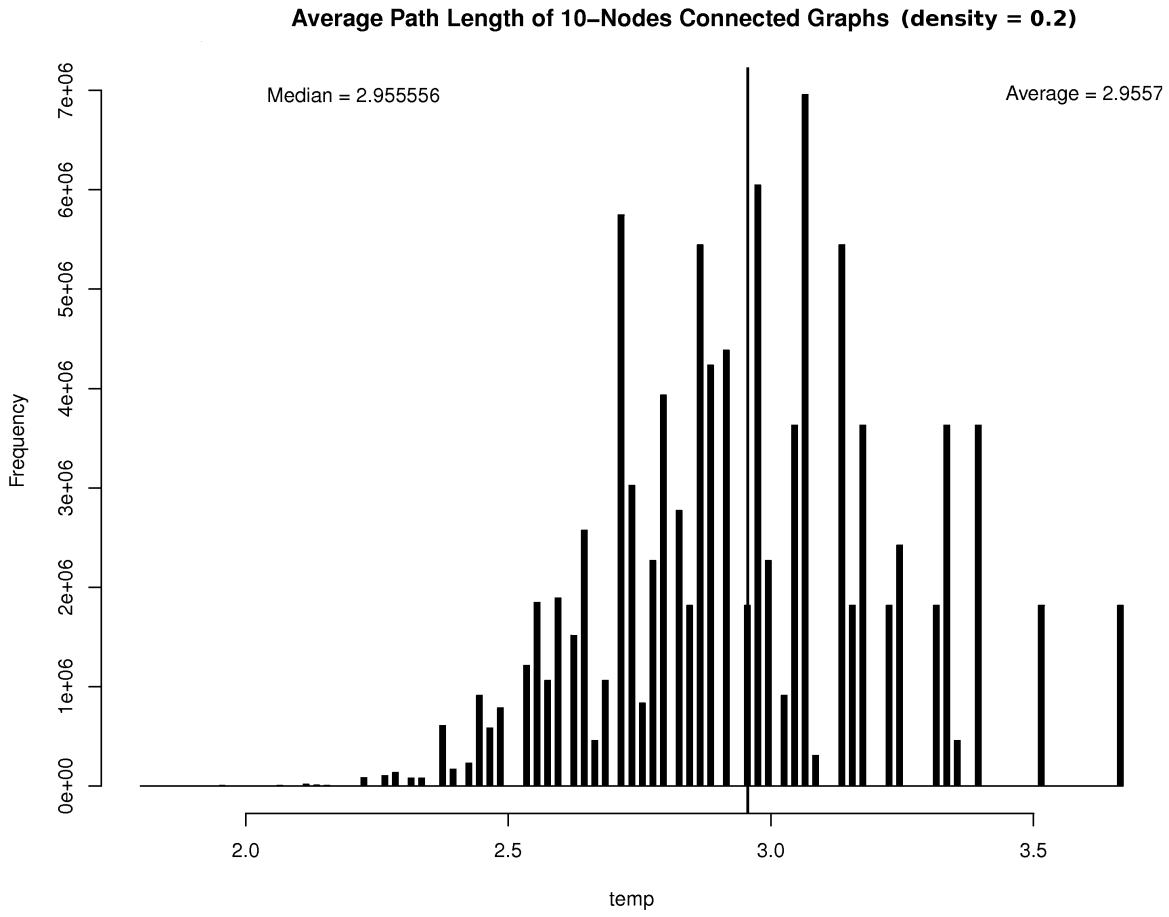


Figure 5: Distribution of average path length in 10-node graphs with d=0.2

References

[1] Agata Fronczak, Piotr Fronczak, Janusz A. Holyst: "Average path length in random networks", *Physical Review*, 17 November 2004, Warsaw University of Technology

[2] Albert, R., Jeong, H., Barabási, A.-L.: "Error and attack tolerance of complex networks", *Nature* 406, 378-382 (27 July 2000)

[3] Kampis, G., Gulyás, L., Szászi, Z., Szakolczi, Z.: "Dynamic Social Networks and the Textrend / CIShell Framework", *Applied Social Network Analysis Conference*, 27-28 August 2009, University of Zurich and ETH.

[4] Girvan M. and Newman M. E. J.: "Community structure in social and biological networks", *Proc. Natl. Acad. Sci. USA* 99, 78217826 (2002)

[5] Granovetter, M.: "The Strength of Weak Ties", *American Journal of Sociology* 78 (6): 13601380, May 1973

[6] Gulyás, L., Kampis, G., Szakolczi, Z., Jordán, F.: "Finding Trendsetters in Longitudinal Networks", *COINS conference*, http://www.coins2009.com/

[7] Kampis, G., Gulyás, L., Szakolczi, Z., Szászi, Z. and Soós, S.: "Mapping Science: Bio-Inspired Methods for Dynamic Network Analysis", *Modeling Science Conference*, Oktober 6-9, Amsterdam, http://modelling-science.simshelf.virtualknowledgestudio.nl/

[8] Sloot, P.M.A., Ivanov, S., Boukhanovsky, A., van de Vijver, D., Boucher, C.: "Stochastic simulation of HIV population dynamics through complex network modeling", *Int. J. Comput. Math.*, 85, 1175-1187, 2008.

[9] Vasas, V., Jordán, F.: "Topological keystone species in ecological interaction networks: considering link quality and non-trophic effects", *Ecological Modelling*, 196:365-378, 2006.

[10] Erdős, P.; Rényi, A.: "On Random Graphs. I.", *Publicationes Mathematicae* 6: 290297, 1959.

[11] Barabási, A.-L., Albert, R.: "Emergence of scaling in random networks", *Science* 286: 509512, 1999.

8 Appendix: Generating all possible graphs with given size and density

We mentioned above in Section 1 that generating all possible graphs with given size and density gives astronomical numbers. Let N the number of vertices in the graph and let d the density of the graph. So the graph has $d\frac{N(N-1)}{2}$ edges (working with undirected graphs). In an undirected graph the largest number of edges can be $\frac{N(N-1)}{2}$. This means if we want to generate a graph with N vertices and d density, we have to select $d\frac{N(N-1)}{2}$ edges from $\frac{N(N-1)}{2}$ (see Equation 1).

If we have only 10 vertices and 12 edges (d=0.266667) this formula gives that there are 28 billion 760 million 21 thousand and 745 different graphs. Another example of how large is the number of possible graphs with given size and density is if we have 100-nodes and 495 edges (d=0.1). The formula above gives $1.3367 \cdot 10^{697}$. Put in context, in theory our universe contains "only" 10^{80} atoms, which is about 10^{627} less than the number of possible 100-nodes graphs with 0.1 density.

This number highly depends on density. Although it is symmetric in the function of density because of binomial numbers. However this symmetry is surprising, because neither ℓ or other measured values show this symmetry. Figure 6 shows the symmetry on 10-node graphs.

Sampling of ℓ is difficult. ER graphs are not sampling well the ℓ if the graph is connected. As we can see in Figure 3 the ℓ of ER graphs is much shorter than expected. After that, sampling ℓ in a non-biased way is not trivial. The sampling algorithm can

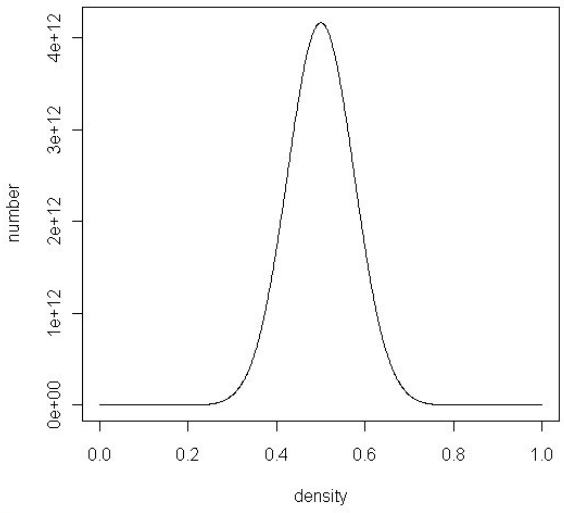


Figure 6: Number of possible graphs with given size

be very complicated and slow. In addition, the trivial enumerating of possible graphs is also wrong, because of very large numbers of permutation. For example if we want to enumerate all of 100-nodes graphs with 0.1 density. The number of possible graphs is $1.3367 \cdot 10^{697}$. During the enumeration the first 80 million or more graphs the result would be identical, they give the same ℓ value, because we rewire only the last edge and it does not cause any change in ℓ.

We enumerated all the 10-node graphs with density 0.2 and it took a week CPU time to perform this task. This is "only" 886 163 135 graphs. Let us imagine how much time it takes to enumerate all the 100-nodes graphs with density 0.1.

Part III

Social System Simulation

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

EMERGENT HUMAN BEHAVIOUR DURING A DISASTER: THEMATIC *VERSUS* COMPLEX SYSTEMS APPROACHES

Damienne Provitolo*, Edwige Dubos-Paillard† and Jean-Pierre Müller‡

Abstract. Disasters or catastrophes engender social and spatial disorganization of the territories affected by these events and specific human behaviour. In this paper, we look both at the responses of societies in terms of specific human behaviour in times of disaster or catastrophe, which may be either a form of vulnerability or on the contrary of social resilience, and at the forms of emergence associated with such exceptional events. The first part proposes a typology of behaviours observed at times of catastrophe and identifies properties common to all such behaviours. Parts two and three ask whether behaviours observed at times of disaster or catastrophe (behaviours that stand apart from everyday behaviours and that can be observed both individually and collectively) can be characterized as emergent behaviours. Answering this involves weighing up inputs from the science of risk and the science of complexity. We present the different properties for characterizing emergent human behaviour, in order to recombine the disciplinary approach of scientific communities. This knowledge of the conditions underlying the emergence of behaviours points the way to how the phenomenon might be modelled.

Keywords. Human behaviour, emergent group, global behaviour, catastrophe, disaster, complex system, emergence, qualitative change.

1 Introduction

Disasters or catastrophes engender social and spatial disorganization of the territories affected by these events. Such disorganization is usually temporary and seldom irreversible. Fritz (1961) lists four main properties characterizing catastrophes or disasters: (a) events that are identified in space and time (date, frequency and duration) (b) that have impacts, (c) on social units, (d) which in return come up with responses or adjustments to those impacts. Fritz thus makes connections between the event, the impacts and the responses to those impacts. The social unit is variable in scale as it encompasses individuals, families, groups, institutions or the entire society. Lares (1992) enhances this concept of disaster by integrating the idea of individual and group behaviour, which are an atypical, immediate and contextualized response to the event: 'El desastre o la emergencia masiva se puede definir come un peligro que se fue mas alla y que incide en la conducta individual y de grupos, ya que los diversos sectores de la poblacion deben dejar de lado sus meta individuales et sociales que configuran su marco de referencia, y formular de manera inmediatamente nuevas metas que deben ser resueltas aqui y ahora '[1].

In this paper, we look both at the responses of societies in terms of specific human behaviour in times of disaster or catastrophe, which may be either a form of vulnerability or on the contrary of social resilience, and at the forms of emergence associated with such exceptional events.

This research postulates that a large majority of societies, no matter their level of economic development, are unprepared to face man-made or natural disasters and domino effects. Paradoxically, the Japan, which is among the most advanced countries in terms of risk culture, has just suffered a chain of events that led to one of the most severe catastrophe that has ever occurred. Despite this exceptional event, recent years have been characterized by the multiplication of major catastrophic phenomena. This tendency is not likely to reverse in the futur because of population growth and the densification of infrastructures and population in high risk areas. If the trend continues, thousands of human lives will be lost and the costs of disasters should exceed USD 300 billion a year (Groupe URD-Urgence, réhabilitation, développement-, 2010). Moreover, as stated by Lagadec 'shocks are becoming very rapidly sytemic, they are evolving quickly from the local level to the national and the international ones' as it has been demonstrated by the eruption of the Eyjafjallajökull volcano in 2010 (Iceland) which caused air traffic disruption in our globalized world, or by the dissemination of an avian influenza virus (H5N1virus) in 2004. Some organizations are now using the devastating natural catastrophe term to describe an event were

*Damienne Provitolo: UMR 6526 Géoazur, CNRS, Sophia Antipolis, France. E-mail: provitol@geoazur.unice.fr
†Edwige Dubos-Paillard: UMR 6049 ThéMA, CNRS , Besanǫn, France
‡Jean-Pierre Müller: CIRAD, Montpellier France

[1] The massive disaster or emergency can be defined as a danger that went beyond and that affects the individual and group behaviour because the various sectors of the population must put aside their individual and social goals that form their framework and they must immediately formulate new goals so that must be resolved "here and now"

the number of fatalities exceeds 500 and/or overall loss exceeds USD 650 millions. (Munich Re, 2011). Moreover, great natural catastrophe term is now used if a regions ability to help itself is distinctly overtaxed, making supraregional or international assistance necessary. As a rule, this is the case when there are thousands of fatalities, hundreds of thousands are left homeless and/or the overall or insured losses are of exceptional proportions given the economic circumstances of the country concerned. (Munich Re, 2011 on the basis of the United Nations definition).

Our societies have to enhance the culture of risk, in order to save human lives but also because the costs of disasters are increasing. The insurance and reinsurance scheme can no longer cover those costs if the increase of "greats" natural or man-made catastrophes observed since early 2010 continues. For example, if a great earthquake, like the one which hit San Francisco in 1906 strikes again in the same region, the number of fatalities would be comparable to what it was in 1906 (due to the growth of population) despite improvements in building codes and construction practices, and the forecast property loss to buildings would be in the range of approximately USD 90 to USD 120 billion (Kircher et al. 2006). Moreover, state governments which until now have always undertaken measures to support financial damage above a certain threshold are actually facing colossal budget deficits (Le Monde, 2011). The challenge is particularly complex in densely populated societies and/or in areas, as in Europe, that have a rich and varied heritage and that have been relatively safeguarded from a great or a devastating catastrophe.

Three levels can be used to reduce the vulnerability of territorial systems : planning policies, civil engineering and awareness raising, education and training for disasters situations. If, this last level is applied in societies that regulary are facing to particular types of hazards (earthquakes, hurricanes, floods), it remains marginally considered in societies where hazards frequency is low. More generally, research related to disasters focus on hazard and its mode of diffusion at detailed levels to improve planning policies and civil engineering. Recent developments in computer science and the evolution of computing power permit to make complex simulations. However, human behaviours modelling is less studied. Yet, this aspect also influences damages and thus their costs all the more when human behaviours are unadapted.

Modelling human behaviours suppose beforehand to identify observed behaviours during catastrophes, their frequency, conditions in which they appear and their effects. It is a necessary precondition to produce simulations as close as possible to reality.

But most often the analysis of human behaviour in disaster situations cover a specific event (many events were described and analyzed by the Disaster Research Center, the Center for the Study of Disasters, or by other authors as Mileti, 1993, Bourque and Russel 1994, who worked on the analysis of behaviours during the 1989 Loma Prieta earthquake ; Ripley, 2008 on the World Trade Center terrorist attacks on September 11, 2001; Morin and al., 2009 on the 2005 volcanic eruption of mount Karthala; Ruin, 2010, on the 2002 flood in the French department of Gard) or even more specifically on a specific behaviour, such as looting or panic (Dupuy, 1991; Quarantelli, 1994a ; Provitolo, 2003; Barsky et al. 2006; Hagenauer et al., 2011).

Analysis on the diversity of behaviours during domino effect are also limited. In addition, the results are generally based on surveys carried out before and/or after a disaster. But the analysis based on different surveys do not always allow a generalization of the results because samples are either too small or not representative of the population (Dauphiné, 2003) or even not reproducible. From the behaviours observed during a specific event, it is then possible to identify the recurring ones for a type of event (flood, earthquake, nuclear or industrial accidents) or for all the disasters. Among the attempts to draw up human behaviour typologies which aim to apply to all disasters, that proposed by Mawson (2005) should be noted.The latter is based on two axes : perceived degree of physical danger, location of attachment figures. But it does not take into account the diversity and change of human behavior according to the dynamics of the event. This typology is not really adapted to our ultimate aim which is to model and simulate human behaviour in spatial and temporal dimensions. That is why, to capture the diversity of behaviours, we adopt an approach based on empirical analysis of different types of disasters.

The originality of this research is to draw up a general typology of human behaviour, not according to the origin of the event (natural hazard, technological or social) but with respect to its temporality and the access to information. This research identifies also the common properties of human behaviours, wheter they are individual or collective behaviours.

Individual and collective behaviours are widely analysed from a disciplinary standpoint (sociological, psychological, economical standpoints) or in terms of complex systems. But collective behaviours are still underconsidered in analysis of human behaviours. However, any disaster creates collective behaviours in the sense that these behaviours are triggered by the same situation and affect the whole community, share common characteristics which are in sharp contrasts to the disparity of individual activities at the normal stage.

The complex systems and the notion of emergence (a notion which refers to two levels of analysis, the individual and the collective level and to the relationship between these levels), seems to be a promising way to identify and modeling the mechanisms underlying these behaviours (Benkirane, 2002; Provitolo and Daudé, 2008; Provitolo, 2009). The issue of emergent behaviours has been rarely considered. This is why, here we review the different points of view (those from risk sciences and those

from complex systems) that will allow to identify a behaviour as emergent. The notion of emergence is open to misinterpretation. For risk sciences the term is generally understood in literary terms, in the sense of the sudden appearance of something, or the arrival at the forefront of a leader or an organization . On the other hand, in the field of nonlinear dynamics and complex systems, emergence makes sense only when adressing levels of organization (Mayet, 2005).

This paper is arranged into three parts. The first part proposes a typology of behaviours observed at times of catastrophe and identifies properties common to all such behaviours. Parts two and three ask whether behaviours observed at times of disaster or catastrophe (behaviours that stand apart from everyday behaviours and that can be observed both individually and collectively) can be characterized as emergent behaviours. Answering this involves weighing up inputs from the science of risk and from complex systems. This knowledge of the conditions underlying the emergence of behaviours points the way to how the phenomenon might be modelled.

2 Diversity of human behaviour in situations of crisis, disaster or catastrophe

Several authors (Quarantelli 1995; Faulkner, 2001; Birkland, 2006) identified criteria for distinguishing disasters, crises and catastrophes : according to the scale or magnitude of the event (the magnitude should be measured in lives or property lost, or by the extent of the failure of the normative or cultural system (Perry, 2006)), according to the actions or inactions of an organization, according to the assistance from regional and national governments or from international or nongovernmental relief organizations, according to the damages, an event can be classified in terms of crises, disasters or catastrophes. But these authors do not specify which threshold or which combinations of variables are used to qualify the event as a crisis, disaster or catastrophe. Quarantelli (2005) identifies six elements to capture the major differences between catastrophes and disasters. In a catastrophe :
- There is massive physical impact (in contrast to the localized impact in disasters).
- Local officials are unable to undertake their usual work roles.
- Help will come mostly from more distant areas (such as international help).
- Most everyday community functions are sharply and simultaneously interrupted.
- International media focus their attention on the event.
- Very high-level officals and governmental agencies from the national level become directly involved in the management of the event.

As for the crisis, it is often defined as an event that exerts heavy and destabilizing pressure on organizations, whether formal organizations or spontaneous. The crisis stakes the adaptive capacity of organizations. Although these authors have distinguished the catastrophe from the disaster and crisis, we use them interchangeably, because they "connote both the idea of a disastrous or catastrophic event, of a disruption, which may involve an infinite variety of situations " (Lepointe, 1991).

Human behaviour in crisis, disaster or catastrophe situations may be analysed over separate periods of time. In this paper, we look at the behaviour of the population before, during and immediately following disaster impact. We try to find out how people react at these times, what they do, and what decisions are made. We make no value judgements of the behaviours and do not look to say whether they are rational or irrational responses (as measured against standard economic yardsticks).

We propose a typology of behaviours observed in the course of various catastrophic events. The events selected (industrial catastrophes of Toulouse (2001) and Bhopal (1984), Three Mile Island nuclear accident (1979), earthquakes of Loma Prieta (1989), Northridge (1998) and Haiti (2010), the string of catastrophes in Japan (2011), mudslide of Vargas (1999), tsunami in Indonesia (2004), volcanic eruption of Karthala (2006), hurricane Katrina (2005), etc.) served as catalysts for observing (from survey data, texts, videos, etc.) how individuals and social units (whether existing before the event or emergent) organize their response and adopt one or more types of behaviour. The observations pertained to natural and technological events or sequences of events in different parts of the world so as to include the diversity of living standards and of cultures in the analysis of behaviour, even if the role of culture is not the specific purpose of this paper.

The typology of human behaviour (Fig. 1) based on observation was drawn up not in terms of the origin of the event but according to its time continuum and access to information. To do this, we have 1) identified (from various data sources) the diversity of behaviours that may occur during a catastrophic event, 2) classified these behaviours depending on the time continuum of the event (time from upstream to downstream of the event) and the information available to the public about the occurrence of a disaster (we take as axiomatic that access to information changes the behaviours), 3) identified the common properties to these behaviours, whether they are individual behaviours or collective behaviours that spread to the whole community (Noto et al., 1994).

Three time periods were distinguished: the phase before the disaster or catastrophe, the disaster/catastrophe phase, and the impact phase. For each of the three dimensions, behaviours of distancing or on the contrary of contact are observed. The pre-disaster phase, which as the name indicates precedes the event, consists of three sequences that are dependent on a system of surveillance and alert:

- When the threat is not announced, it is everyday behaviours that are observed. These are very varied and the travelling that may be associated with them, is spatially very varied.

- When the threatening event is announced but is not imminent, specific behaviours with respect to the event are put in place progressively: spontaneous or organized evacuation behaviour is observed (Three Mile Island, hurricane Katrina), behaviour to counter the potential impact of the danger (e.g. reinforcing defences against tornadoes or flooding), confinement and taking shelter (e.g. going straight back home) or on the contrary gathering close relatives (e.g. collecting children from school). In relation to spontaneous or organized evacuation behaviour, when the potential impact zone is pretty much delineated, a centre-periphery model is progressively put in place. A part of the population leaves the area and institutional emergency crews may move into the area to try to improve security for the population.

- Conversely, when the event is imminent and a personal threat (e.g. danger from a cyclone is confirmed), other behaviours appear: forced or chosen immobility, flight (in panic or reasoned), gathering or mustering of individuals or groups, assistance and emergency services interventions.

The disaster/catastrophe phase consists in a sequence that we call 'foreseen or unforeseen present danger': this is a period (of varying length depending on the type of event) of social and territorial disorganization where everyday behaviours give way to threat-specific behaviours. Most travel is to ensure the protection and survival of oneself or close relatives. Often travel is in groups, sometimes under the authority of a leader, sometimes by the interplay of individual actions. Behaviours of distancing from or coming into contact with the catastrophe are thus observed: flight or fight against the effects of the disaster (during the 1993 Mississippi floods, inhabitants stacked sandbags hoping to keep back the floodwaters), forced immobility (buried under a collapsed building) or chosen immobility (refusal to evacuate), sideration (inability to react), confinement or shelter (travel to get home or get to shelter, etc.), collecting of close relatives (generally within a short radius) or rescue, curiosity (floods at Draguignan in June 2010, France), grouping, assistance and rescue, ensuring public safety. Such behaviours are responses that are more or less adapted to the catastrophe and that change the population's vulnerability or on the contrary its resilience. Their respective importance varies with the type of catastrophe, its space and time dimensions, the alert arrangements, and the visualization of the event.

The phase that follows the event or the impact phase is marked by behaviours similar to those of the event with, in addition, 'antisocial' behaviour, to put a name on looting and theft, etc. Such behaviour is often confined to a minority but is widely reported by the media. Motivations differ greatly between individuals who have lost everything and are trying to survive the event (looking for clothing, water and food) and looters proper looking to get rich quick. On the scene of the catastrophe, organizations emerge to rescue and ensure the survival of individuals. The travel that progressively takes place is closely guided by the centre-periphery model. It intensifies to such an extent that the institutional emergency services very quickly look to set up buffer zones to limit any movement towards the centre other than for the emergency services. The buffer zone is frequently set up within the marginal zone that forms a transitional zone between the affected and unaffected areas. It is also a zone where various rumours may arise that the rescuers and authorities attempt to stifle by their presence and by reactivity to stop them spreading.

From this typology of human behaviours, it is possible to identify three properties common to all of the behaviours observed and identified in the pre-catastrophe, catastrophe and impact phases.

- Behaviours are essentially **non traditional**, they do not correspond to daily behaviours. They generally occur in a threatening environment where **standard methods of acting** cannot be followed or are inappropriate. What characterizes them therefore is that they are at odds with daily behaviour impelled by varied motivations. Such behaviours include sideration, confinement, flight, evacuation, panic, deviant behaviours, etc.

- **Behaviours in a catastrophe period are not specific to a level of analysis**. They may be read with respect to individuals, families, groups or organizations. Studying human behaviour therefore requires us to look at (1) individual, isolated behaviour at odds with traditional behaviour, (2) pre-established collective behaviour (role of collective actors, (3) individual behaviour spreading to the entire group (collective panic). These levels lead us to investigate the concept of emergence.

- **A short-lived behaviours**: behaviours that are limited in time with a return to daily behaviours. Specific behaviours arise with the catastrophe and disappear either with it or during the post-impact phase.

These properties are also presented by some scholars (Rodriguez et al., 2006) as properties allowing emergent behaviour to be characterized.

With respect to the typology of behaviours we propose, the question is whether these behaviours that are at odds with daily behaviours and that can be observed both among individuals and collectively can be characterized as emergent behaviour. To answer this we rely on the

Time continuum of event / Types of human behaviour	Pre disaster phase: Danger announced but not imminent	Pre disaster phase: Danger imminent and personal threat	Disaster phase: Foreseen or unforeseen present danger	Impact phase: Post-disaster phase
Evacuation	x	x	x	x
Fight against the effects of the disaster	x	x	x	
Forced or chosen immobility		x	x	x
Sideration			x	x
Flight (in panic or reasoned)		x	x	x
Confinement/Taking shelter	x	x	x	x
Search for close relatives	x	x	x	x
Looking for help			x	x
Curiosity			x	x
Back to the place of residence/of work			x	x
Grouping		x	x	x
Assistance/emergency services		x	x	x
"Antisocial" behaviour				x

Figure 1: Typology of human behaviour

typologies of emergent behaviours and groups proposed by thematics specialists, on the typology of emergence proposed by J. Fromm (2004) (viewpoint of complex systems) and on its applicability to the analysis of human behaviours. This knowledge of the conditions that underlie the emergence of behaviours allows us to posit the modalities for possible modelling.

3 Emergence of organizations and behaviours: the viewpoint of thematics specialists

Individual and collective behaviour is widely analysed from a disciplinary standpoint (sociological, psychological, economical standpoints) or in terms of complex systems. In sociology, the research initially gave priority to the study of individual reactions in stressful situations (Wolfenstein, 1957), even if from the 1960s some pioneering studies on individual and collective behaviour exist (Beach and Lucas, 1960). Research on collective behaviour has also been developed in the field of economic sociology or economics where the focus is on economic behaviour of individuals (Granovetter, 1978), in the field of psychology with the pioneering work of G. Le Bon (2003), for whom "the crowd is a flock that can not do without a leader". But more recent research has shown that the collective behaviour was not necessarily related to the appearence of a leader. For example, the collective panic would thus emerge from the diffusion of individual panic, without the attendance, the domination of a leader who would call to the panic (Provitolo, 2005, 2009).

It very quickly appeared that there is no consensus as to the definition of emergence. On this question of the emergence of human and organizational behaviour specific to times of danger or disaster, thematics specialists working in the analysis of behaviour often use the concept of 'emergent behaviour' and have proposed various typologies for identifying them. This research has been developed essentially by sociologists and especially by a field of sociology dealing with 'collective behaviour' (Rodriguez et al., 2006). This disciplinary field is more especially interested in the dynamics of social phenomena such as crowd behaviour, looting, panic, revolutions, etc.

Some work (*Disaster Research Center*; Kreps, 1984, 2006; Quarantelli, 1994b; Tierney et al., 2001; Rodriguez et al., 2006) has investigated the emergence of institutionalized groups (routine patterns of action in existing organizations, Weller and Quarantelli, 1973) or informal collective behaviour for managing an event (crisis, disaster, catastrophe).

The Disaster Research Center (*DRC*) proposed back in 1963 a typology of organized behaviours at the time of the crisis period in disasters. This typology consists of a matrix with four quadrants, with one of the two axes defining the Tasks (regular or non regular) and the other the

Structure of the organization (old or new). From these two axes, the Disaster Research Center distinguishes four types of organized response, only one of which is considered emergent:

- **Type I:** Established responses (regular tasks and old structures): this group exists before an event and many of its actions are planned (hospitals, emergency medical services, etc.).
- **Type III:** Extending responses (nonregular tasks and old structures): this group exists before an event but many of its actions are not predetermined (the government agencies that aid in managing the removal of debris and help in rebuilding operations, as in Haiti, for example).

For types I and III, the organization exists before an event and many of its actions are either planned (T. I) or not predetermined (T.3).

- **Type II:** Expanding responses (regular tasks and new structures): many of the actions are planned but the basic structure of the organization shifts from a small group of professionals to a larger group of volunteers (Red Cross, National Guard, etc.).
- **Type IV:** Emergent responses (new tasks and new structures): its existence and activities are ad hoc and therefore unique to the event. These are small or large groups that take shape and carry out tasks or activities that institutionalized groups cannot accomplish. Thus the emergent organized response (people sometimes speak of 'emergent groups' too), is related to the idea of non traditional and new behaviour (example of mutual assistance groups that form just after a catastrophe to look for the injured and help evacuate them). While the informal emergent groups are generally organized in the period after the disaster and more rarely during the event, during which period individuals organize their actions more around their families and friends (Quarantelli, 1988), institutionalized groups, whether emergent or not, act both during and after the event.

Continuing this line of research, keeping as key variables the length of time the organization has existed and its functions, and looking into the question of forms of emergence, Quarantelli (1984, 1994b) distinguishes emergent behaviour from emergent groups. This typology, which applies to the time of crisis, identifies four types of emergent behaviour and considers emergence of the group as on of those types (Fig. 3). The originality of this typology is to reveal that in most of the organizations and groups that were not emergent, there were nonetheless considerable emergent phenomena (Tierney et al., 2001). During a disaster, each of these types of emergence may be present simultaneously.

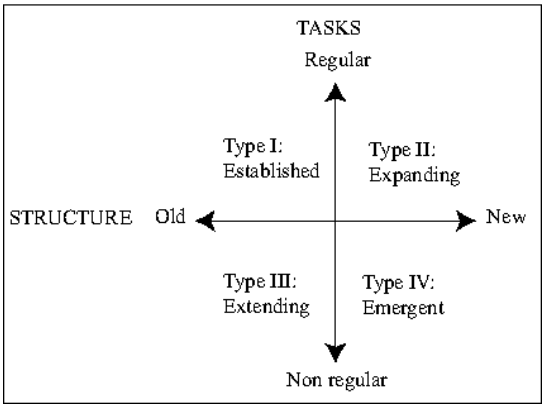


Figure 2: Typology of organized responses established by the Disaster Research Center

		STRUCTURES	
		Old	New
FUNCTIONS OR TASKS	Old	Quasi-Emergent Behavior	Structural Emergent Behavior
	New	Task Emergent Behavior	Group Emergence or Emergent Groups

Figure 3: Four types of emergent behaviour

Quasi-emergent behaviour correspond to established organizations (type I 'established groups' of the DRC typology) that continue to perform their roles and missions as often they do not undergo any major change in their structures or functions. However, there may sporadically be a change in actors so that the organization operates to the full (this is referred to as temporary or minor emergent behaviour).

Structural emergence corresponds to organizations which, to continue to perform their old functions, develop a new structure. Being unable to accomplish their missions, the organizations are then temporarily replaced by individual actors (e.g. amateur radio operators providing a liaison with the Weather Service Office during the New Orleans floods) (extending group by the DRC typology). A temporary link-up or social network is thus put in place.

Task emergence is what happens with organizations whose structure is unchanged in any way but that take on a new task. These, then, are organizations that take on the roles of a defaulting organization. This category replaces what the DRC calls an 'expanding organization'.

Lastly, emergent groups are characterized by both a new structure and function (identical definition to the DRC typology).

The typologies proposed by the DRC and Quarantelli apply to behaviours that emerge within institutional or informal organizations; but these typologies cannot readily encompass the variety of behaviours that can be ob-

served among individuals and which, because of the combinations of actions and interactions, engender emergent collective behaviour (e.g. collective flight, spontaneous gatherings).

Confronted with this observation, we wish to resituate the analysis of emergent behaviour (as meant by thematics specialists, that is, in the sense of non-traditional and unexpected behaviour associated with the performance of a task and the existence of a structure) in the more theoretical context of identification of the properties of emergence (in the sense of complex systems).

4 Human behaviours and organizational responses with respect to types of emergence

Emergence has many meanings within the complex systems. But all of the definitions proposed refer to the connections between the constituents of a system, to relations between the micro and macro levels, and to the appearance of structures, properties or forms at macroscopic level. For some scholars (Bunge, 1977; Searle, 1995), these entities or structures described at macro level cannot be reduced to the composition of individual properties at local level.

The purpose here is to analyse human behaviour and organizational responses in situations of crisis, disaster or catastrophe with respect to one of the emergence typologies used in the area of complex systems. This analysis allows us to identify emergent behaviour from the properties of emergence and to combine perspectives and disciplines by expressing the multiple points of view, those of thematics specialists and those of complex system model makers.

We use the typology of J. Fromm (2004) who, for all types of system, distinguishes four types of emergence (simple, weak, multiple and strong) and dissociates unilateral emergence (bottom-up but no top-down feedback) from bilateral emergence (with top-down feedback).

4.1 Simple Emergence without top-down feedback (Type I)

Type I describes unilateral emergence (Fig. 4), that is, with no feedback loop between macro- and microscopic levels. 'Unilateral emergence arises from the influence of attributes alone, from relations between the microscopic entities and the outside context' (B. Walliser, 2005).

The phenomenon emerges at macroscopic level because of combined actions or interactions between the constituents at microscopic level. Fromm distinguishes (Ia) simple intentional emergence from (Ib) simple unintentional emergence. In type Ib, individuals are unaware of the macroscopic phenomenon they are helping to create.

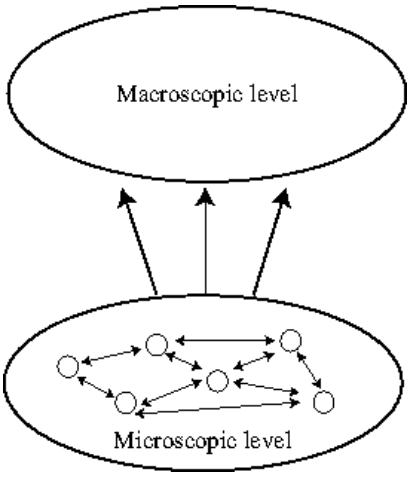


Figure 4: Simple emergence without top-down feedback (adapted from Fromm, 2004)

However, if for type Ia agents are aware of the emerging structure, a specific and fixed role is assigned to each part, and this role does not change in the course of time. Type Ia reveals an absence of flexibility and adaptability, because the roles are fixed and predictable. That means: 'A defect in a single component can bring the entire system to a halt'. The established organization/organized response (regular task and old structure, DRC type I) corresponds to this type of emergence where individual institutional actors take on specific and set roles and interactions among those actors emerge from organized intervention groups (hospital care services, emergency medical services, etc.). However, this type of emergence is rare in situations of crisis, disaster or catastrophe since the actors are generally able to adapt and modify their role so that the organization operates fully (this is known as actor substitution). This is the idea proposed by Quarantelli with the typology entitled Quasi-Emergent Behaviour. For type Ib, the system components have no specific role; emergence is said to be unintentional since it results solely from the interplay of combined actions or interactions among entities (microscopic level). If we refer to the typology of behaviour proposed in part one (Fig. 1), the fight against the effects of the disaster (chain of individuals interacting to fight flooding: e.g. the inhabitants of Des Moines (Iowa) stacked thousand of sandbags in the hope of damming the floodwaters of the Des Moines river in 1993), or mutual assistance behaviours put in place without a leader (as with mutual-help groups that may form spontaneously just after a catastrophe through a combination of individual actions to search for and help evacuate the injured: DRC type IV, emergent organized response) come under simple unintentional emergence. However, because of the principle of non-compositionality (in the sense that descriptions at macro level are not a simple composition of individual properties at micro level, Bunge, 1977), group behaviours such as flight are not emergent behaviour (each individual behaviour is a

flight and the group behaviour is simply the effect resulting from individual flight). There are therefore multiple viewpoints: that of thematics specialists for whom group behaviours are emergent when they break with everyday behaviour, that of model makers of complex systems for whom the concept of emergence can describe how a non-compositional behaviour at macro level rests on a structure of interactions at micro level.

Beyond a critical point, in addition to local interactions, the macroscopic level (e.g. mutual-assistance groups fighting the effects of a disaster) has a feedback effect on the micro level, which then requires such behaviour to be ranked in another category of emergence: weak emergence with top-down feedback.

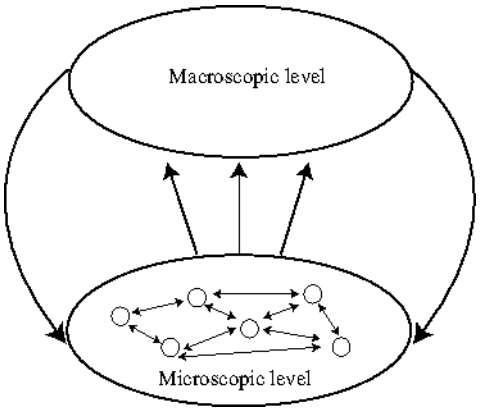


Figure 5: Weak emergence including top-down feedback (adapted from Fromm, 2004)

4.2 Weak Emergence including top-down feedback -Type II-

Weak emergence (type II) (Fig. 5) includes bottom-up (from micro to macro), top-down (from macro to micro) feedback and self-organizing processes. The roles of agents or actors are flexible. This is the framework of bilateral emergence since there is feedback action from the macroscopic phenomenon to the microscopic level. This bilateral emergence may be synchronic when the influences between levels are exerted instantaneously (e.g. actors co-ordinate their behaviours and an evacuation function emerges that vanishes as soon as the actors cease to be co-ordinated) or diachronic when a macroscopic structure arises from the dynamic behaviour of systems. For Fromm, top-down influences are exerted more slowly than bottom-up influences. All individual behaviours observed during catastrophic events (Fig. 1) may be included in this category provided that from these individual behaviours there emerges a new structure at macro level (a spontaneous mutual-assistance group – DRC type IV –, an assembly dynamic, etc.), and the new structure provides feedback for the microscopic entities without these entities having any representation of emergence. We call this 'indirect' or 'physical' feedback in the sense that the aggregate phenomenon exists only for an observer and may cause or modify the conditions that constrain individual behaviours (e.g. when a crowd grows, constraints on movement increase and so reduce the choices as to where to move; individual behaviour is then subjected to the movement of the crowd) in contradistinction to the concept of strong emergence which includes the idea of representation (type IV of Fromm's typology).

In the case of weak emergences, only the observer can 'observe, view' from above (like an airline pilot) what is going on: a common trajectory of population movement towards a precise location, etc. Accordingly the phenomenon can be read in two different ways: that of the individual agent acting locally within his environment, or that of the outside observer. Fromm distinguishes weak stable emergence from weak instable emergence. In both instances, bottom-up and top-down feedback loops come into play. For weak stable emergence, there is positive feedback from micro to macro level and negative feedback from macro to micro level. This double feedback loop regulates the system. However, for weak unstable emergence, there is a double positive feedback loop since the emergence of a new structure tends to reinforce the phenomenon. The emergence of collective panic from individual panic in a crowd made up of panicking and non-panicking individuals may fit in with 'unstable simple emergence' since instantaneous interactions among the microscopic entities (the panicking and non-panicking individuals) spread to the entire crowd (macro level) which, above a certain level of panicking population feeds back onto all of the individuals reinforcing the panic behaviour (process of propagation in space and time).

However, if the actor is aware of the emergent or emerging structure, behaviours no longer fit into the weak emergence but into the strong emergence typology.

4.3 Multiple Emergence with many feedbacks (Type III)

This type of emergence is based on the idea (1) that emergent structures are often a combination of stable and unstable weak emergence and (2) that feedback occurs over the short term (for positive feedback: e.g. when blindly imitating your neighbours behaviour) and the long term (for negative feedback: when adopting thoughtful behaviour, behaviour that is not the outcome of the pressure of conditions that compel individual behaviours in a simple way type II), and (3) that, in addition to these two properties, a form of emergence can be named 'adaptive emergence' for complex systems demonstrating adaptive capacities (Fig. 6). Of course, all complex systems are not adaptive, but all those involving living societies are.

In catastrophe situations, human behaviours are modified in the short term as a function of the cognitive factors and of the adaptive capacity of individuals (idea of appearance of new roles and disappearance of old ones: e.g. organizations that take on the roles of a defaulting

organization cf. typology Task emergent behaviour). This adaptive capacity changes over the long term with experience and learning (long-term memory).

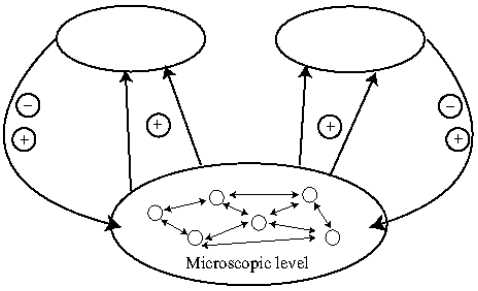


Figure 6: Multiple emergence with many feedbacks (adapted from Fromm, 2004)

4.4 Strong Emergence (Type IV)

Strong emergence (Fig. 7) is bilateral emergence that encompasses the idea of the intentionality and reflexivity of actors (Müller, 2004; Walliser, 2005). The system contains elements that are able to observe the collective effects of their interaction and the representations they derive from those observations influence their decisions and behaviour (Müller, 2004). These actors are more or less aware of the emergent phenomenon they tend to produce and of the mechanism that leads to it, which may modify the phenomenon itself (Walliser, 2005). Many institutionalized behaviours and some collective behaviours (hospital response in disasters) are the result of interaction among actors' individual actions. These actors consciously participate in voluntarily constructing organized groups with roles and missions that are relatively clearly defined beforehand or during the actor-substitution process. The actors are thus aware of the emergent phenomenon but, unlike with simple emergence, the roles and missions they must carry out are flexible (Red Cross, emergency fire services, etc.). Behaviours characterized as 'quasi emergent behaviour', 'task emergent behaviour' and 'structural emergent behaviour' may thus be classed in the strong emergence type when the overall project is explicitly known by individuals. However, spontaneous organizations (search and rescue groups) that appear during the post-shock phase may be part of this category of emergence if and only if the individuals are aware of the emergence process and of the organization that is being put in place. This information can only be extracted by interviews after the disaster or the catastrophe.

To conclude, in order to characterize emergence in all its forms, two levels must be distinguished: a local level of interacting individuals/entities and a global level. Emergence occurs if the behaviour or the global structure cannot be reduced to local behaviours or structures (Bunge,

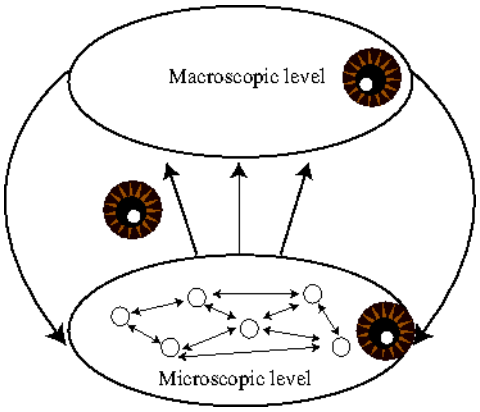


Figure 7: Strong emergence (adapted from Fromm, 2004)

1977; Searle, 1995). We characterize emergence as weak if the phenomenon as such does not have any feedback on the local level. In this sense the global phenomenon exists only for an outside observer. This global phenomenon may, nonetheless, have no influence on the micro level (type I of Fromm's classification), it may cause or modify the conditions that constrain individual behaviours in a simple way (type II of Fromm's classification), or it may be multiple (type III of Fromm's classification). Emergence is strong if the global phenomenon has feedback as such on the local level, which means local entities are 'aware' of the global phenomenon. All of these emergences are synchronic since the local level and global level co-exist at all times simply because they are posited by definition; such emergences are therefore necessarily instantaneous or more specifically a-temporal. We have also been able to see that any group of interacting entities engenders global behaviour (such as collective flight). It is therefore possible to model the dynamics (the behaviour) of the group in the form of a dynamic system whose trajectory in state space is to be observed. We are then interested in the qualitative change of this behaviour. Three situations may be observed:

(1) there can be a move from non-emergent to emergent behaviour (from collective flight to a structured evacuation behaviour, with roles as guides, etc.)

(2) there can be a move from one emergent behaviour to another (from evacuation behaviour to establishing a security perimeter)

(3) the emergent behaviour vanishes (collective flight resumes or everyone goes back about their business)

Cases (1) and (2) are examples of diachronic emergence. In case (1) emergence comes about, in case (2) there is a shift from one form of emergence to another (change of attractor). In case (3) where emergence disappears, it may come about that the global level is no longer identifiable as such. There is therefore a combination of two concepts of complex systems: (1) emergence, (2) the de-

scription of dynamic systems and especially of their characterization in terms of trajectories that make it possible to describe attractors and around them bifurcations and chaotic regimes. One is a-temporal and is a characterization of a system at two levels of observation; the other is temporal and describes trajectories.

5 Conclusion and directions for further research

In this paper, we have proposed a typology of behaviours observed during different catastrophic events, whether natural or technological in origin, and whether local or dispersed. We have constructed this typology not in accordance with the origin of the event but with respect to a time continuum: the pre-catastrophe phase, the catastrophe phase, and the impact phase. For these three dimensions, we have identified a variety of behaviours, some corresponding to distancing from the event, others on the contrary corresponding to coming into contact with it. We have also identified three properties common to all these behaviours: they are essentially non-traditional, they are not specific to a level of analysis (behaviours may take shape for individuals, families, groups or organizations), they are short-lived. These properties are also presented by certain scholars and by risk and catastrophe thematics specialists as properties for characterizing emergent behaviour. Now, in the field of complex systems, to characterize emergence in all its forms, two levels must be distinguished: a local level of interacting individuals/entities and a global level. In the absence of an identifiable global level, there is no emergence, which is therefore contrary to observations reported by pioneer field researchers: during the crisis period of disasters, there was a great deal of emergent behavior, both at the individual and group levels. The emergent quality took the form of nontraditional or new behaviour, different from routine or customary norm-guided actions.

We have therefore presented the different properties for characterizing emergent human behaviour, in order to recombine the disciplinary approach of scientific communities, both those conducting research in the domain of crises and disasters and those developing their research in the domain of complex systems. This first exercise has therefore enabled us to bring together approaches and disciplines and to express the multiple viewpoints. It could be useful subsequently for modelling whereby thematics specialists and model makers can account for emergent human behaviour in situations of crisis, disaster or catastrophe.

Acknowledgments

The research for this work was supported by the *Réseau National des Systèmes Complexes (RNSC).*

References

[1] L. Barsky, J. Trainor, and M. Torres, *Disaster realities in the aftermath of Hurricane Katrina: Revisiting the looting myth.* Natural Hazards Center Quick Response Report Number 184, February, 2006. Boulder, Colorado: Natural Hazards Center, University of Colorado at Boulder. Miscellaneous Report, number 53, 2006.

[2] H.D. Beach and R. A. Lucas, "Individual and Group Behavior in a Coal Mine Disaster", *Disaster Study*, number 13, Washington, National Academy of Sciences, National Research Council, 1960.

[3] R. Benkirane, *La complexité, vertiges et promesses*, Le Pommier, Paris, 2002.

[4] T.A. Birkland, *Lessons of disaster: policy change after catastrophic events*, Georgetown University Press, Washington D.C., 2006.

[5] L.B. Bourque, L.A. Russell, with the assistance of G.L. Krauss, D. Riopelle, J.D. Goltz, M. Greene, S. McAffee, S. Nathe, *Experiences During and Responses to the Loma Prieta Earthquake*, Research Reports, Governor's Office of Emergency Services, State of California, July 1994.

[6] M. Bunge, "Emergence and the mind", *Neuroscience* 2, pp. 501-509, 1977.

[7] A. Dauphiné, *Risques et catastrophes : observer, spatialiser, comprendre, gérer*, Armand Colin, Paris, 2003.

[8] J.P. Dupuy, *La panique*, Les empêcheurs de penser en rond, Paris, 1991.

[9] B. Faulkner, "Towards a framework for tourism disaster management",*Tourism Management* 22, pp. 135-141, 2001.

[10] C.E. Fritz, Disasters. In R.K. Merton and R.A. Nisbet (Eds), *Contemporary social problems. An introduction to the sociology of deviant behavior and social disorganization.* University of California Press, Riverside, 1961.

[11] J. Fromm, *The Emergence of Complexity*, Kassel University Press, 2004.

[12] J. Hagenauer, M. Helbich, and M. Leitner, "Visualization of crime trajectories with self-organizing maps: a case study on evaluating the impact of hurricanes on spatio-temporal crime hotspots", *25th Conference of the International Cartographic Association*, Paris, July 2011.

[13] M. Granovetter "Threshold Models of Collective Behavior", *The American Journal of Sociology*, vol. 83(6), pp., 1420-1443, 1978.

[14] G.A. Kreps"Sociological inquiry ad disaster research", *Annual Review of Sociology*, vol. 57, pp. 309-314, 1984.

[15] G.A. Kreps and S.L. Bosworth, Organizational Adaptation to Disaster. In H. Rodriguez, E. L. Quarantelli, and R. R. Dynes (Eds.), *Handbook of disaster research*, Springer, New York, pp. 297-315, 2006.

[16] C. A. Kircher, H. A. Seligson, J. Bouabid, and G. C. Morrow, "When the Big One Strikes Again. Estimated Losses due to a Repeat of the 1906 San Francisco Earthquake," *Earthquake Spectra*, Special Issue II, vol. 22, Oakland, California: Earthquake Engineering Research Institute, 2006.

[17] A. Lares, "Problemas relacionados con los disturbios civiles y el analisis de la conducta humana en situaciones de emergencias masivas," *Conferencia Internacional Manejo de Emergencias Masivas*, Carabobo, Venezuela, Sept. 1992.

[18] G. Le Bon, *Psychologie des foules*, Presses universitaires de France, Paris 13, 2003.

[19] Lepointe E., "Le sociologue et les désastres", *Cahiers internationaux de sociologie*, volume XC, pp. 145-174, 1991.

[20] A.R. Mawson, "Understanding mass panic and other collective responses to threat and disaster", *Psychiatry*, vol.68(2), pp.95-113, 2005.

[21] L. Mayet, "L'émergence d'une nouvelle science", *Sciences et Avenir, Hors Série*, L'Enigme de l'émergence, July, August 2005.

[22] D. Mileti (dir), "The Loma Prieta, California, Earthquake of October 17, 1989. Public response," *US Geological Survey*, Reston (Virginia), Professional Paper 1553-B, 69 p., 1993.

[23] J. Morin, F. Lavigne , P. Bachèlery , A. Finizola, N. Villeneuve, "Institutional and peoples response in the face of volcanic hazards in island environment: Case of Karthala volcano, Comoros Archipelago. Part I - Analysis of the May 2006 eruptive crisis". *SHIMA The International Journal of Research into Island Cultures*, vol. 3, Issue 1, pp. 33-53, 2009.

[24] J.P. Muller, "Emergence of Collective Behaviour and Problem-Solving",*Engineering Societies in the Agents World IV*, Fourth International Workshop ESAW-2003, Revised Selected and InvitedPapers, LNAI 3071, Springer Verlag, pp.1-20, 2004.

[25] Munich Re Group, "Catastrophes naturelles 2010 Analyses-Évaluations-Positions" *Topics Geo* vol. Munich Re Group, 2011.

[26] R. Noto, P. Huguenard, A. Larcan, *Médecine de catastrophe*, Masson, Paris, 1994.

[27] R. Perry, What is a disaster? In H. Rodriguez, E. L. Quarantelli, and R. R. Dynes (Eds.), *Handbook of disaster research*, Springer, New York, pp. 1-15, 2006.

[28] D. Provitolo, "Un exemple d'effets de dominos : la panique dans les catastrophes urbaines", *Cybergéo*, 328, 2003. `http://www.cybergeo.eu/index2991.html`

[29] D. Provitolo, E. Daudé, "System dynamics versus cellular automata in modelling panic situations," *Actes du colloque ESM08, The 2008 European Simulation and Modelling Conference*, Le Havre, France, pp. 535-540, Oct. 2008.

[30] D. Provitolo, A new classification of catastrophes based on Complexity Criteria. In M.A. Aziz Alaoui and C. Bertelle (Eds) *From System Complexity to Emergent properties*, Series Understanding complex systems, Springer-Verlag, pp. 179-194, 2009.

[31] E.L. Quarantelli, *Emergent behavior at the emergency time periods of disasters*, Final Report 31, DisasterResearch Center, University of Delaware, 1984.

[32] E.L. Quarantelli, "Looting and antisocial behavior in disasters". *Preliminary Paper 205*. Newark, DE: University of Delaware Disaster Research Center, 1994a.

[33] E.L. Quarantelli, *Emergent behaviors and groups in the crisis time periods of disasters*. University of Delaware, DisasterResearch Center, 1994b.

[34] E.L. Quarantelli, *Catastrophes are different from disasters: some implications for crisis planning and managing drawn from Katrina*. Social Science Research Council. http://understandingkatrina.ssrc.org/Quarantelli, 2005.

[35] E.L. Quarantelli, "What is a disaster?", *International Journal of Mass Emergencies and Disaster* vol. 13(3), pp. 221-230, 1995.

[36] A. Ripley, *The unthinkable. Who survives when disasters strikes-and why*, Three Rivers Press, New York, 2008.

[37] H. Rodriguez, E.L. Quarantelli, R. R. Dynes, *Handbook of disaster research*, Springer, New York, 2006.

[38] I. Ruin, "Conduite à contre-courant et crues rapides, le conflit du quotidien et de l'exceptionnel", *Annales de géographie*, vol. 4(674), pp. 419-432, 2010.

[39] J.R. Searle, *La redécouverte de l'esprit*, Gallimard, Paris, 1995.

[40] K.J. Tierney, M. K. Lindell, R.W. Perry, *Facing the Unexpected: disaster preparedness and response in the United States*, Joseph Henry Press, Washington D.C, 2001.

[41] B. Walliser "Emergence synchronique et émergence diachronique", *Sciences et Avenir, Hors Série*, L'Enigme de l'émergence, July, August 2005.

[42] J.M. Weller, and E.L. Quarantelli, "NeglectedCharacteristics of Collective Behavior", *The American Journal of Sociology*, vol. 79(3), pp. 665-685, 1973.

[43] M. Wolfenstein, *Disasler : A Psychological Essay*, Glencoe, III, 122, Free Press, 1957.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

FROM COMPLEX NETWOKS TO URBAN MOBILITY MODELING: THE RETICULAR MODEL FOR URBAN SIMULATION (REMUS)

Dong-Binh Tran, Rassil Frigui, Diego Moreno, Arnaud Piombini and Dominique Badariotti *†‡

The REMUS model is based on the concept of graph theory which allows modeling the network accessibility of time/distance-dependent multimodal transport and urban pattern. The urban graph is represented by buildings and transportation network.

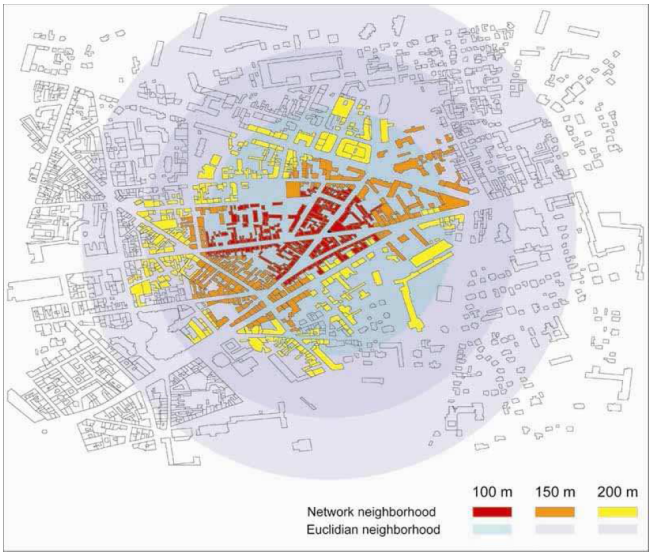


Figure 1: Comparison of network and Euclidian neighbourhoods of a given building at the distance of 100, 150 and 200 m (computed with REMUS)

The REMUS model allows the extraction of neighborhood proximity of complex networks which represent the accessibility in terms of network-time-distances between buildings, according to a given modal choice and a given time/distance. It visualizes the neighborhood of each spatial unit (building) (Figure 1), and calculates topological graph-based indicators characterizing the real distribution of spatial units, depending on the local urban pattern.

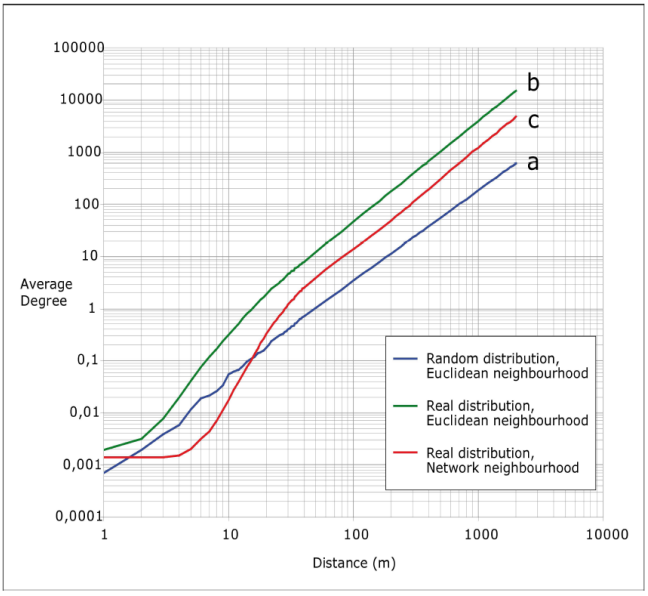


Figure 2: Average degree comparison of neighbourhood graphs: (a) random distribution and euclidian-defined, (b) real distribution and euclidian-defined, (c) real distribution and network-defined

The REMUS indicators are comparative and could be used to compare the topological properties of different urban patterns (quarters or whole cities) (Figure 2). They have been applied to cities and urban quarters, demonstrating that the real distribution of buildings and the introduction of network accessibility in urban metrics induce an important anisotropy in urban space at local scales, which is obviously impacting mobility and modal choice. In a second step, these topological indicators, summarizing the morphological properties of local urban patterns, are compared to mobility indicators extracted from surveys. The main objective is to respond how

*Dong-Binh Tran, Rassil Frigui, Arnaud Piiombini and Dominique Badariotti are with Image, Ville, Environment Laboratory ERL 7230, Faculty of Geography and Planning, University of Strasbourg, 3 rue de l'Argonne, 67000 Strasbourg, France.

†Diego Moreno is with ESPACE Laboratory UMR 6012, University of Nice Sophia Antipolis, Campus Carlone, Hall H, Bureau 429, 98 bd Edouard Herriot, BP 3209, 06204 Nice Cedex, France

‡Dominique Badariotti is the corresponding author. email: dominique.badariotti@live-cnrs.unistra.fr

and why urban mobility is influenced by the local urban pattern.

Keywords: Accessibility, Complex networks, Mobility, Urban Neighbourhood.

References

[1] Badariotti Dominique, Banos Arnaud and Moreno Diego (2007) *Conception d'un automate cellulaire non stationnaire à base de graphe pour modéliser la structure spatiale urbaine : le modèle REMUS*, Cybergeo Journal, Article 403, October 3, 2007, 16 p.

[2] Badariotti Dominique, Banos Arnaud et Moreno Diego (2009) *Morphologie urbaine et réseau. Etude des discontinuités et des ruptures induites par le réseau de circulation à laide du modèle Remus*, Revue internationale de géomatique, vol 19, n1/2009, p. 45-66.

[3] Moreno Diego, Badariotti Dominique, Banos Arnaud (2009) *Integrating morphology in urban simulation through reticular automata modelling* Chap 7, in Handbook of theoretical and quantitative geography , F. Bavaud and C. Mager (eds), University of Lausanne, p. 261-310.

[4] Moreno Diego (2009) *Une approche réseau pour lintégration de la morphologie urbaine dans la modélisation spatiale individu-centrée*. 208 p., Phd of UPPA within the project Comportement des acteurs individuels et émergences sociales et spatiales (Scholarship from Région Aquitaine). Defended on December 9, 2009.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

ROUANTS SIMULATION PLATFORM TO MODEL SERVICE-USER DYNAMICS OF CULTURAL SITES WITHIN URBAN AREA

Rawan Ghnemat [*], Françoise Lucchini [†] and Cyrille Bertelle [‡]

Abstract. *Rouants* simulation platform concerns spatial analysis of urban dynamics described here by services development and their practice by users. A study case allow to analyse cultural sites development applied to the French urban area of Rouen.

Keywords. Urban Dynamics, Cultural Sites, Swarm Intelligence, Geographical Information System, Urban Spatial Environment.

1 Introduction

This paper emphasizes the interest of understanding the functioning of dynamics of collective cultural forms in urban area, as social reality for everyone [5]. Geographers have focused on understanding the interrelationships that people forged with their territory. They work on experiments on social systems belonging to specific territories, such as cities [2, 18, 25]. City is generally understood as a complex system that can be described with several organizational levels [3, 17]: micro, meso and macro levels. At the micro level, the system is the set of interrelationships linking intra-urban elements such as individuals, households and institutions. At the meso level, the city is considered as a systemic entity with functional responses generally included in the mechanisms of competition and/or complementarity with other neighboring entities. Finally, at the macro level, urban systems are composed of interdependent cities belonging to national and international territories. We position this research at the intra-urban level.

It is possible to describe cultural behavior of citizen according to cultural sites on a real urban environment. We focus on describing how the spatial configuration of a city may compel individual behavior within social systems [19, 21, 22]. In this paper, we explore, for a specific intra-urban area (Rouen, French urban area), the complex spatial mechanisms linked to an urban activity development [20]. We specifically study the development of cultural sites in time and on a specific city space, as well as the practice of social individuals to these cultural sites. We are also interested in understanding and modeling the adaptive mechanisms of these cultural sites according to their user practices [14].

To reflect the complex reality of the creation and practice of these cultural sites, we have built a theoretical model of relationships between actors who produce and use the cultural sites. For this purpose, the model proposes to experiment a spatio-temporal system consisting of three groups of actors; the social individuals, the cultural sites and the urban spatial environment, as described on Figure 1. The complexity results from the interleaving of different types of interactions:

- The complexity coming from interactions between social individuals and producing self-organization [3];
- The inherent complexity of social individuals characterized by multiple criteria which are themselves in interaction (age, gender, educational level, social level, etc.);
- The game of competition and/or complementarity in the attractiveness of cultural sites varying on several levels of attractivity depending on their nature and their location (for example, the attractive power of a cinema in comparison with the attractive power of an opera, or, as another example, the attractive power of a site at central urban location in comparison with another site at peripheral urban location);
- The effect of the general spatial configuration of a city being more or less constraining and so acting on the formation of self-organized dynamics (as examples, the barrier effect from a river or from an industrial area, the attractor effect coming from a municipal center or from a high density residential area, the accessibility produced by road transport planning and by public transport planning).

Spatial dynamics and temporal dynamics are mixed. It is usual to observe that, after some periods, bifurca-

[*]Rawan Ghnemat is with German-Jordanian University, Amman, Jordan. E-mail: rawan.ghnemat@gju.edu.jo
[†]Françoise Lucchini is with UMR IDEES/MTG, University of Rouen, France. E-mail: francoise.lucchini@univ-rouen.fr
[‡]Cyrille Bertelle is with LITIS, University of Le Havre, France. E-mail: cyrille.bertelle@univ-lehavre.fr

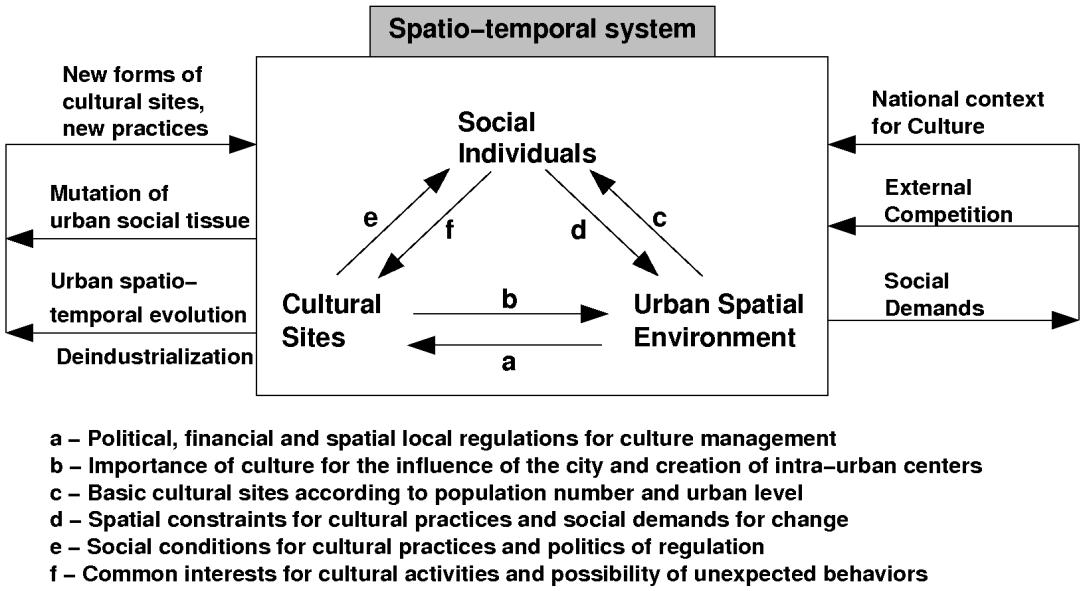


Figure 1: Spatio-temporal system to model the actors involved in urban cultural site development

tions appear in urban planning: for instance, it is the case of the reorganization of brownfields into cultural centers, within spaces previously dedicated to productive activities and so repulsive to cultural sites.

This works results from dialogue between geographers and computer scientists, in order to understand territory functionnalities. Territory is usually grasped by geographers as a system of interrelations linking itself with its inhabitants. "How territory works as system?" is the core of our approach. First, we need to understand what are the relevent components of this system, then to highlight the associated mechanism of actions on them to finally being able to design operational objectives of management. Temporal and spatial dynamics act jointly on all territories. For instance, urban change is a permanently renewed process that geographers have integrated since long time ago within their analysis of urban territories. Dynamics of such changes can be expressed in following terms: how urban territory evolves when a new activity takes place on a specific urban location? For instance, what will be the impact of new cultural structures like a cinema multiplex or a new music scene in requalifying urban zone produced by industrial and harbour brownfields? It is precisely what's happend on the urban area of the French city of Rouen. The process of requalifying old sites into new uses began in Rouen since 2009, particularly in an important part of its central area, previously devoted to intensive industrial and harbour activities on each side of the Seine river. "Seine Ouest" project deals with the reality of urban rehabilitation by the development of cultural centers and by their attractive potentiel they are able to generate in terms of visitor number on this urban zone.

ROUANTS simulation platform has been created to allow the visualization of the attractivity of cultural centers within the urban area of the French city of Rouen. This platform open ways of "explanation" about the most discriminent criteria, about competitive games involving cultural centers, about mobility questions and strategies of adaptation. We let organization emerging from a world of uncertainty.

2 Modeling service-user dynamic of cultural sites

2.1 Urban Modeling

The first models designed for urban modeling are mostly based on dynamical systems. One of the first famous approaches was proposed by I.S. Lowry [18] and J.W. Forrester [11] and consists in describing the city by global dynamics on socio-economic indicators. This kind of models are decomposed with stocks and flows. For instance, based on this approach, Dendrinos and Mulally [9] use a prey-predator model, assuming that the increase of city population makes the economic status decreasing. The predators represent the urban population and the preys, the per capita income. Like in other disciplines, during the last two decades, important efforts have been made to develop, beside these previous top-down models, decentralized methods called bottom-up models. In this new categories of models, we focus on local knowledge, like people behavior, or on spatial local information. The increase of information systems and specifically the accurate geographical databases included in Geographical Information Systems (GIS) allow to combine spatial management and local information, as a major contribution

to these decentralized models [21]. Moreover, intensive developments concerning diffusion in spatial process [15] had been achieved on cellular automata simulations for various purposes, facilitated by the development of computer power leading to grid computation of large dimension [10]. Since the last decade, geographical and social systems have seen their greatest development on modeling by the use of coupling GIS, cellular automata and multiagent systems, allowing this last concept to describe with accuracy the behavior of the inhabitants of these spaces and especially their social interactions [1]. A classical example of the need of extended cellular-based with agents comes from T. Schelling's model describing the segregation process [24]. More recently, the development of multiagent platform engineering [23], allows to integrate dynamical simulations within GIS [7]. The study presented in this paper is based on such technologies, mixing GIS and simulation within the platform Repast.

2.2 Collective intelligence models

Multiagent systems are a very promising concept for decentralized computations concerning wide range of problems solving modeled by distributed artificial intelligence approaches. Multiagent systems concept consists in describing a phenomenon by a set of autonomous entities, called intelligent agents, interacting in social way and interacting inside an environment [8]. If the interactions are complex, self-organization phenomena can appear or emerge as additional properties of the system interaction, not described from the entity behavior level. Such emergent high-level organizations coming from the low-level entity interaction system are frequently observed in human society but also in society of insects, each of them having a very simple behavior. To represent and to model such property transfer from individuals to systems, a specific classe of agent computing algorithms have been developed and lead to the concept of collective intelligence or swarm intelligence models. The current two major developments of swarm intelligence are (i) a class of methods based on the bio-inspiration of bird flocking or fish banks, the particle swarm optimization [16] and (ii) a class of methods based on the bio-inspiration of insect society, like bees or termites, called ant systems models [4]. The study presented in this paper is based on this second class category and is based on a combination of two algorithms, ant clustering and ant nest building [6].

2.3 Urban cultural dynamics modeling

The goal is to understand and to analyse the cultural practices in city. *Rouants* is a dynamical model for studying attractiveness of cultural sites in city. It is built on the achievements of the sociology on culture in order to identify the behavior of people depending on cultural sites. The model was applied to the design of cultural offer for the city of Rouen. We use swarm intelligence methods and stigmergy concept to model the dynamics of cultural center development. Specifically, we rebuild, by dynamical simulation, a multi-centre and multi-criteria system. The people dynamics, movement and decision, are inspired by self-organized processes existing in collective movements designed by swarm intelligence and stigmergy [4]. These processes correspond to the dialogue between individuals and environment for a collective benefit. The model includes a mixture of deterministic and probabilistic behaviors, corresponding to the part of unpredictability in human behavior.

2.4 Specification of multi-criteria multipole services-users dynamics model

Each center has several attractive criteria. Specific colors of pheromone are associated to these criteria and allow to model attractive mechanism according to age, gender educational level and social level of each individual.

Figure 2 is a schematic representation of an easy-to-understand simulation composed of two cultural centers, a cinema and a theater. We represent on this figure, the pheromone/attraction functions as circles around the centers and we label them with the associated characteristics/colors. Materials represent here users which will be carried and moved by ants, following the complex spatial system of attraction functions.

A spatial multi-criteria multi-center simulation is described by a set of centers and by a set of colors, each color corresponding to a specific criteria which can be expressed as a attractive function by a cultural center. Statistic values of national uses of cultural centers translate the capability for a center to attract a percentage of people belonging to specific class of criteria. This attractive phenomena are modeled by a colored template function emitted by each center. The mathematical expression of the attractive function of color c_j expressed by the center P_i, is given by

$$\Phi_{ij}(M) = a_{ij} \exp\left(-b_{ij}\left(d(M, P_i) - r_i\right)^2\right)$$

Where the amplitude a_{ij} is proportional to the capability of the center to attract people carrying the corresponding color/criteria; b_{ij} is the template slope, corresponding to the spatial diffusion of the attraction phenomenon; r_i is the radius defining the area of attraction of the center P_i. In the model, people are described by material including a table of characteristic colors which corresponds to the criteria that the individual is carrying.

The dynamics of users attraction is computed by (i) ant ability to pick or deposit material and by (ii) ant movement when carrying material.

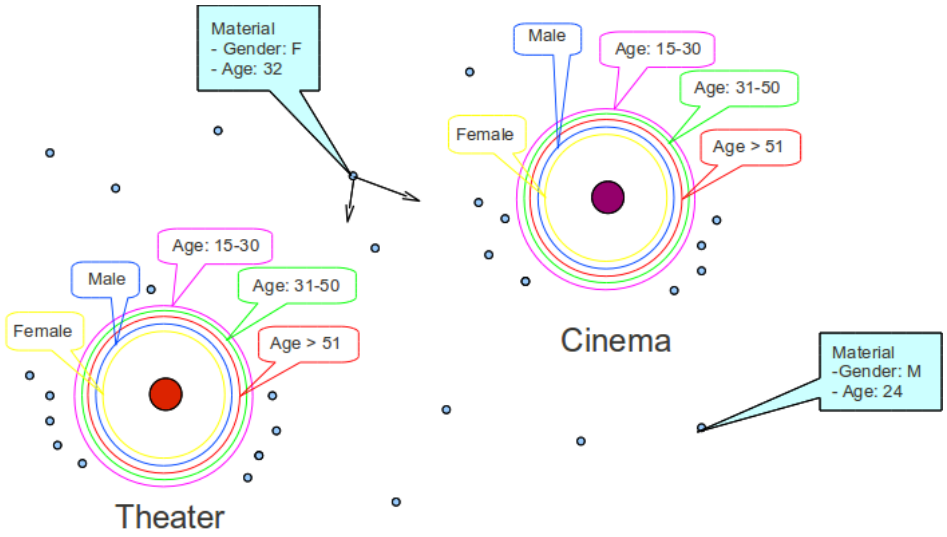


Figure 2: Cultural equipment dynamics modeling

Ant picking and deposit material is a composition of ant clustering computation under the influence of a template function Φ, that is:

- The probability for an ant, at the location M, to pick a material when it is not yet carrying one but are crossing one, is:

$$P_p(M) = (1 - \Phi(M)) \left(\frac{k_1}{k_1 + f} \right)^2$$

- The probability for an ant, at the location M, already carrying a material, to deposit it is:

$$P_d(M) = \Phi(M) \left(\frac{f}{k_2 + f} \right)^2$$

Where f is the material density perceived by the ant, k_1, k_2 are two thresholds.

The movement of an ant which is carrying a material is computed by a ranking process. The ranking computing corresponds to the weighted average of some selected pheromones. The selected pheromones are the ones whose color belongs to the table of characteristic color of the carried material [14]. This ranking process movement computation is alternated with random movements, allowing ants to explore the solution space and to not be blocked by other ants, producing in that way, some noise in the linear movement along the pheromone gradient.

Even if ants are used in this bio-inspired model, we don't consider that ant behavior is able to represent human behavior. Ant behavior is used to implement self-organization processes; human behavior is introduced in complex mechanism linking many criteria and specific knowledge expressed by statistical values of social practice described by the pheromone attraction function.

3 Modeling engineering and simulation platform

In this section, we first present a prototyp model built on a very simple spatial situation but easy to analyse. In the second subsection, we describe a more sophisticated simulation platform, integrating a detailed and specific GIS over the French city of Rouen.

3.1 Virtual urban area prototyp and first analysis

Figure 3 is describing Virtual urban area prototyp as an experimental configuration with seven centers and initially random places for the peoples/materials and the ants. Centers positions are described on the left top sub-figure. Each center is emitting heigh attractive pheromone functions associated to a color labeled from 0 to 7. On figure 3, we also represent the amplitude of the colored pheromone functions for the centers labeled from 0 to 2.

On figure 4, we show the result of one simulation where ants progressively aggregate the materials around the centers, following pheromone trails and clustering algorithm. On the left top sub-figure, we see the initial distribution of materials and ants. In the other sub-figures, we see successive steps of simulation. We can observe the formation of material affectation to each center in order to respect the attraction process, according to the material characteristics. In [14], the analysis of the attraction phenomena is detailed, allowing to better understand how the computation generates self-organization and distributes the material between the centers.

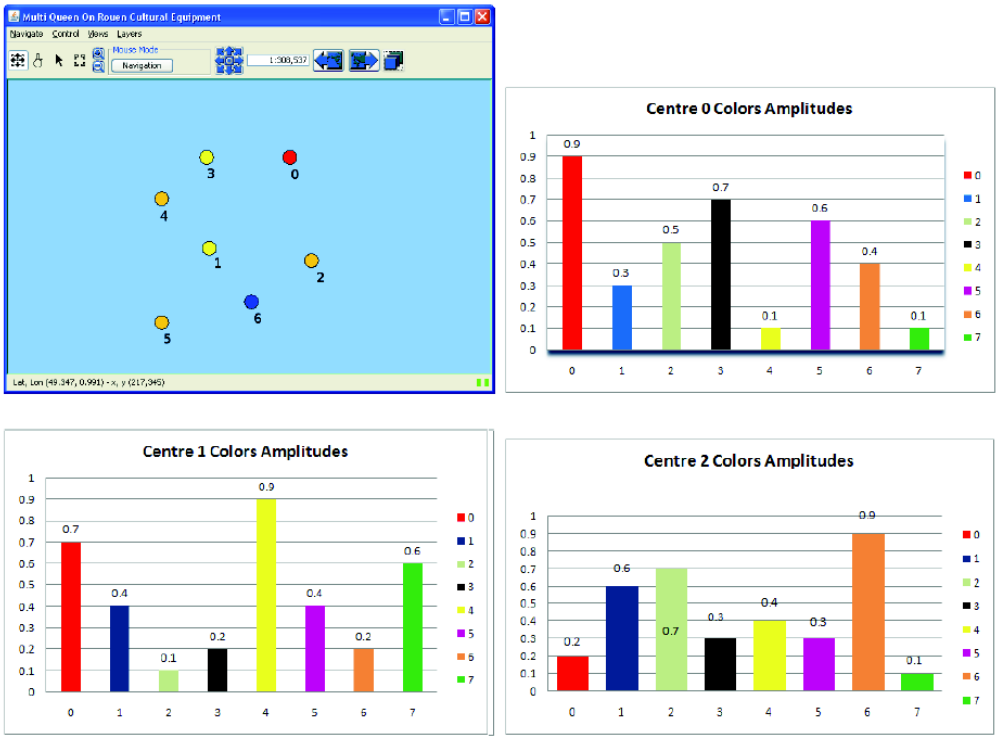


Figure 3: Virtual urban area prototyp on Repast: initial configuration based on seven centers, each of them emitting heigh attractive criteria represented by colored pheromones

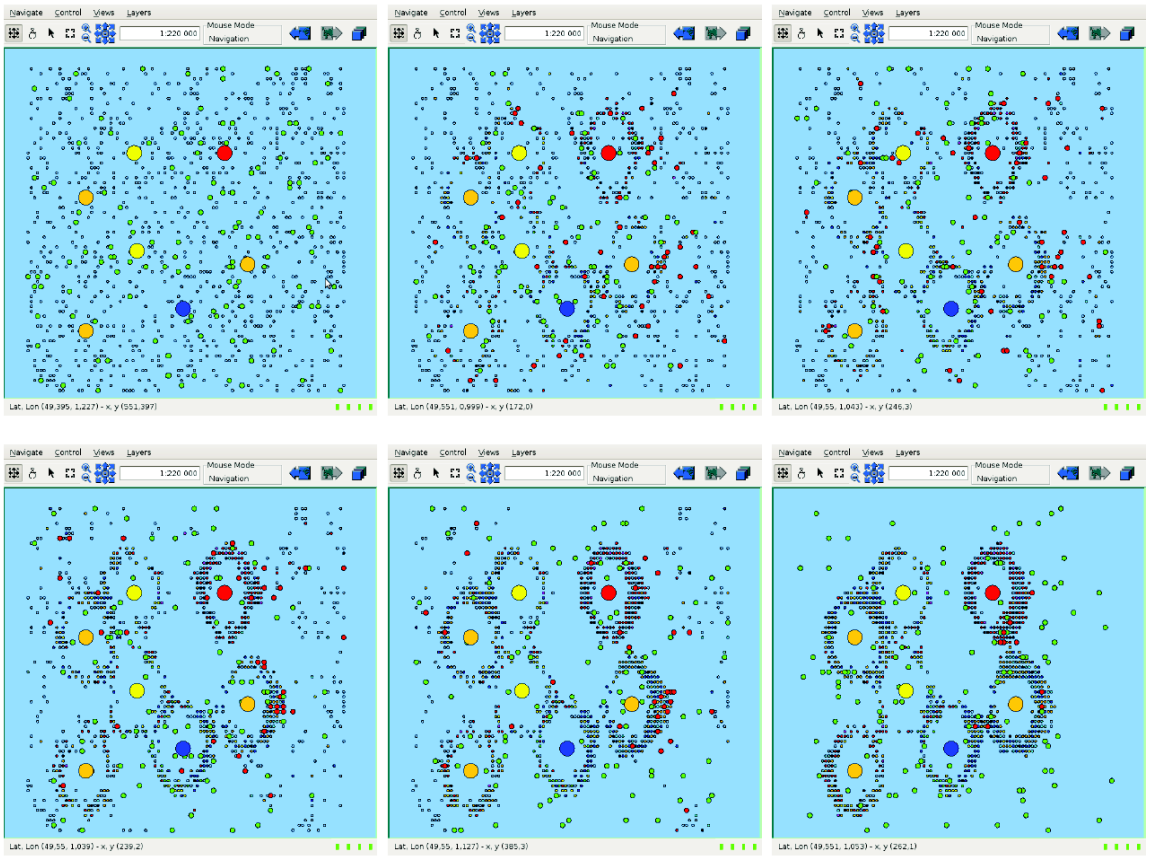
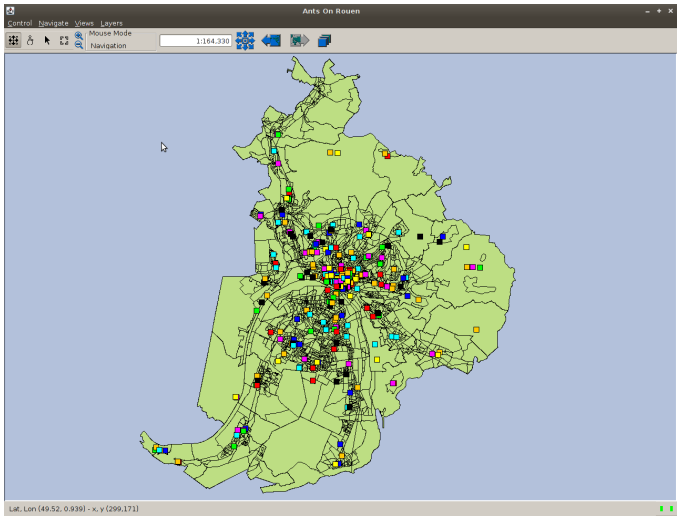


Figure 4: Virtual urban area prototyp simulation computation on Repast, at successive steps from top to bottom and left to right: iteration 0, iteration 152, iteration 250, iteration 370, iteration 601, iteration 1601.

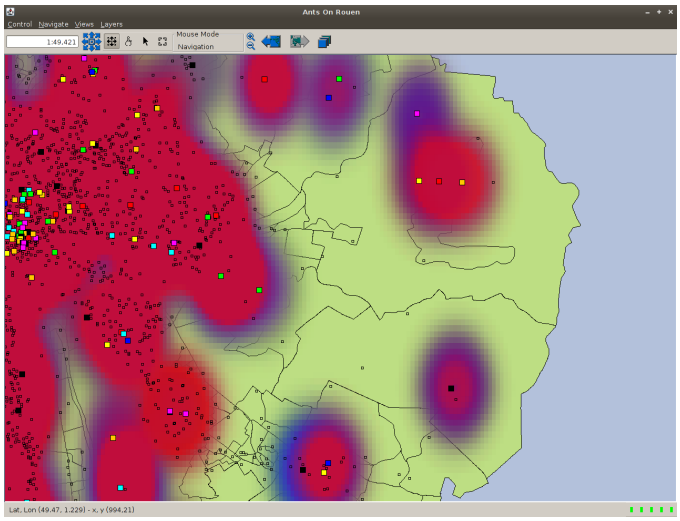
3.2 *ROUANTS* simulation platform

Rouants simulation platform allows to dynamically experiment the complex interrelationships mixing the three types of actors previously described: social individuals, cultural sites, urban spatial environment. This framework implements decentralized approaches and so is devoted to built self-organized mechanisms of collective cultural behavior. The results of the simulation allow to reveal the emergence of spatio-temporal forms from the interleaving of multiple types of interactions and from the spatial constraints coming from urban system[13, 14, 23].

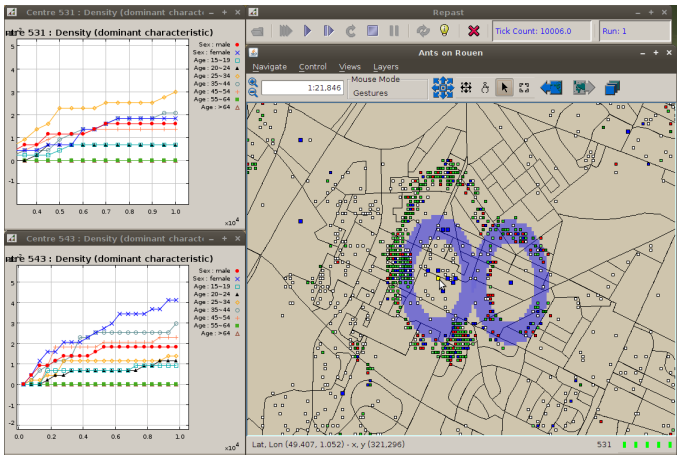
Rouants implements the previously described algorithms of collective intelligence and integrates them inside a specific GIS, representing the cultural sites and the block housing over the urban area of Rouen in High-Normandy. Figure 5 shows different visualizations produced by this platform. On the sub-figure (a), we see the whole distribution of cultural sites and centers over the urban area; the sub-figure (b) is a sumple vue allowing to observe how the pheromones are spreading over the space; the sub-figure (c) described an analysis of some specific places in this agglomeration, tracing on the graphes on the left side of this sub-figure, the temporal series of number of materials being attracted in two zones, according to the dominant color/criteria of the materials.



(a) Rouen Cultural equipments GIS



(b) pheromone density visualization



(c) result output on GIS and graphics

Figure 5: Visualization of *ROUANTS* simulation platform

4 Conclusion and Perspectives

We illustrate in this paper the possibility of experimenting dynamics of multiple services-users in city. Rouants simulation platform is an effective tool for approaching cultural practices in urban situations. Decentralized mechanisms reinforce the suitability of the type of data and methods used to the study of socio-spatial complex systems like cities. We observe emergence of multiple cultural aggregation phenomena in relation to the sociology of culture and to the specific urban and social environment of the city of Rouen. The simulation highlights territorial situations of rivalry and complementarity in cultural practices. The modeling concept of the platform allows also to experiment adaptive capacity and feedback process like politics of regulation of cultural centers, according to the user practices. The social and territorial impacts of these mechanisms are one of our further perspectives and analysis. There is also a general evidence: the use of cultural heritage is a vector for both cultural policy and cultural development.

Acknowledgements

This work is part of the multidisciplinary project RISC (French acronym for Complex Networks and Complex Systems), supported by the French Ministery of High-

Education, the Region High-Normandy and its Research Network on Transportation, Logistics and information Processing (GRR TLTI) and by European Union with FEDER grants. With this support, a part of the final implementation of the framework *Rouants* has been made by Florent Marchand de Kerchove. Docteur Rawan Ghnemat was granted by a French Government Scholarship.

References

[1] I. Benenson, P.M. Torrens (2004) *Geosimulation - Automata-based modeling of urban phenomena*, Wiley.

[2] M. Batty, Y. Xie 1994) *From cells to cities*, Environment and Planning B, 21, pp 531-548.

[3] C. Bertelle, G.H.E. Duchamp, H. Kadri-Dahmani (eds), *Complex Systems and Self-Organization Modelling*, "Understanding Complex Systems" series, Springer, 2009.

[4] Bonabeau, E., Dorigo, M., Theraulaz, G. (1999) *Swam Intelligence, from natural to artificial systems*, a volume in the Santa Fe Institute Studies in the Sciences of Complexity, Oxford University Press.

[5] P. Bourdieu, A. Darbel (1969) *L'amour de l'art, les musées d'art européen et leur public*, 255 p., Les Editions de Minuit, Paris.

[6] S. Camazine, J.-L. deneubourg, N.R. Franks, J. Sneyd, G. Theraulaz, E. Bonabeau (2001) *Self-Organization in Biological Systems*, Princeton University Press.

[7] A.T. Crooks (2007) *The Repast simulation/modeling system for geospatial simulation*, Technical reportr, UCL Working Papers Series, Paper 123.

[8] E. Daude, P. Langlois (2006) *Introduction à la modélisation multi-agents des systèmes complexes en géographie*, in F. Amblard and D. Phan (eds) *Modélisation et simulation multi-agents*, Lavoisier, Hermès Science, Paris, pp 357-441.

[9] D. Dendrinos, H. Mullaly (1985) *Urban Evolution: Studies in the Mathematical Ecology of the Cities*, Oxford University Press.

[10] E. Dubos-Paillard, Y. Guermond, P. Langlois (2003) *Analyse de l'évolution urbaine par automate cellulaire. Le modèle SpaCelle*, L'espace géographique 32 (2003/4), pp 357-378.

[11] J.W. Forrester (1969) *Urban Dynamics*, Cambridge, MA, MIT Press.

[12] C. Fuchs (2003) *Concepts of Social Self-Organization*, Research paper INTAS Project "Human Strategies in Complexity".

[13] R. Ghnemat, C. Bertelle, G.H.E. Duchamp (2007) *Adaptive Automata Community Detection and Clustering, a generic methodology*, in *Proceedings of World Congress on Engineering 2007, International Conference of Computational Intelligence and Intelligent Systems*, pp 25-30, London, U.K., 2-4th July 2007.

[14] R. Ghnemat (2009) *Adaptive modeling for spatial emergence within complex system*, Ph.D. in Computer Science, University of Le Havre, France.

[15] T. Hagerstand (1953) *Innovation Diffusion as a spatial process* (translated in English by Allan Pred, 1967) Lund, Sweeden, University of Chicago Press.

[16] J. Kennedy, R.C. Eberhart (1995) *Particle Swarm Optimization* In *Proceedings of IEEE International Conference on Neural Networks (Perth, Australia)*, IEEE Service Center, Piscataway, NJ, 5(3), pp 1942-1948.

[17] J.-L. Le Moigne (1999) *La modélisation des systèmes complexes*, Dunod.

[18] I.S. Lowry (1964) *A model of metropolis*, Santa Monica, CA, the RAND Corporation: 136.

[19] Lucchini, F. (2010) *La fabrique des lieux culturels*, HDR, University of Rouen, France.

[20] F. Lucchini, W. Hucy (2007) *L'image de la ville* in *Rouen, métropole oubliée*, Y. Guermond (ed), l'Harmattan.

[21] F. Lucchini (2006) *Un SIG pour la culture : un instrument de réflexion sur la localisation des équipements culturels dans une agglomération* in Géographes Associés 30, pp 237-243.

[22] F. Lucchini (2002) *La culture au service des villes*, collection Villes, Anthropos, Economica.

[23] Repast web site (2008) `http://repast.sourceforge.net`.

[24] T.C. Schelling (1971) *Dynamic Models of Segregation*, Journal of Mathematical Sociology, Vol. 1:143-186.

[25] D. Unwin (1981) *Introductory Spatial Analysis*, Methuen and co Ltd, London & New-York.

Part IV

Complex Systems and Methods

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

TOWARDS A SWARM OPTIMIZATION ALGORITHM WITH "LOGISTIC" AGENTS.

Rodolphe Charrier *

Abstract. Adaptation and self-organization are two main aspects of swarm mechanisms and collective intelligence. We adress here the global question of controlling the coordination of groups of mobile agents in order to achieve optimization processings. We study in this paper a simple case involving "logistic" agents –whose internal decision is governed by a logistic map, that is a discrete parametrized quadratic map– slaved to a stochastic environment through their control parameter. The proposed algorithm enables agents to find local minima in the environment. We show that the adaptation process on the control parameter leads to this local optimization. Applications may be envisaged for multi-objective problems.

Keywords. Swarm Intelligence, Reactive Multi-Agent System, Multi-objective Optimization, Logistic Agent, Adaptation, Self-organization, Decentralized Control.

1 Introduction

The root principle of Collective Intelligence consists in considering intelligence no more as an individual characteristic only, but also as an emergent phenomenon involving (self-)adaptation and self-organization processes. This type of intelligence may be massively distributed and seems very robust to perturbations and failures. Research on Collective Intelligence has provided efficient meta-heuristics for approximating large sets of problems, notably in the Swarm Intelligence field [3]. The challenge is to understand and model properly the involved mechanisms so as to build appropriate and efficient algorithms.
The majority of proposals in modeling Swarm Intelligence in particular, is based on specific probability laws used as decision functions like Ant Colony Algorithms for instance [10], or based on randomized algorihtms like Particle Swarm Algorithms [11].

We address in this paper the challenging problem of controlling the collective motion within a reactive multi-agent system to achieve optimization processings. In this way, we have followed the Particle Swarm Algorithms evolution by focusing first on flocking-like phenomena. Flocking has a great interest firstly because of its importance in nature –many species adopt such a coordinated form of collective motion–, and secondly because it may have many applications in complex network studies such as social networks, ad hoc networks, swarm robotics, ... The two most used modeling frameworks for achieving flocking simulations are currently Reynolds's approach [1] and Vicseck's model [2].

Reynolds has based his approach on three main rules –collision avoidance, velocity matching and flock centering– without giving any mathematical specifications. On the contrary, Vicseck has expressed an explicit dynamical system. Both derives from particle systems and provide a synchronization process: Reynolds uses a velocity matching rule while Vicseck involves the average of the neighbors' velocity angles to compute the velocity angle of the current particle.

Our approach to tackle the flocking phenomenon is quite different. On the one hand, our mobile entities are not particles but agents. This enables to distinguish the geometrical/kinematic characteristics from the internal state of entities. The encapsulated internal state gives to agents an autonomy of decision for any type of actions, not only for moving. On the other hand, we stress on the synchronization processes, which must be involved in the internal decision making to our opinion. Our main working hypothesis is that many swarm phenomena may be described and analyzed in terms of oscillator synchronization and decentralized control, in the spirit of [8]. In that way we have proposed a particular model which is a reactive multi-agent system (MAS) model called Logistic Multi-Agent System (LMAS) [5], because of the internal logistic map governing the decision process of each agent. It refers also to the Coupled Map Lattices (CML) family of computational models, particularly the ones involving logistic maps [7]. Logistic maps are one-dimensional deterministic quadratic maps controlled by a single parameter, which can generate chaotic series. These maps are very interesting because of their wide range of dynamical behavior –from complete randomness to single fixed points–, and their facility to be coupled and controlled. This specific dynamical behavior enables to use such a deterministic series generator instead of

*Rodolphe Charrier are with LITIS, University of Le Havre, France. E-mails: rodolphe.charrier@logistic-swarm.net

a probabilistic one. This change opens a new research perspective which constitutes the heart of our studies.

In this paper, we firstly recall the ability of the LMAS to produce flocking simulations by synchronization caused by individual couplings. This constitutes an introduction for the main case we want to tackle. We intend to show how one can apply the self-organization and adaption capabilities of the LMAS to find minimum values in a stochastic environment. A decentralized and adaptive control mechanism is used to achieve the motion coordination and the convergence towards the local minima. This mechanism is based on control variations of the logistic map, caused by the local agents' perceptions of random data stored in the environment. Simulations show the agregation of agents on the same minimum field locations. The discrete dynamical system theory enables to partially explain this resulting phenomenon. In particular, we stress on the fact that the whole group of agents converges to a decision fixed point. This also shows that dynamical modeling approaches are appropriate for analysing phenomena and for monitoring the dynamics.
The paper is organized as follows. The first section recalls the formulation of the LMAS model. We then some collective motion results by showing how a flocking may appear within a LMAS in a first step. In a second step, we explain the algorithm for finding local minima in a data field of the environment. Then we discuss results and conclude this work.

2 The Logistic Multi-Agent System (LMAS)

This section begins with a short description of the model of coupled map lattice, which is the underlying mathematical basis of the LMAS model. The global coupled instance of the model is presented. We then show how we have turned it into a multi-agent system.

2.1 A CML foundation

CML models are used by physicists to study spatiotemporal chaos phenomena in the field of hydrodynamics (simulation of turbulent flows) or condensed matter physics, and more generally in theoretical physics, as a computational model. Our interest focuses on CML using nonlinear quadratic maps, namely logistic maps. These types of CML have been widely explored by Kaneko since the 80's [7].
A CML is a discrete time and space cellular model in which cell states x take their values in a continuous domain. A CML is sometimes considered as a cellular automaton with continuous states. Let us focus on a mean-field instance of the CML class, namely the globally coupled map lattice (GCM) [6]. A GCM is governed by the master equation:

$$x_i^{t+1} = (1-\epsilon) f(x_i^t) + \frac{\epsilon}{N} \sum_{j=1}^{N} f\left(x_j^t\right) \tag{1}$$

where N is the total site number in the lattice, x_i^t is the state variable of the cell on site i at time t and ϵ the diffusive coupling coefficient. CML and GCM have been widely studied when f is the well-known logistic map usually defined on the interval $[0,1]$ ($[0,1]$ is invariant through f) by the following recurrent equation:

$$x^{t+1} = f(x^t, a) = 4\ a\ x^t\ (1 - x^t) \tag{2}$$

where $a \in [0,1]$ denotes the control parameter of f. We intentionally use the notation $f(x,a)$ to stress on the importance of the control parameter a in our case studies. This map is known to generate chaotic series as t tends to ∞, in particular if $a = 1$, or to converge to some fixed points or periodic cycles in function of the value set for a.
The asymptotic behavior of this nonlinear map is completely described by its bifurcation diagram on Fig. 3, which plots $x^\infty = f^\infty(x^0)$ as a function of the control parameter a.

Shibata and Kaneko have anticipated the evolution of this model to fit with the Swarm Intelligence field. They proposed a specific CML instance called the CML gas in 2003 [9]. A CML gas is a CML-like model in which chaotic cells are motile on the lattice. But to our knowledge, there was only one paper about this research perspective. The LMAS may be considered as an agent based evolution of the CML model specifically designed for Swarm Intelligence. The next section describes this evolution.

2.2 Conception of the LMAS

Contrary to CML, a multi-agent system distinguishes the group of agents from the environment. The environment is the space where agents can perceive data and make actions (move for instance). Consequently, each agent encapsulates its internal state, which enables autonomy in acting.

Definition 2.1. The logistic multi-agent system (LMAS) is composed of N agents called logistic agents. A logistic agent is a reactive agent whose internal decision is governed by a logistic map.

The internal state s of a logistic agent is composed of three variables corresponding of the three quantities of CML:
$s = \langle x, a, \epsilon \rangle$

- x is the decision variable computed by the logistic map characteristic of the agent

- a is the control variable of the agent, which is also the control parameter of the logistic map. It governs the type of dynamical series generated by the map given by the bifurcation diagram in Fig. 3.

- ϵ is the coupling variable of the agent which quantifies the strength of the influence of its neighbors' decision.

All these three variable have values in the real interval $D = [0, 1]$. The logistic agents can locally perceive their environment. In the most general case, each of these quantities can be modified by the environment through perception functions, that is why they have a variable status since they may change with time.

Here is summarized the decision-making of the agent i in one master equation derived from the formula 1:

$$x_i^{t+1} = (1 - \epsilon_i^{t+1}) f(x_i^t, a_i^{t+1}) + \frac{\epsilon_i^{t+1}}{|V_i^{t+1}|} \sum_{j \in V_i^{t+1}} f(x_j^t, a_j^{t+1}) \tag{3}$$

where all parameters may depend on time. The influence of the other agents is reduced to the current neighborhood V_i of the agent i, which contains $|V_i|$ other agents at the current time. This master equation gives the rooting principle of the model and the scheduling of the computation: all quantities are computed before the decision variable x during the perception process. This equation corresponds to the decision process.

The action process consists in using the decision variable x to perform actions like moving or updating local data in the environment (a pheromone field for instance). Finally, the computation process is clearly divided in three parts — perception-decision-action — which is coherent with a cybernetic approach. In the next section, we apply this model to collective motion simulations.

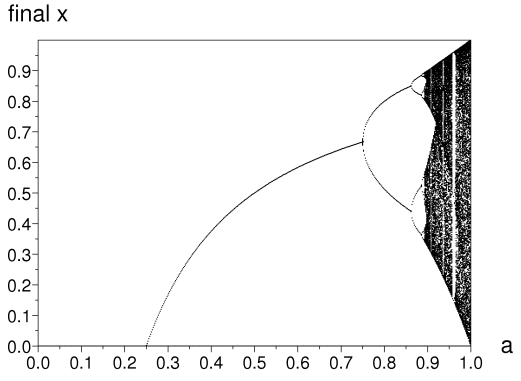


Figure 1: Bifurcation diagram for the logistic map $x_{t+1} = 4\ a\ x_t\ (1 - x_t)$, calculated with 500 iterations on 500 samples of the interval $[0, 1]$

3 LMAS based simulations

3.1 The flocking phenomenon

In this subsection, we describe a very simple implementation of the LMAS which leads to flocking simulations. The root principle is to interpret the decision variable x of the agents as a moving angle in the geometrical space. This case shows a simple example of the LMAS self-organization capabilities resulting from the synchronization processes which occur between the internal states of agents. The action phase of moving will "express" this synchronization in the environment space as a flocking phenomenon, that is a grouping of coordinated agents. This implementation has been completely presented and the synchronization process analyzed in [4].

Let us shortly describe the specifications of the modeling for this case:

- The state of the agent i at time t turns to the tuple: $s_i^t = \langle x_i^t, a_i, \epsilon \rangle$

- a_i and ϵ are constants in the current case.

- The value of a_i is randomly chosen in $[0, 1]$ for each agent, that is each agent may have any type of dynamical behavior according to the bifurcation diagram 3.

- The coupling variable ϵ is a global constant as well, but may be set with various values initially.

- The environment is a discrete 2D torus where agents are located randomly initially.

Let us comment the perception-decision-action processes for the agent i:

- The perception gives the second term of equation 3: $P_i = \frac{\sum_{j \in V_i^{t+1}} f(x_j^t, a_j)}{|V_i^{t+1}|}$

- The decision is calculated by means of equation 3).

- Then the moving action consists in updating the direction of the velocity (cf. fig 2) according to the updated decision value:

$$\begin{cases} \alpha_i^{t+1} & = 2\pi\ x_i^{t+1} \\ \mathbf{r}_i^{t+1} & = \mathbf{r}_i^t + v_0\ \mathbf{u}_{\alpha_i^{t+1}} \\ \theta_i^{t+1} & = \arg(\mathbf{r}_i^{t+1}) \end{cases} \tag{4}$$

gives the new velocity direction of agent i in a simple way.
The velocity magnitude remains a constant here.

Thanks to this modeling, we can control the move in a $2D$-space with a $1D$-quantity only, which is an important feature of this model.

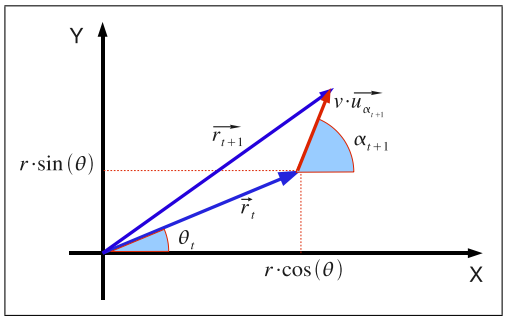


Figure 2: Description of the agent's motion in cylindrical coordinates.

The final master transition equation for the decision x is summarized in the following expression for the agent i:

$$x_i^{t+1} = (1-\epsilon) f(x_i^t, a_i) + \frac{\epsilon}{|V_i^{t+1}|} \sum_{j \in V_i^{t+1}} f(x_j^t, a_j) \quad (5)$$

Figure (3) shows a snapshot of such a flocking simulation involving 50 agents whose initial control has been chosen randomly in $[0,1]$ and with $\epsilon = 0.96$. The environment is a toric grid of 100×100 cells. The velocity magnitude is set to 1.0. Simulations lead to the formation of clusters of partial synchronization which gives the shape of flocks. These clusters are quite unstable: they split as soon as they cross the path of another cluster.
Another type of flocking simulation consists in involving only chaotic agents, namely with $a = 1.0$ in each agent. We let the reader refer to the paper [4] to get the whole analysis on the synchronization process for this type of flocking. The following case study is founded on the same type of formulation and dynamical mechanisms.

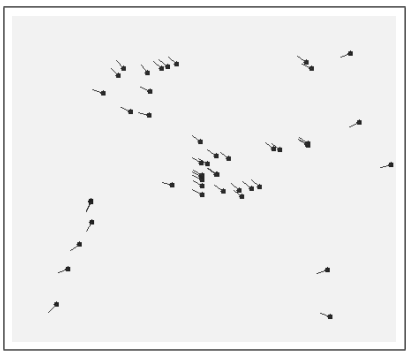


Figure 3: Snapshot of a flocking simulation: $n = 50$, $\epsilon = 0.96$, radius of neighborhood= 20

3.2 Towards optimization

3.2.1 Control by adaptation

In the preceding flocking simulation, we stressed on the group of agents: each agent was influenced by others in a limited neighborhood and no external data stepped in. Now we intend to show a specific effect of the influence of initial data marked down in the environment. Whereas the a-variable was a constant in the preceding case, it is now modified through a perception function. The aim for the agents is now to find local minima of data in the environment by stopping at the corresponding locations. This aim is inspired from some optimization problems in the Particle Swarm Optimization (PSO) domain, where particle final positions indicate the best solutions for finding the zero values of a positive nonlinear function [12].
Let us summarize the main changes compared to the preceding situation:

- The state of the agent i at time t turns to the tuple: $s(t) = \langle x_i(t), a_i(t), \epsilon \rangle$
  a depends now on time. As before, ϵ remains a constant chosen initially.
- The environment holds now a discrete field denoted $\mathcal{D}$, which maps a random value in $[0,1]$ onto each cell of the environment —let us recall that the environment is a discrete grid here—. These data are persistent in the field, that is they are not modified by agents. The field values are indicated in our simulations by color levels from light gray for the value 1 to full white for the value 0.
- The perception function for the a component is quite simple: an agent can read local data in the $\mathcal{D}$ field and returns the mean value of these data in its perception vicinity. The updating proceeds in a convex coupling with the same coupling coefficient ϵ.
- Moving and updating actions:
  - as before the new direction of the velocity is set by the updated decision $x(t+1)$,
  - the velocity magnitude changes as well in function of the decision x according to the updating equations:

$$\begin{cases} \alpha_i^{t+1} & = \alpha_i^t + 2\pi\delta\ (0.5 - x_i^{t+1}) \\ v(t+1) & = x(t+1) * v_0 \\ \mathbf{r}_i^{t+1} & = \mathbf{r}_i^t + v_0\ \mathbf{u}_{\alpha_i^{t+1}} \end{cases} \quad (6)$$

  where v_0 is the initial velocity magnitude and δ is a coefficient on angle variations.

The transition system for the agent's internal state may be written as before, added with the updating of the control variable a:

$$\begin{cases} a_i^{t+1} & = (1-\epsilon) a_i^t + \epsilon\ \langle \mathcal{D} \rangle_{V_i^{t+1}} \\ x_i^{t+1} & = (1-\epsilon) f(x_i^t, a_i^{t+1}) + \\ & \quad \frac{\epsilon}{|V_i^{t+1}|} \sum_{j \in V_i^{t+1}} f(x_j^t, a_j^{t+1}) \end{cases} \quad (7)$$

where $\langle \mathcal{D} \rangle_{V_i^{t+1}}$ denotes the average operator on the perception neighborhood V_i^{t+1} applied on the data field $\mathcal{D}$.

The control variable is computed before the decision variable since it is needed for the logisitic map calculation. We can easily verify that a_i always belongs to $[0, 1]$.

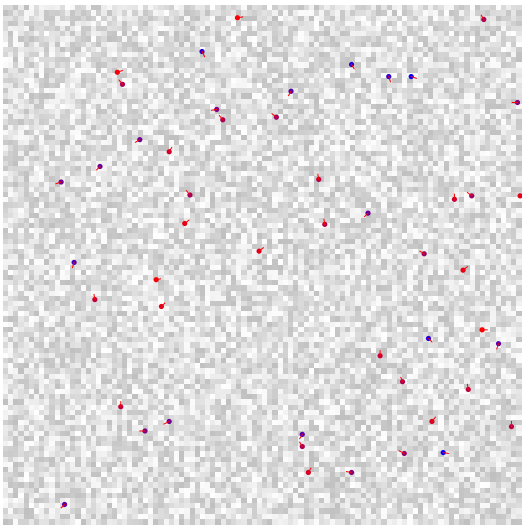
3.2.2 Simulation and results

Simulations have been settled as follows:

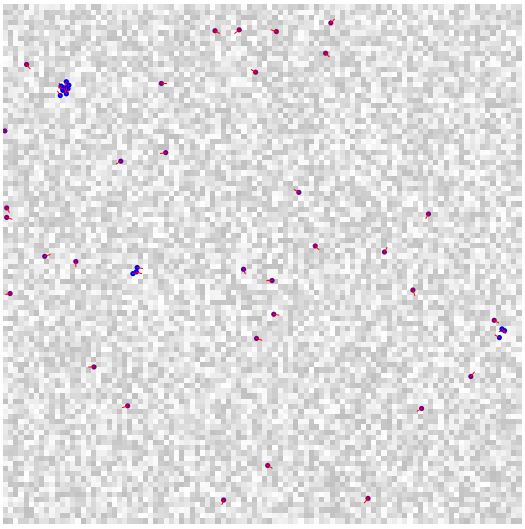
Environment	100×100 torus
Data field $\mathcal{D}$	100×100 grid
Weighted cells	at random in $[0, 1]$
Agent set	$N = 50$
Radius of perception	$R = 1.0$ ($\simeq 8$ cells)
Initial velocity magnitude	$v_0 = 1$
Angle increment	$\delta = \frac{1}{20}$
Initial agent's decision x	at random in $[0, 1]$
Initial agent's control	$a_i^0 = 1$
Global coupling	$\epsilon = 0.3$

The figure (3.2.2) shows three snapshots taken at different evolution stages of a simulation leading to three main agregation points. The agents' coloring reveal their internal decision making: from light red for a chaotic behavior to light blue for a fixed point behavior. At time $t = 0$, every agent has a different velocity direction. Then at time $t = 2000$, three agregation areas begin to emerge. At time $t = 3700$ these agregation groups are definitively fixed and correspond to local minima given in the table (5(c)). Then each agent will progressively join one of these groups untill there are no moving agents left. This is caused by the attraction effect of the agregation of many agents on the same location: each agent extends the attraction area with its coupling potential.

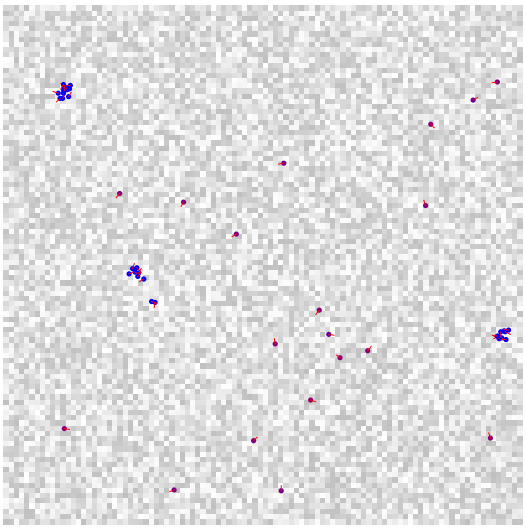
The figures (5(a)) represents the chart of the internal control (in red) and decision (in blue) variables of a randomly chosen agent. Its perception values are also plotted in green, showing the range of variations perceived in the environment. All these series have a mean value of 0.5 at the beginning, that is why the updating equation for the velocity angle (6) involves this particular value. The decision variable x becomes therefore the amount of variations around the initial velocity angle. This moving regime continues untill a minimum value or an agregation group is perceived. Then the agent's decision converges dramatically towards zero, as it is shown in the figure 5(b) at time $t = 3250$. This evolution is qualitatively the same for each agent, depending on the agregation point.



(a) t=0



(b) t=2000

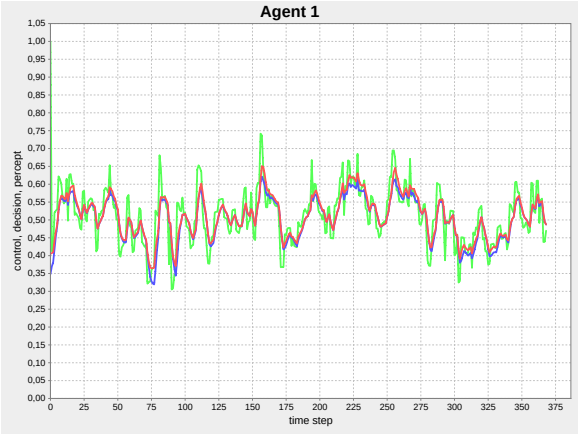


(c) t=3700

Figure 4: Snapshots of the simulation at different time steps.

4 Discussion

Now we have to comment on the simulation results.
First of all on the convergence process. To explain this convergence, we must look again at the bifurcation diagram of the logistic map (3). As long as the agent's perception function returns values close to the mean 0.5, the



(a) Variations of control (in red), decision (in blue) and perception (in green) with time for agent 1.

(b) Dramatic convergence to a local minimum.

Group 1	minima: 0.02, 0.06
Group 2	minima: 0.008,0.02
Group 3	minima: 0.04, 0.06

(c) Values of the minima detected in the field by each group.

Figure 5: Variations of the internal agent's variables and detection of local minima.

control parameter stays as well in this range of values and so does the internal decision variable of the agent. This is reinforced by the fact that the point $(a = 0.5, x = 0.5)$ is a superstable fixed point of the bifurcation diagram, that is $f'(\frac{1}{2}) = 0$ for the logisitc map. In other words, the variations on a are systematically counterbalanced by the induced variations on x.
But as soon as the perception operator –which is an average operator on about 8 cells– returns values close to 0.25, the agent's decision can converge rapidly towards the fixed point $(a = 2.5, x = 0)$. At this stage, the considered agent does not move anymore because of the moving laws (6), and can not leave this point, except if another agent is sufficiently near to draw it away by the coupling strength.

The second comment is about the other parameters in the simulation. In particular, we may wonder how the tuning on the radius of perception or on the global coupling parameter impacts the performances. This study has not been completely achieved so far. Some testings suggest tendencies:

- the perception neighborhood has to be as smaller as possible to capture the best minima,

- the coupling coefficient is the same for the coupling of internal agents' decisions and for the updating of the internal control parameter (cf. equations (7)). These two couplings are antagonist at first: while the control coupling tends to stop agents rapidly on field sites containing low values, the decision coupling can produce perturbations for already stabilized agents. This antagonism prevents to converge too soon toward unsatisfactory locations. The tuning of this parameter seems therefore essential for the quality of the convergence. Another way we do not have explored yet, consists in tuning the two coefficients separatly so as to give more tuning possibilities.

A last comment may be made on the quality of the minima found. At this stage of the study, our objective is not the absolute performance in the results but the proof of the concept. To improve performances, we have to compare the proposed algorithm with particle swarm approaches on benchmark problems [13].

5 Conclusion

This paper has proposed an algorithm based on the logistic agents' model which derives from the CML model, so as to tend to a multi-objective problem solving. After recalling the genesis of our logistic multi-agent system, we have applied this model to a small and basic problem of finding local minima in a random field within a toric environment. Although we can not conclude to the design of a new swarm optimization algorithm at this stage, we have laid its foundations.

The main interest of our approach lies in its nonlinear dynamical system theory, which provides strange but usefull mathematical maps, like the logistic map. We have exploited the interesting properties of this map, notably by linking the control parameter of the map to the data stored in the environment. The agents' decision process is therefore indirectly controlled and governed by the environment. It is also quite simple to explain by means of the bifurcation diagram of the logistic map.
This algorithm has to be improved to be more efficient to detect the best local minima. Testing on benchmark problems could be a future development strategy. Transforming the algorithm for high dimensional spaces may be a future prospect as well.

References

[1] C. W. Reynolds,"Flocks, herds, and schools: A distributed behavioral model", *Computer Graphics*, 21(4):25–34, 1987.

[2] T. Vicsek, A. Czirok, E. B. Jacob, I. Cohen, O. Schochet, "Novel Type of Phase Transitions in a System of Self-driven Particles", *Physical Review Letters*,75:1226–1229, 1995.

[3] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence. From natural to artificial systems*, Springer-Verlag, Berlin Heidelberg New York, 1999.

[4] R. Charrier, C. Bourjot, and F. Charpillet, "Flocking as a synchronization phenomenon with logistic agents", *ECCS'07, European Conference on Complex Systems, Dresden*, Dresden, October 1-5 2007.

[5] R. Charrier, F. Charpillet, and C. Bourjot, "A Nonlinear Multi-agent System designed for Swarm Intelligence : the Logistic MAS", *SASO'07, International Conference on Self-Adaptive and Self-Organizing Systems, Boston*, 2007.

[6] K. Kaneko, "Clustering, coding, switching, hierarchical ordering, and control in network of chaotic elements", *Physica D*, 41:137–172, 1990.

[7] K. Kaneko, *Theory and applications of coupled map lattices*, John Wiley & Sons, 1993.

[8] S. C. Manrubia, A. S. Mikhailov, and D. H. Zanette, *Emergence of Dynamical Order*, volume 2 of *World Scientific Lecture Notes in Complex Systems*, World Scientific, 2004.

[9] T. Shibata and K. Kaneko, "Coupled map gas: structure formation and dynamics of interacting motile elements with internal dynamics", *Physica D*, 181:197–214, 2003.

[10] M. Dorigo, C. Blum, "Ant colony optimization theory: A survey", *Theoretical Computer Science*, 344:243–278, 2005.

[11] J. Kennedy, R. C. Eberhart, "Particle swarm optimization", *Proc. of the IEEE Int. Conf. on Neural Networks*, 1942–1948,1995.

[12] R. C. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory", *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, 39–43,1995.

[13] K. De Jong, "An analysis of the behaviour of a class of genetic adaptive systems", *PhD Thesis*,University of Michigan,1975.

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

EXPLORATORY ANALYSIS OF WEB DATA: METHODS, TOOLS AND GEOGRAPHICAL DISTRIBUTION

Fabien Pfaender, Nicolas Esposito*, Mathieu Jacomy†

Abstract. We propose a methodology for the exploration of Web data, built on the principles of exploratory data analysis and analytical visualisation of information. Our approach aims at combining these two approaches in order to benefit from both of them. This allows us to explore heterogeneous complex dynamic systems such as the Web, and to construct emergent structures and indicators without getting lost. By studying the geographical dimension for a specific Web locality, which is exemplary in many ways, we were able to test our methodology and various visualisation tools, thus validating our theoretical proposals.

Keywords. Exploratory Data Analysis, Web, Geography, Information Visualisation, Methodology

1 Introduction

At a time when we are faced with systems which are becoming ever more complex, dynamic and heterogeneous, we propose a method for engineering the emergence of structures and indicators in complex systems. This engineering comprises a set of original methods and tools which are able to extract the relevant properties. This should enable us to explore any system whatsoever; the Web, which is the object of the first part of this paper, is a prototypical example. As developed in sub section 2.1, constituents of this complex system are not all discovered and known and their interactions are subject to a lot of analysis without leading to an end even if some emergent properties seems to emerge[10]. At the heart of the Web, the geographical dimension is in full boom, and contributes a means for users to anchor themselves in reality while increasing the possible uses. However it is not simple to mobilize the spatial register on the Web, especially when this dimension has not been originally conceived for it. It is then necessary to give birth to meaning, by playing on whatever geographical references lie to hand (geolocalisation of resources, positions of actors, spatial references in the content, etc.); this is exemplary of a heterogeneity which the method we propose makes it possible to deal with, both theoretically and practically, as we shall see in sections 3 and 4.

*Fabien Pfaender and Nicolas Esposito are with Department of Cognition Research and Enactive Design, Compigne University of Technology, France. E-mails: fabien.pfaender@utc.fr, nicolas.esposito@utc.fr

†Mathieu Jacomy is with Department médialab, SciencesPo, France. E-mail: mathieu.jacomy@gmail.com

2 Exploring the Web

The figures that are used to characterize the Web are enough to make one dizzy. It is a question of tens of thousands of millions of resources that are indexed and accessible by research engines, and even this is only the visible part of the iceberg [5]. A short time ago, the total amount of information contained on the Web was estimated to be 17 times the size of the Congress Library [14]; but the figures so defy the imagination that even this analogy is difficult to grasp. Moreover, these figures are based on a conception of the Web as mainly occupied by static documents, which are relatively simple to capture and to analyse. However, what was still the case five years ago is no longer valid. Thus, we have passed from a documentary model to a model of dynamic resources with a wide variety of formats (social networks, sites which regroup contents, micro-publications, etc.) to which one has access according to diverse modalities , and which render the task of systematic exploration even more arduous.

Faced with this mountain of information, which by its size, by its form, by the means which are necessary to measure up to it has the stature of an Everest, any attempt to take the Web as an object of study leaves one with the choice of two attitudes. The first attitude consists of considering the Web system in its totality. In doing this, one equips oneself with a panel of general indicators which require an enormous computational power, and which even so will have difficulty being locally precise [16]. This is the case in particular for those research engines which have to evaluate and to classify the relevance of any resource whatsoever for a given user request. All the potential resources have to be considered as comparable, and the calculations thus use an identical algorithm of classification for each of them; this is the case for example with the Pagerank of Google [9] . By doing this, one transcends local organisations for the benefit of a destructuring universality, somewhat like a satellite view of the Web which reveals the general large-scale principles.

Another vision of the Web consists of viewing it as a very large structure composed of smaller localities, noteworthy organisations of resources belonging to different registers. Each locality is organized in a specific way and emergent properties varies from one locality to another. In adopting this vision, one seeks not so much to elucidate general indicators, that can be true for the entire system but too vague to be truly useful, but rather to formulate hypotheses and discover properties concerning each of the sub-spaces encountered. This model gains in complexity (more properties and more interactions to explain that are heterogenous between subspaces) what it loses in universality (as the general model is abandoned); it has the advantage of not distorting its components, making it possible to provide a fine analysis of a locality or a group of localities [11]. We will refer to this below as the "second approach", and it is the one which we will develop in this article.

2.1 Structures with multiple dimensions

Whatever posture be adopted, the reality of the Web translates into a multitude of resources on several intertwined levels. What at the start is only a Web resource, i.e. an object accessible via the HTTP protocol with the help of a URL address, can thus be described on five different levels which, once combined, form the localities that will be mentioned below. These levels are not complex properties in themselves, they are primary categories in which complex properties can be classified.

We first distinguish the level of the resource. Here one finds what the server sends back when one calls a particular URL. The content can take various forms, the most common being text, images or audiovisual documents. An HTML resource (a hypertext page) commonly possesses a structure composed of several linked resources (scripts, style pages, media, etc.) each of which, taken one by one, constitutes a full resource in its own right. The second level is that of the link. This level is naturally very present on Internet, but it is important to distinguish intra-level and inter-level links. The former consists of links between elements of the same level, such as hypertext links between Web resources, links of relationship or friendship on the social level, or yet again temporal links of succession or simultaneity on the temporal plane. By contrast, inter-level links connect together two different levels: for example a link can connect the author of a Web-page to the page in question, just as a link can connect a resource to a geographical site in a number of different ways that we will examine below. These links are very often the source of inaccuracies, since they associate two elements which are heterogeneous by nature. We also will return to this aspect in the case of geo-localisation. The social level is the third level of description that we can invoke. This level describes the human individuals who gain access or who publish the content and the relations they hold between them. This level is very much in fashion nowadays when social networks flourish on the Web; not content with putting actors in touch, this also makes it possible to describe their relations with an ever-increasing richness of detail. Next, the geographical level integrates a territorial component. If the graph of links constructs a multi-dimensional space without any physical referent, the geographical reference is more and more used on the Web. This makes it possible to answer questions such as the physical location of resources, the routes by which information passes, the positions of authors and readers, or yet again the places mentioned in the content itself. Finally, the temporal level describes the evolution of resources, of links, of geography or yet again of actors over time. In a system where the dynamics of resources and links is so strong, there is a lot at stake in integrating this temporal dimension in order to study the system. However, here we come up against technical limits which are still very strong, in particular because of the great mass of data which prohibits taking regular snapshots.

In order to study and explore the Web, we must be able to combine all of these levels, in order to make sense of what is going on. The resources are inscribed in these different levels which all together compose a system which is complex both by its size and by its structure. And even if the number of combinations is so great that it is not conceivable to completely integrate all these levels, we can nevertheless seek to grasp some organisational features in order to highlight the localities we mentioned above as being at the heart of the second approach we adopt here. Of course the crucial question is the method to be employed to reach this objective; but before revealing this method and putting it into practice, we must first specify more completely on one hand the importance of the geographical aspect which will be at the centre of our preoccupations, and on the other hand the choice of the supposed locality that we will analyse on an experimental basis.

2.2 Spatiality in full development

Among the different levels evoked in the vast variety of possibilities available to describe a Web resource, it is difficult to ignore the rise in strength of the level of geography: from participatory cartographies to contents that are geo-tagged , by way of open data that refer to towns and the localisations of resources or actors [19], the geographical reference has become unavoidable. This phenomenon is due both to technical advances which have made geolocalisation easier, and by the consequent fleshing out of the services provided . But it is important to keep in mind that this progress answers a demand which was not hitherto satisfied for a strong spatial anchoring which would bring the virtual reality of the Web back to a more concrete reality. On the Web, spatial information benefits from the natural cultural familiarity that users have with it, and which renders it special. In particular this helps to fight against the disagreeable or painful dis-

orientation that threatens when surfing on the Web [7]; it helps to recover an orientation, to get ones bearings. It is true that the culture of the Web bathes in the idea that the hypertext link always prevails, because it was present right from the conception of the network and largely explains its success. There is thus a tendency to consider the hypertext link as the unique structuring principle of the network; but this is not necessarily the case. Indeed, even if the geographical structure starts out as a mere addition, a supplementary indicator that is not essential, it has the potential to transcend the hypertext link. Geographical information thus becomes a way of structuring this space, and of explaining it [17]; the ideal being to couple together as many dimensions as possible, and in particular hypertext and geography.

Moreover, integrating the geographical dimension is particularly interesting for testing methods of analysing the Web, because it brings into play a great many different aspects. Thus, one can investigate the localisation of the material which makes up the contents, or of the persons in charge of these contents, as well as elucidating the geographical references in the contents themselves. The richness and diversity of the means of analysis renders the integration of this dimension perfect for a heterogeneous methodological study where one wishes to reveal the structure of several different levels present in the same set of data.

2.3 Choice of a set of data

In order to test our second posture and the methodology which will allow us to reveal a certain sort of organisation, we have chosen a particular set of data in order to limit the problems of capture and to focus on the exploration. The question of capture on the Web is a whole problem in itself, once again because of its size and its heterogeneous and dynamic nature. While one may postulate the existence of distinct aggregates of data, it remains problematical to capture them while mastering all the different dimensions. In order to eliminate a part of the problem, the system we have chosen is the on-line shop for Apples iOS software programmes `http://itunes.apple.com/us/genre/ios/id36?mt=8`. One can find there software for iPhone, iPad and iPod touch. This is a system which represents a bounded portion of the Web with known limits, since it contains a list of software units which evolve with a dynamic that can be measured. Thus, for a very large number of pages (close to 500,000 unique addresses) one can extract a title, a description, user evaluations, a price, a classification by category; links to other similar software, links with Internet sites, etc. We thus have a relevant terrain, very broad, with information of all sorts, hypertext links towards the exterior of the Apple site, and internal links towards other software which make it possible to efficiently test an exploratory chain.

3 A hybrid methodology

The exploration of complex systems and the emergence of knowledge concerning them is a fundamental problem which is born from the confluence of two factors: on one hand, the abundance of data in a digital form which means that they can be computed and exploited using algorithms; on the other hand the wish to consider ever more factors in the study of phenomena, in a quest for exhaustive measures taken as a synonym for truth. It is a question of understanding systems as wholes, facing up to their complexity. Two major routes have been considered for studying complex systems in this way, and they echo two conceptions of weak emergence[4]: low-level first or micro-level first emergence and high-level or macro-level first emergence.

The first route is thus to consider a low-level first emergence, where the property of emergence is carried by the elements which make up the system and deducible from them. It is a question of putting the autonomous elements that one masters into interaction with each other, so as to observe their behaviour compared to that of the natural system that one wishes to analyse. However, this approach does not make it possible to discover new hypotheses on the basis of the real system. It is rather a question of simulating the real situation by trial and error, making hypotheses and testing them, which produces more hypotheses.

The second route, by contrast, considers a high-level first emergence where the emergence is hardly reducible to the components of the system. The major drawback is that in this case, it is necessary to discover and to reveal the structure by considering the system as a whole (or what one imagines may be its whole), in favour of a model which is certainly approximate, but functional. This allow to discover low level properties than can be simulate in a second time. The majority of studies which concern networks (be they the Web, co-citations in scientific articles, or social networks) [2] take this route, using visualisation by means of graphs as revealing notable patterns and as generating indicators which are specific to the network under observation. For this route there are very few tried, systematic methods. The EDA (Exploratory Data Analysis) is one of them, whereas a method with similar objectives appeared at the same time as the visualisation of information [20] under the name of Visual Analytics [21].

3.1 Augmentation of the EDA and combination with visual analyses

One of the rare existing methodologies, the EDA [24], thus allows for the instrumentally equipped discovery of new structures in a complex system, by means of a systematic recourse to the visualisation of information in all the phases of the exploration of an unknown complex system; this has the effect of maximizing intuitions during

the formulation of hypotheses [1]. In other words, it is a question of providing the user with the means of grasping a mass of data and becoming familiar with it, in order to discover noteworthy patterns which can be the source of questioning and the elaboration of hypotheses concerning the system. The exploration is divided into precise subtasks in order to end up with a relevant, coherent model of the original system. We can summarize the succession of tasks as follows: visualize the whole; zoom and concentrate on a part; pay attention to particular details. This makes it possible to overcome the classical dichotomy between the analysis of macro-structures and the analysis of micro-interactions. Thus, the juggling between the aggregation and the disaggregation of the data, and the back-and-forth between the global and the local, makes it possible to detect and to follow the emergent patterns.

This process of generating hypotheses concerning the system by abduction finds a similar echo in Visual Analytics where the summary of tasks becomes: first view the general level; then zoom and filter; finally enter into detail as needed. This method, scientifically less rigorous because it contents itself with an overview of the data before filtering it in an analytical process, is nevertheless very useful to equip one-self with innovative visualisations which are lacking with EDA. To the extent that visualisations are at the heart of the process, it is important to have a maximum of choice in the conceptions or in the possibilities of interactions, that one will then adapt as needed. Moreover, visualisations of information are particularly well adapted to the heterogeneous, dynamic data that they make it possible to present together, whereas the EDA is difficult to use on non-homogeneous digital data. On the other hand the visualisations of information are generally ad hoc constructions, conceived with a unique aim which makes it possible to explore a particular aspect in a single interface; whereas ever since its origin the EDA pleads in favour of a pluralistic use of statistical visualisations [15] in order to cover a maximum of cases.

The EDA thus brings a degree of rigour to the process and the faculty of linking simple statistical visualisation; whereas the visualisation of information and its analytical branch contributes complex visualisations more adapted to exploration. There is thus an interest in combining the advantages of both these two domains without losing either rigour, or creativity.

3.2 A semiological component as an integral part of the analysis

Since the visualisations are at the basis of the approach, in order to be effective they must be framed by a theoretical scheme of graphical semiology. The choice of the visualisations, which are the principal link between the human and his object, is crucial. The EDA is based on existing semiologies to help it make this choice [22, 6]; but the latter are limited to informational graphics and are only really effective for systems that are already constituted. Thus they are of little help when faced with novel systems of which one knows little or nothing.

In order to solve these problems and to dispose of a basis as solid as possible in order to construct a process for the exploratory analysis of data, we have developed and used our own semiology based on experimental studies [18]. We have thus succeeded in showing that according to the structures used, the visualisation in question does not lead to the same reasoning and knowledge. According to their degrees of constraint, they can either offer a wide range of possibilities in perceptual terms and thus in terms of information and knowledge, or on the contrary restrict the range of possibilities in order to focus on a precise element of information. It is thus possible to combine different types of visualisation, as a function of their characteristics, according to the needs of different stages in the elaboration of indicators. We can distinguish two extreme cases: on one hand one wishes to have an open visualisation in order to formulate hypotheses concerning the global organisation and to maximize the insights; on the other hand, one may wish to perceive and to use the result concerning the organisation, of a hypothetical particular descriptor while minimizing fresh insights. The

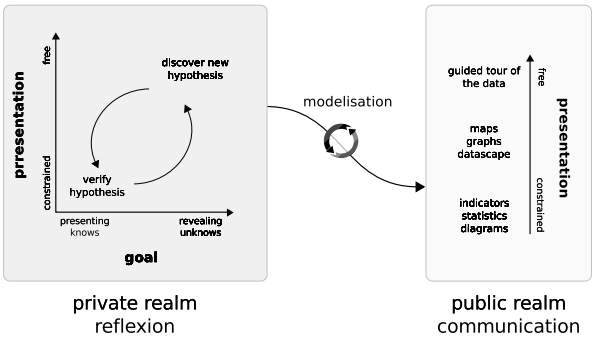


Figure 1: The EDA process adapted to very large sets of heterogeneous data. The exploration is presented on the left and is conducted by a single explorer. The presentations, free and constrained, result from a taxonomy of visualisation as a function of the degrees of freedom that they allow in their interpretation

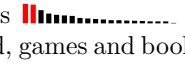
open visualisations which maximize the insights will be those endowed with structures that are as free as possible. Structures such as graphs or maps, for example, fill this role very well. A balance between global reading and local reading will provide insights of different sorts. If regularities are perceived in these visualisation and one finds points of perceptual anchoring whatever their nature, be they the orientation of links, groupings, spread-out displays, routes, incongruous patterns, etc. they become potential indicators that will have to be analysed in order to formulate a clear hypothesis. The choice of the visualisation and its mode of presentation correspond already to a generic and very imprecise way of formulating a hypothesis which amounts to a vague intuition. The

formulation of a clear, precise hypothesis should have the aim of verifying this vague hypothesis by refining it and presenting it by means of the second sort of visualisation, the constrained visualisations. The latter are constructed from the basic data to which one applies filters or data processing algorithms, corresponding to a specific hypothesis formulated in advance which is spatialized in order gain knowledge of it. It is a question of transmitting a specific element of knowledge to the system, and the presentation must be perceptually constrained so that it is not possible to draw erroneous conclusions, but only an indication concerning the specific hypothesis that one wishes to test. The structure of a diagram in two dimensions is an example of a visualisation of this type, on condition that it respects the semiological canons of the genre.

We have been able to test these phases, and the proposed methodology as a whole, on the set of data Apple iOS Application Store US. This provided the opportunity to confront widely diverse tools and visualisations with a real system.

4 Heterogeneous tools for exploration

Now that we have laid down the theoretical basis, it is possible to begin the exploration we have chosen to carry out, by testing the numerous tools which are available free; these run from visualisations to data bases, including in between data-processing algorithms and other programs which are necessary for carrying out this study. There exist today a large number of means, which are made available to research scientists, journalists and others interested in the exploration of data [12, 25]; the drawback is that each uses its own format of exchange or posting. It is therefore necessary to jungle permanently with heterogeneous tools; the toolkit becomes a heterogeneous assembly of diverse scripts and it is necessary to continually pass from one to another. In this context, it is important to keep a dynamic connection between the intuitions one may have and their application, in order to conserve efficiency. Thus, latency resulting from concentrating on the technique leads us away from understanding the system and from an intimacy with the data, which is a key to success in exploration.

Before plunging into the exploration, it is necessary to carry out a capture of the set of data which is as correct as possible. This means, concretely, launching an organizing robot which will recuperate all the links to pages, and which will then take care of launching a robot to capture the content of each page. The problem here is to recuperate all the pages while adapting to their structures which often vary, and to do this in a minimum time in order to have a snapshot which is not too extended in time, and without being banished from the site under visit (if one generates too much traffic on a site, the latter can choose to no longer serve the pages). The clich of iOS which stores US Apple, slowing down the capture robots as little as possible, nevertheless takes seven complete days for near to 500 000 referenced pages and 277 997 unique applications. Between the beginning and the end of the operations, 15 000 applications have disappeared. Each page follows an identical pattern, with minimal variations which contain structural elements that are stored as one goes along in a free relational data-base (PostgreSQL) which will subsequently allow for easier extractions. Each capture and each insertion is visualized in the form of a very long list, where in the case of an error the cause is immediately apparent (change in colour and length of line) in order to permit visual surveillance. The crawler robots (robots that download a web resource to analyze its content and extract relevant information) are programmes developed in Java especially for the occasion in order to render them more precise, because specifically adapted to a particular site, but which are not re-usable on other portions of the Web. Once this work is completed (provisionally at least), it is time to pass to the exploration itself. To do this, we have recourse to the methodology which consists of getting to grips with the system with the help of free visualisations of the graph type, and to validate preliminary intuitions by constrained visualisations. In order to see the large picture, we have tried out two approaches. The first was to observe the distribution of applications in the various categories, to see which categories were the best represented and if there were substantial disparities. A histogram, resumed here as a sparkline[23], makes it possible to isolate two categories  which are particularly well represented, games and books, with close to 40 000 applications each. We then find categories that are less well furnished, such as news, finance or weather, with an average of only 5 000 applications each. Concerned to see how the applications within each of these categories were organized, we opted for a second approach in parallel which consisted of viewing the whole set of applications in the form of a graph, considering the suggestions for purchase situated at the bottom of each application as a link towards another application. The nature of the link is a matter of discussion and cannot be completely trusted as this is a black box process provided by Apple. Our goal here is as much to reveal or at least have some hints about the suggestion algorithm as to describe the system structure in itself. Further work should include more links like similar authorship or hypertextual link between applications but outside of the Apple IOS store. An interactive visualisation of data of such size in the form of a graph was only possible using the free software Gephi [3], and only after waiting 40 minutes for displaying the latter.

4.1 Division in levels

This graph revealed two notable patterns. The first was a heart which was very dense, but relatively disconnected

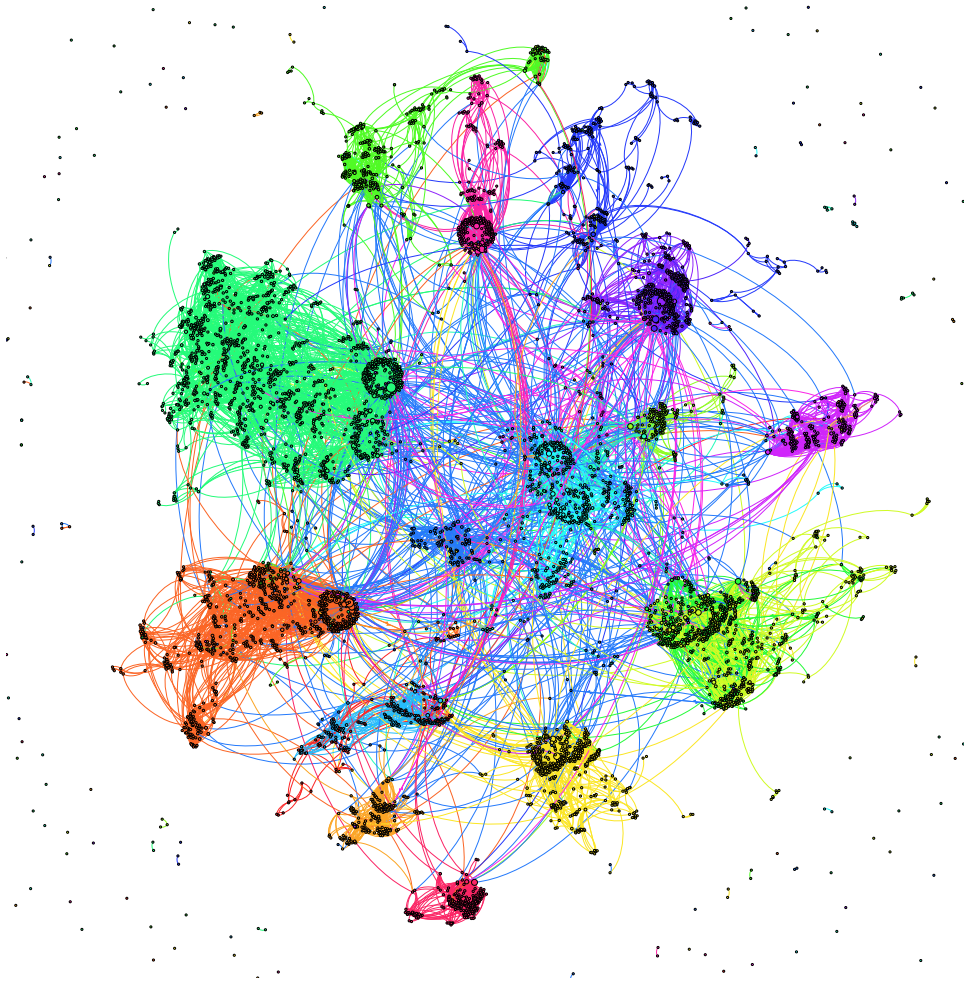


Figure 2: Graph of explorations of the heart. The size of the nodes depends on their degree (number of entering and leaving links), whereas the colour is associated with a category. It will be seen that the category indicators are relevant, since one finds localities on the graph that correspond to them.

from the rest of the graph. In fact this corresponded to applications destined for iPad, an Apple tablet for which applications can be dedicated. Their small number (10 000) leads to a large number of redundant links, which thus creates a sub-graph within the complete graph. In order to try and see more clearly what is happening, it was thus necessary to remove from the graph these specifically dedicated applications which obscure the overall exploration. This made it possible to discover an impressive quantity of applications which were not linked (too few purchases or not well displayed), others which were never visited but which had exit links, and finally a relatively small number which had both entry and exit links. This made it possible to visually exhibit three levels, which are found pretty systematically, in terms of thematic localities [13]; these are an exterior, a periphery, and a heart. We then confirmed this visual hypothesis by another visualisation, constrained this time in order to analyse the composition of the system with respect to these three levels. This figure was constructed with the visualisation library Web Protoviz [8], which allows a rapid and easy display of dynamically generated diagrams. This is useful for rapidly conceiving visualisations which respect elementary semiological principles and which concentrate on seeing rather than doing.

It turns out that the heart here is constituted of applications which are promoted by Apple, since we find many links towards them. This heart represents only 2.5% of the whole. A back-and-forth at the micro level for the on-line Apple shop reveals that a good number of these applications are promoted (or had been promoted at some previous time) on the home-page of the dedicated software iTunes. The visualisation revealed that the heart was composed of thematic localities corresponding to the categories, the most central one being the category utilities . Next, only about a quarter of the applications are on the exterior. It is interesting to note the criteria which led to their exclusion. The Figure 2 reveals an interest-

ing pattern according to which one particular category, books, has a very large number of these excluded applications (more than 25,000). At issue here are books edited in the form of applications, which compete with general applications which are free and very successful. In a general way, we find on the exterior applications which are too specialised, which have therefore not attracted the crowds and which are thus lost in the mass. By the way, we may remark that actually they are not particularly less well noted than the others.

This characterisation of the system in three levels of different natures makes it possible to concentrate the analysis on a particular zone, as recommended by EDA when it comes to passing from the micro-level to the meso-level. Among the multitude of questions which arise at this stage of the analysis, we have chosen to concentrate on a connection between the social plane and the geographical plane.

4.2 Fragmented geographical information

Indeed, one of the questions at the origin of this study is to understand how the developers of applications are distributed geographically, in order to couple this information with a study on the social plane. The system is closed and the applications are not free, so that we cannot content ourselves only with the pages that have been captured. Matters would be easy if the system already included geographical information or information concerning the authors of the applications, but this is unfortunately not the case. One finds neither geo-localisations, nor the names of authors or societies, only a pseudonym which renders a research of the Yellow Pages type inoperative. We thus come up against the problem of geography on the Web, and methodologies making it possible to obtain information concerning localisation. In the absence of any data of this sort in our original system, we are forced to have recourse to alternative indicators to obtain information concerning geographical distribution. The first of these indicators is to systematically localise the Website of the application. To do this there are mainly two techniques which function differently: the Geoip and the Whois. The Geoip consists of finding the place which could correspond to the physical location of the server which sends back the Webpage of the application. For a Web address of the type www.utc.fr, we create a correspondence with an IP address (role of the Domain Name Server) which follows a route made of various network equipments around the world and from which one can induce a localisation. For this we used the free correspondence base of Geoip which is reliable at the level of the country (99.5%), less so at the level of the town (79% in the USA) (`www.maxmind.com/app/geolitecity`). Thanks to the latter we were able to obtain in 24 hours (the time the program takes to run) a localisation for each of the applications, on condition that the site was correctly informed on the Apple page of the application.

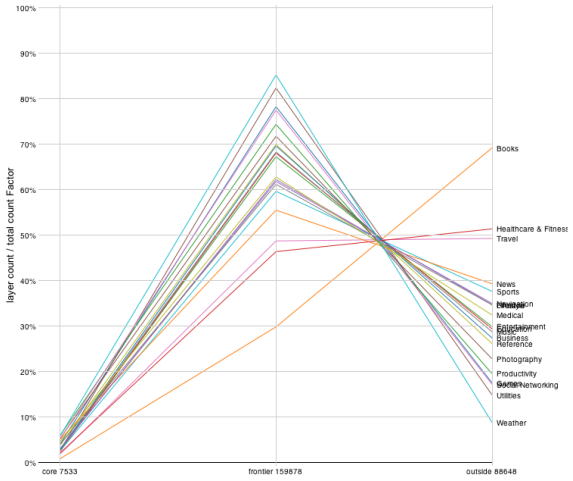


Figure 3: Distribution of applications in the heart, the border and the exterior, according to their category visualized in parallel co-ordinates.

We find in Figure 4 that the countries of the G8 and China are the most represented, and that the United States is far in front of the others. (They represent as much as the rest of the world combined although we should recall that we are on a US shop). The distribution also shows the domination of the East coast. There is very little material to compare with other web geoip datasets distribution. Nevertheless preliminary studies indicate that applications web sites are spread all over the world with a rather strong concentration in western Europe wereas applications are proposed in a US only store. It suggests that developers are concentrated in highly developed countries with a good engineering level and that we have to face an internationalization of developing applications in the analysis.

However, this visualisation must be taken with caution and must be confirmed, because there are many sources of inaccuracy: the first is that anyone at all can buy a name of a domain anywhere. We can count on the wish of the authors to choose a provider near to them or who speak the same language, but this is not at all certain. Next, some hosts of content optimize the requests by orienting towards a server close to the applicant with replicas all over the world. The simple fact of requesting the IP address of a site from a given site, as is the case here where one computer requested 277,000 IP addresses from the same place, is already enough to tamper with the results. Finally, the components of the sites (style page, pictures, etc.) can be hosted on several servers, and the Geoip only localises the master-page, and thus reveals only a part of the truth.

In order to face up to these limitations, it is possible to obtain the address of the person responsible for the Web

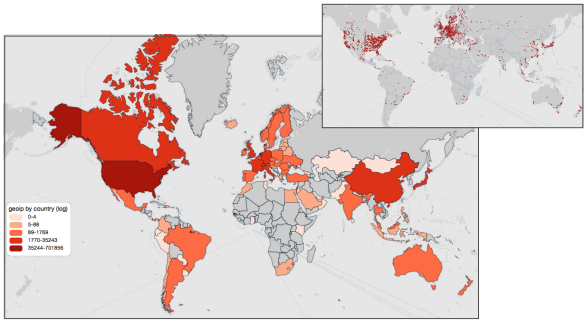


Figure 4: Chloropleth map of the geographical distribution of servers which host the Web sites of the applications (logarithmic shading scale). These visualisations are realised in SVG on an underlying map openstreetmap (www.openstreetmap.org) and generated dynamically by the javascript library polymaps (polymaps.org).

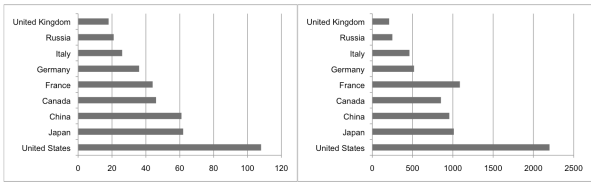


Figure 5: Number of citations of the names of G8 countries plus China in titles (on the left) and in the descriptions (on the right).

site by means of a service called Whois. This service provides information of a better quality, since it links a Web site to a real entity (a society, a person). For example, the applications of the game editor Gameloft reply via Geoip with a localisation in Canada, whereas Whois indicates that the seat of the enterprise is in France. On the other hand, nothing makes it possible to know directly whether the IP corresponds to a Canadian subsidiary of the enterprise. There remains the case where the provider of access puts up a screen with his own name in order to protect the information concerning his clients, which can happen. Moreover, whereas the Geoip base is available for free, a large number of Whois requests cannot be obtained for free. Thus for example it would cost about 800 euros to localize the 277,000 applications of our system. All the more so since after that it is necessary to convert the addresses into geographical locations with other services such as Google Maps, which also have limitations on the use that can be made of them. All this does not totally forbid the undertaking, but reduces the scope of our actions. A compromise thus consists of using the division in layers that we mentioned previously to focus on the 7000 largest applications, but we lose a lot of information. This study is still being carried out. Another approach to studying a geographical distribution was to note the various languages in which the applications are available, since this information can be related to the country in question. Of course since we are dealing here with the iOS Store US, English is used in 95% of cases. But among the other main languages used we find again the countries of the G8 together with China (in 4-6% of applications). This information was confirmed by analysing the text available in the applications themselves. Since the G8 countries turn up often, we looked to see how these countries are quoted in the applications, in the titles and descriptions. This makes it possible to measure the interest of the country in the system in general. Figure 5 shows the results of this distribution, which is very close to the distribution calculated in a sample of 5 million books (see `http://bit.ly/eulTDK`).

5 Conclusion

It is clear that we have not yet exploited all the information that could be extracted from this data-set, particularly concerning geographical information. The method has nevertheless allowed us to test the exploratory chain, and to explore a complex system quite effectively. Some emergent properties have been hypothetized and validated with visualizations like the heart, periphery and exterior. However, the use of the various softwares for visualisation and data-processing remains delicate when spatiality comes into play, because of the very great heterogeneity in format or in precision of the data that are available. Geographical properties seems to exists and the analysis will be pushed forward to let them emerge. This highlights how useful it would be to have a software capable of aggregating the various tools, in order to allow an exploratory analysis of data that would be less costly and less dependent on computer skills. As a result of the present study, our team has recently undertaken the development of software of this sort.

Acknowledgements

The research for this work was supported by the Scientific Council of the Compigne University of Technology.

References

[1] Andrienko, N., Andrienko, G., *Exploratory analysis of spatial and temporal data: a systematic approach*, Springer, 2006

[2] Barabasi, A. L., *Linked : The New Science of Networks*, Perseus Publishing, 2002

[3] Bastian M., Heymann S., Jacomy M., *Gephi: an open source software for exploring and manipulating networks*, International AAAI Conference on Weblogs and Social Media, 2009

[4] Bedau, M. A., *Weak Emergence*. In James Tomberlin, ed., Philosophical Perspectives: Mind, Causation, and World, vol. 11 (Blackwell Publishers), 1997, pp. 375-399.

[5] Bergman, M. K., *The Deep Web: Surfacing Hidden Value*, The Journal of Electronic Publishing, volume 7, issue 1, 2001

[6] Bertin, J.,*La smiologie graphique*, Mouton-Gauthier Villard, 1963, Paris

[7] Boullier, D., Ghitalla, F., Gkouskou-Giannakou, P., Le Douarin, L., Neau, A., *L'Outre-lecture,Manipuler, s'approprier, interprter le Web*, Centre Pompidou Library Editions, 2003, Paris

[8] Bostock, M., Heer, J., *Protovis: A Graphical Toolkit for Visualization*, IEEE Transactions on Visualization and Computer Graphics, 2009, p. 1121-1128

[9] Brin, S. and Page, L., *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Seventh International World-Wide Web Conference (WWW 1998), 1998, Brisbane, Australia.

[10] Cilliers, P., *Complexity and Postmodernism: Understanding complex systems*. Taylor & Francis, first edition, 1998

[11] Ghitalla, F. *La gographie des agrgats de document sur le Web*, Research White Paper, Published online, 2004, `http://www.scribd.com/doc/44666231/Geographie-des-agregats-Web`

[12] Janert, P. K., *Data Analysis with Open Source Tools*, O'Reilly Media, Inc, 2010

[13] Kleinberg, J., M., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A. S., *The Web as a Graph: Measurements, Models and Methods*, Lecture Notes in Computer Science, Vol. 1627, 1999

[14] Lyman, P. and Varian, H. R., *How Much Information*, 2003. Published online `http://www.sims.berkeley.edu/how-much-info-2003`

[15] NIST/SEMATECH *e-Handbook of Statistical Methods*, `http://www.itl.nist.gov/div898/handbook/`, 2010

[16] Pfaender, F., Jacomy, M., Fouetillou, G., *Two Visions of the Web: from Globality to Localities*, Proceedings of Information and Communication Technologies, 2nd edition, 2006, Damas, p. 566-571.

[17] Pfaender, F., Jacomy, M., Ramm, M., *Percevoir le territoire Web Picard*, in Proceedings of SAGEO06 , 2006, Strasbourg, France

[18] Pfaender, F., *Spatialisation de l'information : Lire, inscrire et explorer les systmes informationnels*, PHD thesis, Universit de Technologie de Compigne, 2009

[19] Scharl, A., Tochtermann, K. (diteurs), *The Geospatial Web, How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*, Advanced Information and Knowledge Processing Series, 2007, London: Springer

[20] Shneiderman, B., Card, S.-K., MacKinlay, J.-D., *Readings in Information Visualization, Using Vision to Think*, Morgan-Kaufmann Publishers, 1999, New York

[21] Thomas, J. J., Cook, K. A., *Illuminating the Path: The R&D Agenda for Visual Analytics*, National Visualization and Analytics Center, Published online, 2005, `http://nvac.pnl.gov/docs/RD_Agenda_VisualAnalytics.pdf`

[22] Tufte, E., *Visual Display of Quantitive Information*, second edition, Graphic press, cheschire, Connecticut, 1993

[23] Tufte, E., *Beautiful Evidence*, Graphic press, cheschire, Connecticut, 2006

[24] Tukey, J. W., *Exploratory data analysis*, Addison-Wesley, 1977

[25] Warden, P., *Data Source Handbook*, O'Reilly Media, Inc, 2011

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

EFFECTS OF PARTNER SELECTION ON THE EMERGENCE OF CULTURAL DIVIDES IN MIXED POPULATIONS

Elpida Tzafestas *

Abstract. In this work, we are exploring spatial dynamics and clashes in cultural simulations involving multicultural populations with partner selection. We are using as basis an Axelrod model [1][2][3][4] extended with a Moore neighborhood, heterogeneous sets of cultural features per agent and a number of psychologically realistic, basic and more advanced, conceptual models of cultural affinity perception and imitation [5]. Elsewhere we have shown that in many cases the population stabilizes to multi-cultural configurations and that in cases of population clashes where two or more culturally contiguous populations meet the cultural divide may persist, albeit in a relatively weaker form [5] [6]. In this paper we repeat our previous experiments to investigate the effects of partner selection on these configurations.

Keywords. Cultural simulation; Axelrod model; Selfishness; Indifference; Cultural clash; Social network; Partner selection.

1 Models

First, we are briefly summarizing the extended models that we are using (for an in-depth account, see [5]). Our models have been inspired by the observation (also put forward by other authors, for instance [4]) that the original Axelrod result, where cultural grouping and polarization emerges in an initially diverse society, is a combined side-effect of the model assumptions of 4-connectivity and fewer features than traits per feature. If instead, more features are used with fewer traits each, all systems eventually lead to monoculture. On top of this, if 8-connectivity (Moore neighborhood) is assumed, systems converge to full affinity substantially faster. It is this modeling intricacy that led us to reflect initially on the factors that may be responsible for the emergence of diverse cultural groups.

Heterogeneous Axelrod model: It is an Axelrod model, where each agent may have a variable number of cultural features, that are initially ordered. For example, agent-i may have (the first) five features whil agent-j may have (the first) nine features etc. This way, some features are widespread within the population and others are not.

Selfish model: Each agent has a feature vector, defining a trait value for a feature or none (absence of the corresponding feature). The obvious option is to have agents compute affinity on common features and define 0 affinity for uncommon features. Thus for two agents with features $[T_1, T_2, --, --, T_5, --, T_7, --, T_9, T_{10}]$ and $[--, T_2, T_3, --, T_5, --, T_7, T_8, --, T_{10}]$, both agents will perceive affinity as [–,aff,–,–,aff,–,aff,–,–,aff], where aff denotes the regular Axelrod affinity for the corresponding trait and is computed only on common features and any feature perceived as not shared by an agent does not count. A **second selfish model** more psychologically realistic is to allow an agent A to compute affinity with agent B only on As features (and affinity will be 0 if B does not possess the corresponding feature), while B will compute affinity with agent A only on Bs features. In the previous example, the perceived affinity vector for the two agents will become respectively [0,aff,–,–,aff,–,aff,–,0,aff] and [–,aff,0,–,aff,–, aff,0,–,aff]. Both models have been found to lead to a single value for shared features within the population (affinity for shared traits is called actual affinity), while the latter model results in lower perceived affinities by agents.

Indifferent model: It is implemented via the use of an additional dont-care vector storing a boolean value foreach feature (true for indifference). For instance, for two agents with 10 features and dontcare vectors respectively [t,f,f,t,t,f,f,t,t,t] and [f,t,f,t,t,t,f,f,f,t], the corresponding partial affinity vector will be computed as [aff,aff,aff,–,–, aff,aff,aff,aff,–]. Again, this supposes an homogeneous feature handling method, whereas features for which both agents are indifferent do not count. As before, we define a second indifferent model, where the dont-care values are treated in an individualistic manner; in the previous example, the partial affinity vectors of the two agents become [–,aff,aff,–,–, aff,aff,–,–,–] and [aff,–,aff,–,–,–,aff,aff,aff,–], respectively.

Complex model: A model that uses degrees of indif-

*Elpida Tzafestas is with Cognitive Science Laboratory, Departement of Philisophy and History of Science, University of Athens, University Campus, Ano Ilisia 15771, Athens, Greece. E-mail: etzafestas@phs.uoa.gr

ference toward cultural features. Each cultural feature is assigned a real-valued weight between 0 and 1: the lower its value, the more indifference the agent will show toward the feature. The perceived affinity of an agent with another one is defined accordingly as $\sum w_i \text{aff}_i / \sum w_i$, while the actual affinity is defined as usually. Apparently, with this model it is very common to have agents that perceive each other very differently due to their different weight vectors.

2 Experiments

We have performed a series of experiments with and without partner selection to understand the effect, if any, of partner selection. Partner selection uses a simple probabilistic preference scheme, previously introduced in [7], where each agent maintains a set of probabilities of interaction with each one of its eight neighbours. All partners are equiprobable in the beginning and the corresponding preferences develop after each interaction according to the following reinforcement algorithm:

```
Let aff be the affinity against partner i
   (with whom interaction just occurred)
   and avg be the average affinity
   with all neighbours.
If (s > avg) then increase preference
   for partner i,
Else decrease preference for partner i.
Re-normalize all preferences so that
   they express probabilities
```

Experiment 1. Modified Axelrod model (homogeneous) results with an initially diverse population (fig. 1). Partner selection does not change the final outcome which is full monoculture in the population, but it substantially retards the cultural contagion process. On the other hand, partner selection induces extremely fast local polarization as is depicted in fig. 1(left) where an intermediate configuration is shown, that is not possible without partner selection in this model. Fig. 1(right) shows the large speed of local convergence in the beginning of the experiment that slows down in what follows compared to the convergence speed without partner selection.

Experiment 2. Modified Axelrod model (homogeneous) with two initial populations where a cultural clash is expected. Again, partner selection does not change the final outcome which is full monoculture in the population, but it substantially retards the cultural contagion process. This is due to the border agents between the two populations that create individualistic partnerships and thus hinder the fast spreading of cultural information.

Experiment 3. Modified Axelrod model (heterogeneous) results with an initially diverse population. Partner selection leads generally to fairly lower final affinities and is slower as well, as in previous cases. However, as in experiment 1, partner selection induces extremely fast initial development as shown in fig. 3(middle). Fig. 3(right) shows how the average preference bias (preference bias = maximum preference minimum preference for a neighbour within an agent) develops. As expected, the bias rises quickly to a high value, whereas it remains random in the case without partner selection. The unstable nature of the latter measure is due to the presence of intrinsic noise to the agent preferences.

Experiment 4. Heterogeneous selfish 2 vs. indifferent model results with an initially diverse population. As before, partner selection does not change substantially the final outcomes (it may lead to slightly lower affinities), but is faster in the beginning and overall slower (fig. 4).

Experiment 5. Heterogeneous selfish 2 vs. indifferent model results with two initial populations where a cultural clash is expected, but border agents may form individualistic partnerships, as in experiment 2. As in previous experimens, partner selection does not change the final outcomes. In the case of two selfish 2 populations that meet, affinity with partner selection develops higher values on its way to final convergence (fig. 5 left), while in the case of two indifferent populations that meet, affinity with partner selection develops much slower showing a final avalanche effect (fig. 5 right).

Experiment 6. Heterogeneous complex model results with an initially diverse population or with two initial populations where a cultural clash is expected. In the case of the mixed population, as before, partner selection does not change the final outcomes but cultural convergence slows down (fig. 6 left). In the case of two complex populations that meet, stabilisation is almost immediate to substantially lower affinity levels, because of the potential of agents to quickly build and maintain individualistic partnerships with some of their neighbours (fig. 6 right). The disturbances found in the graphs with partner selection are due to the intrinsic noise to the agent preferences.

3 Conclusion

Overall, partner selection has been found not to alter substantially final expected outcomes without partner selection, but to slow down cultural convergence because of the initial construction of partnerships that constrain further development. From an evolutionary point of view, a partner selection mechanism would be promoted by evolution, because real biological and social systems generally do not have all the time needed to develop final, stable configurations, but operate in response to newly encountered situations. In such cases, it makes sense to dispose of

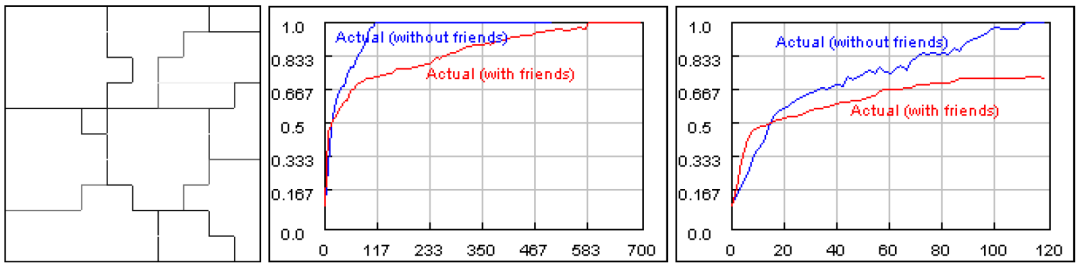


Figure 1: **Experiment 1** - (x: time in 1000s of cycles, y: actual affinities) Typical outcome for a modified homogeneous Axelrod model. (Left) An intermediate configuration in the case of partner selection. (Middle) Actual affinities with and without friends (partner selection). (Right) Zoom to the first part of the previous graph (first 120 cycles).

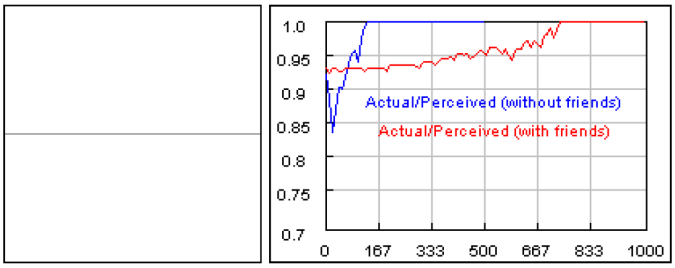


Figure 2: **Experiment 2** - (x: time in 500s of cycles, y: actual affinities) Typical outcome for a modified homogeneous Axelrod model. (Left) The initial configurations, with two homogenous populations occupying each half of the space. (Right) Actual affinities with and without friends (partner selection).

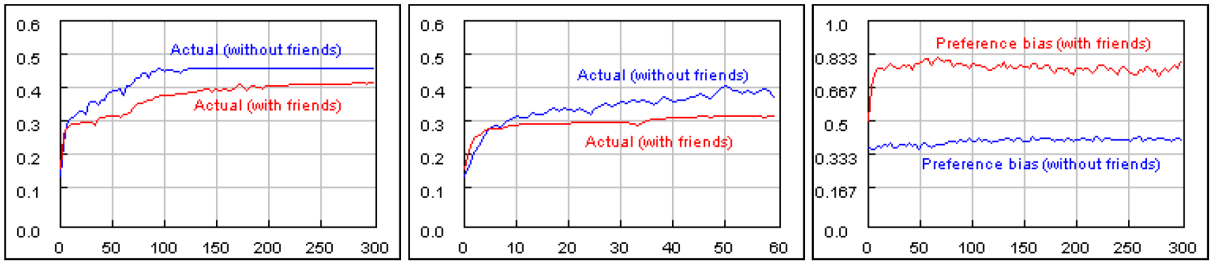


Figure 3: **Experiment 3** - (x: time in 1000s of cycles, y: actual affinities) Typical outcome for a modified heterogeneous Axelrod model. (Left) Actual affinities with and without friends (partner selection). (Middle) Zoom to the first part of the previous graph (first 60 cycles). (Right) Preference bias with and without friends (partner selection).

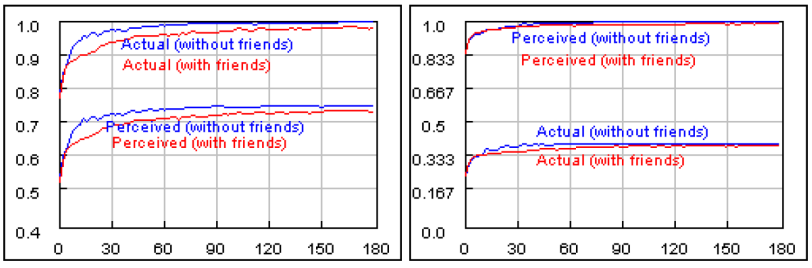


Figure 4: **Experiment 4** - (x: time in 1000s of cycles, y: actual affinities) Typical outcome for an heterogeneous selfish 2 or indifferent model. (Left) Actual and perceived affinities for the selfish 2 model with and without friends (partner selection). (Right) Actual and perceived affinities for the indifferent model with and without friends (partner selection).

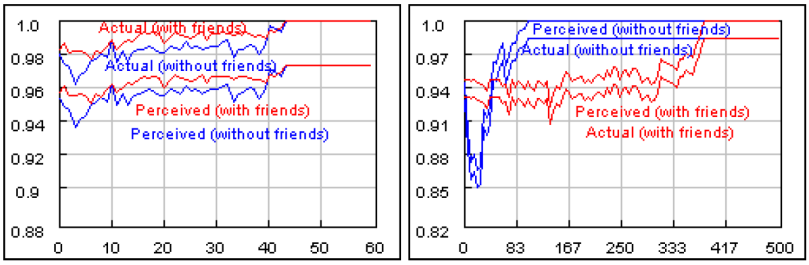


Figure 5: **Experiment 5** - (x: time in 500s of cycles, y: actual affinities) Typical outcome for a selfish 2 and a indifferent model starting from an initial configuration as depicted in fig. 2 left. (Left) Actual and perceived affinities for two selfish 2 populations with and without friends (partner selection). (Right) Actual and perceived affinities for two indifferent populations with and without friends (partner selection).

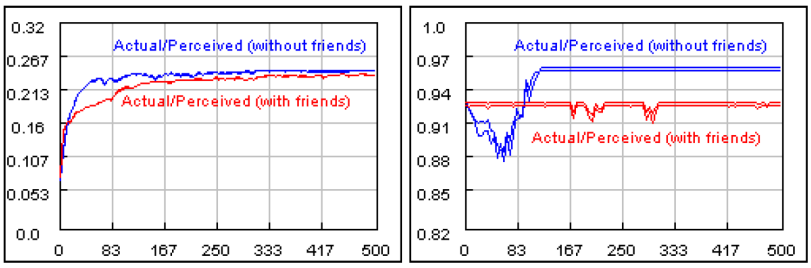


Figure 6: **Experiment 6** - Typical outcome for a complex model. (Left) (x: time in 1000s of cycles, y: actual affinities) Actual and perceived affinities in an initially mixed population with and without friends (partner selection). (Right) (x: time in 500s of cycles, y: actual affinities) Actual and perceived affinities starting from an initial configuration as depicted in fig. 2 left with and without friends (partner selection).

means to find quick, working solutions to problems, and partner selection appears to be exactly such a means to form quickly small groups of high affinity, although unstable in the longer term.

References

[1] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, John Wiley and Sons, 1974.

[2] Axelrod, A. (1997) *The dissemination of culture: A model with local convergence and global polarization.* The Journal of Conflict Resolution, 41(2), 203-226.

[3] Shibanai, Y., Yasuno, S., Ishiguro, I. (2001) *Effects of global information feedback on diversity: Extensions to Axelrods adaptive culture model.* The Journal of Conflict Resolution, 45(1), 80-96.

[4] Trigg, A.B., Bertie, A.J., Himmelweit (2008) S.F. *Modelling Bourdieu: An extension of the Axelrod cultural diffusion model.*CRESC, Open University, Working Paper No. 49.

[5] Lanchier, N. (2010) *The Axelrod model for the dissemination of culture revisited.* arXiv:1004.0365v1, 2 Apr.2010.

[6] Tzafestas, E.S.(2011) *Cultural diversity dynamics*, Proceedings IEEE Symposium on Artificial Life, Paris.

[7] Tzafestas, E.S.(2011) *Cultural dynamics, clashes and stability*, Proceedings Indian Conference on Artificial Intelligence, Tumkur, India.

[8] Tzafestas, E.S.(2007) *Emergence of social networks in systems with attraction.* Proceedings Satellite Workshop Interacting Agents, Complexity and Interdisciplinary Applications (IACIA), European Conference on Complex Systems, Dresden.

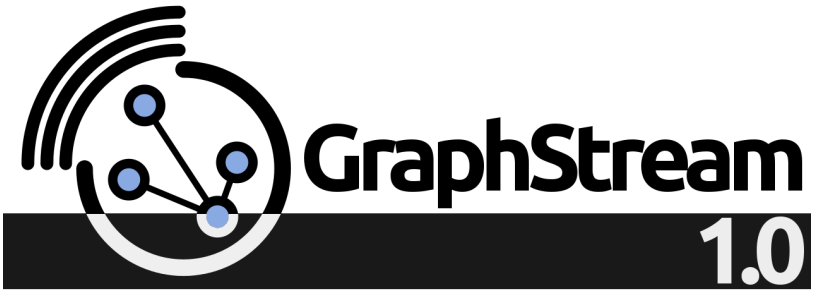
Part V

GraphStream

Proceedings of EPNACS 2011 within ECCS'11
Emergent Properties in Natural and Artificial Complex Systems
Vienna, Austria - September 15, 2011

GRAPHSTREAM
A JAVA LIBRARY FOR DYNAMICAL COMPLEX NETWORKS

Eric Deussé *



1 GraphStream Description

GraphStream is an open-source java library allowing to manage graphs describing complex networks and their evolutive dynamics.

GraphStream goals are:

- To give practical tools to dynamically manage graphs evolution and associated data for:
  - New measure indicators for complex networks including dynamics;
  - Robust algorithms suited to dynamics.
- To give basic development platform for various applications:
  - Urban mobility and transportation simulations;
  - Adhoc networks;
  - Social networks;
  - Adaptive distributions of dynamical simulations or dynamical systems;
  - Analysis of network morphodynamics.

*Eric Deussé is a collective name emerging from the swarming team RI2C from LITIS laboratory, University of Le Havre, 25 rue Philippe Lebon, BP 540, 76058 Le Havre Cedex, France

GraphStream is based on event model describing graph evolution as stream between a source that produces events and one or more sinks that receive and process these events. Events can be changes in the structure of the graph or in attributes of elements :

- addition or removal of nodes;
- addition or removal of edges;
- addition, update or removal of data attributes;
- time steps.

GraphStream provides a visualization tool to display a graph and its evolution. This tool is able to use existing coordinates of nodes (looking in node attributes), else an automatic force-based layout is provided to compute these coordinates. Style of elements can be customized using a CSS-like stylesheet. Beside the huge family of existing graph libraries, GraphStream focuses on providing additional tools to manage dynamics, but the integration of GraphStream with other libraries is eased by a great number of import/export modules including GraphViz and Pajek file formats, shapefile data format used in geographic information systems, and more other file formats.

A dynamical web site, intended to users from various disciplinaries, provides documentation, tutorials and videos. It can be found at :

`http://graphstream-project.org`

2 Practical Lab Schedule

During the GraphStream practical lab session, provided in EPNACS satellite meeting within ECCS'11, Vienna, Austria, on September 15 2011, two parts will be developped:

- Part 1: GraphStream, a java library for dynamical complex networks, concepts and introduction to elementary implementations;
- Part 2: Application to simulations for geographical systems.

3 Gallery

In the following various figures, extracted from the video file produced to illustrate the version 1.0 of GraphStream, are presented and commented.

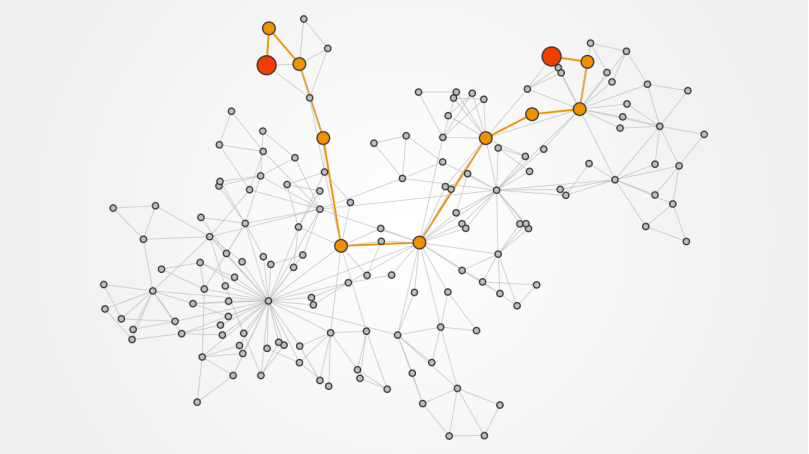


Figure 1: GraphStream algo core: shortest path computation

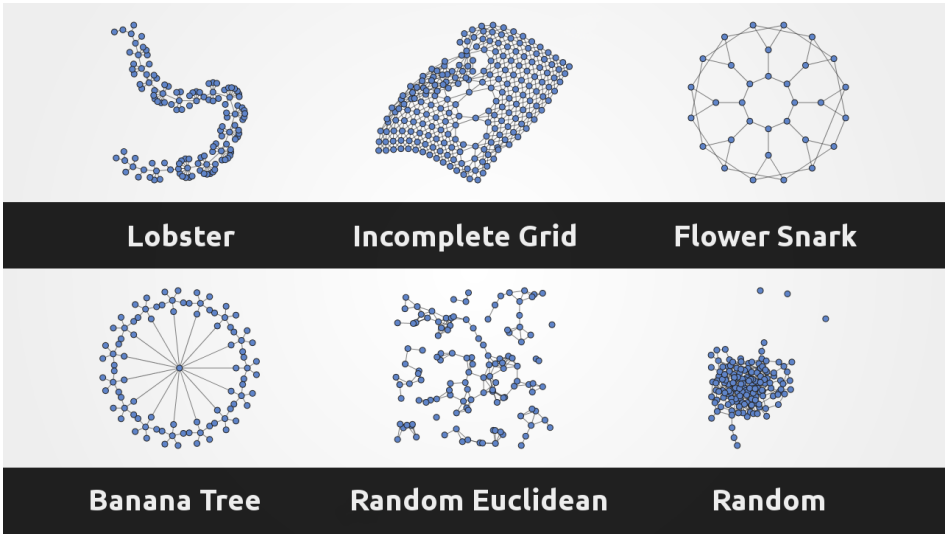


Figure 2: Various graph generators

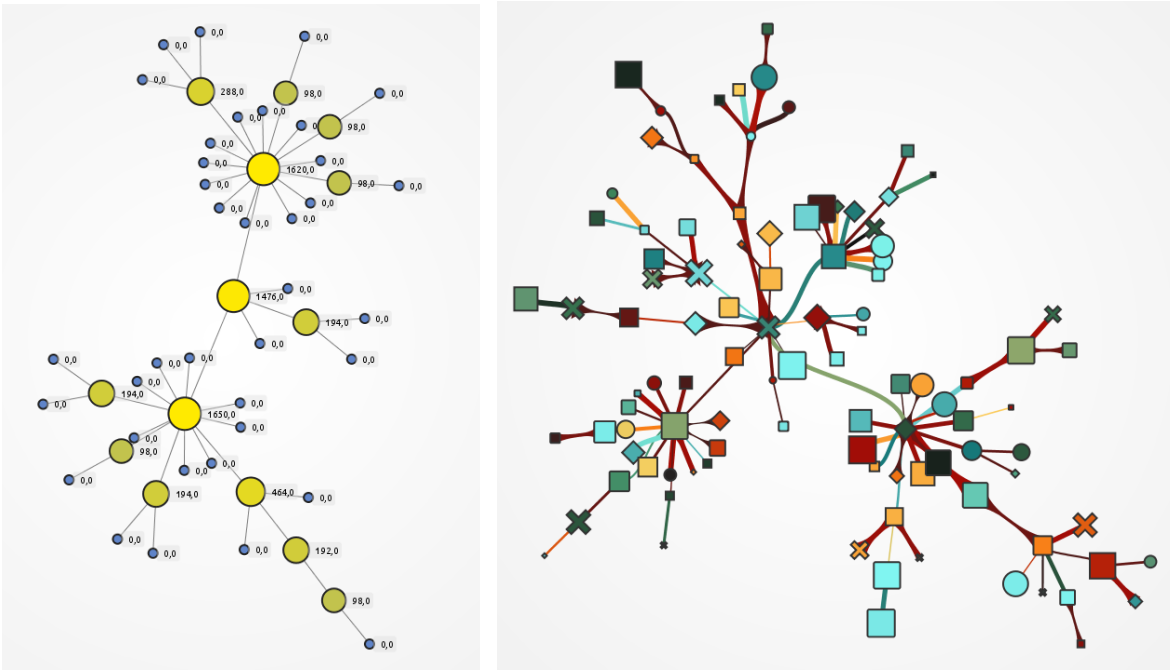


Figure 3: GS algo core; betweenness centrality computation; GS visualization with CSS

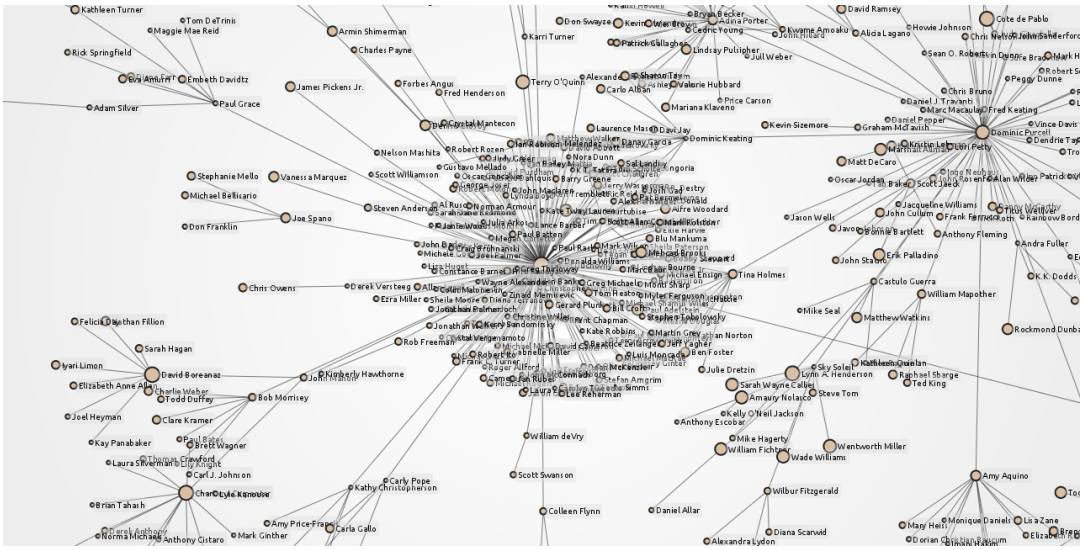


Figure 4: Complex network visualization

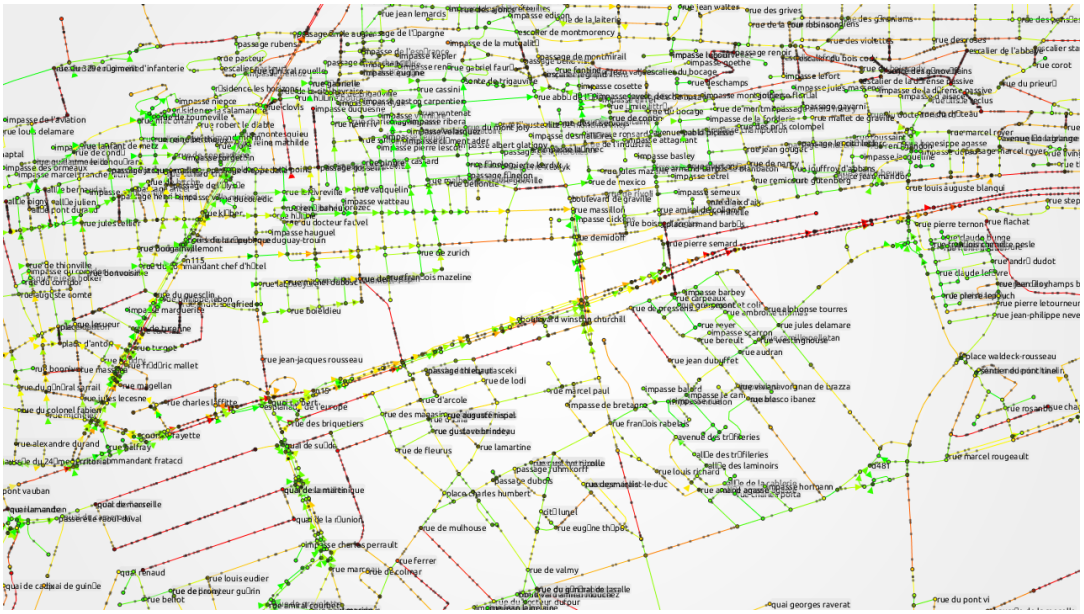


Figure 5: Simulation running on road network extraction from the shapefile of Le Havre city

Index