Transversal Problems in Complexity

MIR-IEA, Université de Cergy-Pontoise, May 23rd, 24th, 2018

Abstracts

Wednesday 23rd, Morning Session: Complex Dynamics I. Evolutionary Game Theory, Population Dynamics applied to Economy and Biological Systems

WM1	Mark Broom	City, University of London
Title	**Game-theoretical modelling of a dynamically evolving network**	

Animal (and human) populations contain a finite number of individuals with social and geographical relationships which evolve over time, at least in part dependent upon the actions of members of the population. These actions are often not random, but chosen strategically. In this talk we introduce a game-theoretical model of a population where the individuals have an optimal level of social engagement, and form or break social relationships strategically to obtain the correct level. This builds on previous work where individuals tried to optimise their number of connections by forming or breaking random links; the difference being that here we introduce a truly game-theoretic version where they can choose which specific links to form or break. This is more realistic and makes a significant difference to the model, one consequence of which is that the analysis is much more complicated. We discuss some general results and then focus on a particular example.

WM2	Yannick Viossat	CEREMADE, Paris-Dauphine
Title	*Learning by imitation and survival of irrational behaviour*	

The replicator dynamics and other imitative game dynamics eliminate irrational behaviours, in the sense that pure strategies strictly dominated by other pure strategies go extinct. By contrast, strictly dominated strategies may survive under a large class of innovative dynamics that is, allowing initially absent strategies to appear. We argue that the fundamental reason for these results is that innovative dynamics favour rare strategies, while currently studied imitative dynamics do not. It is however easy to imagine imitation processes favouring rare or frequent strategies. The resulting dynamics allow for survival of pure strategies strictly dominated by other pure strategies.

WM3	Floriana Gargiulo	GEMASS, Université Paris-Sorbonne
Title	*Driving mechanisms and emerging effects in reputation dynamics*	

Several theories have been developed concerning the microscopic mechanisms driving the emergence of reputation stratification in the society. In particular, some micro-phenomena are particularly important to understand how the micro-interactions among the actors can construct the macroscopic reputation scenarios: homophily, cumulative advantage, memory effects, reciprocity and the heterogeneity of the activity patterns. The data from online social platforms provide an optimal playground for testing these hypotheses. We work with four different datasets containing the detailed information on ratings exchanges between users: Epinions, Slashdot, Advogato and the Bitcoin-OTC. We first verify the hypothesis of the presence of homophily in the peer selection mechanism. Secondly, we analyze the role of social influence, namely the role of the previous micro-interactions on the future score attribution, with a particular attention to the cumulative advantage mechanism.

Third, we show that in symmetric peer interactions a first rating (from i to j) strongly influence on the response rating (from j to i).
At the same time, the dynamical evolution of reputation shows the emergence of interesting hierarchical patterns, at the global scale, and complex behaviors concerning the growth mechanisms of in individual reputations. We will analyze the similarities and the difference among these patterns in the aforementioned databases.
Finally, I will discuss on the structuration of a data driven modelling framework able to make the connection between the microscopic interactions and the emergent effects

WM4	Agnieszka Rusinowska	CES, Université Paris 1
Title	*Modeling Contagion by Aggregation Functions*	

We use the approach based on aggregation functions to investigate the contagion phenomenon in a countably infinite society of individuals interacting with their neighbors. The contagion process is defined as a Markov process with an uncountable set of states, and hence the analysis requires the involvement of sigma -fields. We study in details two types of aggregation functions – strict, and Boolean – and discuss briefly the case of non-strict and non-Boolean aggregation functions. We provide the convergence analysis, determine absorbing and transient sets, and irreducible classes. When considering Boolean aggregation functions, the contagion process becomes deterministic, and the contagion model of Morris (2000) can be seen as a particular case of our framework with aggregation functions. In this case, consensus and nontrivial absorbing states as well as cycles and infinite trajectories can exist.

Joint work with Michel Grabisch and Xavier Venel

WM5	Aymeric Vié	Sciences Po, Saint-Germain-en-Laye
Title	*Selecting Information in Financial Markets: Herding and Opinion Swings in a Heterogeneous Mimetic Rational Agent-based model*	

As expectations are driven by information, its selection is central in explaining common knowledge building and unraveling in financial markets. This paper addresses this information selection problem by proposing imitation as a key mechanism to explain opinion dynamics. Behavioral and cognitive approaches are combined to design mimetic rational agents able to infer and imitate each other's choices and strategies in opinion making process. Model simulations tend to reproduce stylized facts of financial markets such as opinion swings, innovation diffusion, social differentiation and existence of positive feedback loops. The influence of imitation reliability and information precision on opinion dynamics is discussed. The results shed light on two competing aspects of imitation behavior: building collective consensus and favoring innovation diffusion. The role of contrarian and individualistic attitudes in triggering large-scale changes is highlighted. From the results, some policy recommendations to reach better financial markets stability through opinion dynamics management are finally presented.

Wednesday 23rd, Afternoon Session: Complexity in Social Systems

WA1	Jean-Pierre Nadal	CNRS, LPS-ENS, CAMS-EHESS
Title	*Dynamics of the 2005 French riots: a data-driven epidemiological modeling*	

During autumn 2005, after two youth died while trying to escape a police patrol, riots started in a poor suburb of Paris, spread around and then in all France, hitting more than 800 municipalities and lasting about 3 weeks. Remarkably, although there were no displacements of rioters, the riot activity did travel. Thanks to an access to daily national police data, we analyzed the dynamics of riot propagation.
In this talk I will show that a parsimonious data-driven epidemic-like model, taking into account both local (within city) and non-local (through geographic proximity or media) contagion, allows reproducing the full (day by day)

time course of the riots at the scale of the country. I will make explicit the specificity of the model as compared to the modeling of the spread of infectious diseases.
Moreover, I will show that our analysis allows to give a precise mathematical characterization to the expression "wave of riots", and to provide a visualization of the propagation around Paris, exhibiting the wave in a way not described before. The remarkable agreement between model and data demonstrates that geographic proximity played a major role in the propagation, even though information was readily available everywhere through media. This work is the result of a multidisciplinary collaboration involving expertise in sociology, computer science, physics, and mathematics.

Reference: Laurent Bonnasse-Gahot, Henri Berestycki, Marie-Aude Depuiset, Mirta B. Gordon, Sebastian Roché, Nancy Rodriguez & Jean-Pierre Nadal, "Epidemiological modelling of the 2005 French riots: a spreading wave and the role of contagion", Scientific Reports 8, Article number: 107 (2018) http://rdcu.be/H801

WA2	Dimitrios Kotzinos	ETIS-CNRS-UCP
Title	*Understanding Online Personal Networks*	

Online social networks play an important role on how we get our information, make friendships, join groups and hold a significant part of our personal information (like when and where we were born, where we live, what schools we attended, etc.) and our opinions on various subjects. Traditionally the personal information was shared among a preselected group of individuals (usually called "friends", "followers", "connections", etc.). But the emergence of the online social networks gave this group of individuals diverse characteristics. It is no more required that the individuals are personally acquainted or even connected; one can reply to a post (e.g. a tweet) from and start a dialog with someone that she does not know, never met and is not connected to. This makes the existing works in the area of personal or egocentric networks not directly applicable to the online world. And this affects a lot of the functions taking place: information dissemination and diffusion, definition of influential nodes, creation of communities. So, in this talk, we discuss the extension of the definitions of the personal/egocentric networks for the online world and based on those, we try to analyze the evolution of the online personal networks by providing a set of analytical results. This also provides an alternate view on the evolution of the online social networks, seeing not as a global network that evolves but as a collection of individual networks that change over time both independently and collectively. We discuss the applicability of traditional social network properties in this setting and whether the behavior of the personal online networks is compatible with the behavior of the overall social network.

WA3	Arkadiusz Jędrzejewski	Wrocław University of Science and Technology
Title	*Impact of memory on opinion dynamics*	

We investigate an agent-based model of opinion dynamics with two types of social response: conformity and independence. Conformity is introduced to the model analogously as in the Sznajd model or q-voter model, which means that only unanimous group exerts peer pressure on individuals. The novelty, in relation to previous versions of the q-voter model, is memory possessed by each agent and external noise T, which plays the role of social temperature. Each agent has its own memories of past experiences related to the social costs and benefits of being independent or conformist. If an agent was awarded in past more for being independent, it will have a greater tendency to be independent than conformist and vice versa. We will show that depending on the social temperature T the system spontaneously organizes into one of two regimes.
Below a certain critical social temperature T*, all agents in the society acquire personal traits (so called person state). Some of them become permanent conformists and others start to behave forever independently. This means that initially homogeneous population becomes heterogeneous, and agents respond differently to social influence. For T>T*, all agents with equal probabilities behave independently or conform to peer pressure (so called situation state). This regime change between person and situation state, which reminds the idea of an annealed vs. quenched disorder, affects also public opinion. Particularly interesting results are obtained for individualistic societies, in which public opinion is non-monotonic function of T, which means that there is an optimal social temperature for which an agreement in the society is the highest.

WA4	David Chavalarias	CNRS, ISCpif, CAMS-EHESS
Title	*What can be learned from the multi-level reconstruction of online political landscapes dynamics? The French Presidential Elections through the lens of Twitter data*	

We will show how to qualify and quantify the activity of political communities in a multi-polar political environment as well as their temporal evolution through the study of their digital traces.
From the analysis of a corpora of 60 million Twitter exchanges between more than 2.4 million users who interacted with political figures during the 2017 French presidential elections, we characterize the socio-semantic networks of the French political environment, as well as their development over a period of eleven months preceding the election.
This reconstruction provides unprecedented insight into the opinion dynamics and the reconfigurations of political communities, giving access to an intermediate level of resolution, between traditional sociological surveys and large statistical studies (such as those conducted by national or international organizations).
We will show how this type of reconstruction, that are intended to constitute input for social systems modeling, can provide insights into some important societal issues.

Website: http://chavalarias.com & http://politoscope.org

WA5	Radosław Michalski	Wrocław University of Science and Technology
Title	*Sequential Seeding as a way of Reaching the Boundaries of Information Spread*	

We consider here information spread which propagates with certain probability from nodes just activated to their not activated neighbours. A novel approach based on sequential seeding [1] is fundamentally strengthened here with newly proposed coordinated execution for precise comparison of different methods and the formal proof that sequential seeding delivers at least as good results as the single stage seeding does [2]. We have executed experiments that compare single stage and sequential approaches on one directed and six undirected real world graphs to greedy approach [3] and to results obtained with optimal seed set (maximal possible coverage).
The gain in terms of coverage arises when the saved seeds are assigned to nodes not yet activated.
The utilized coordinated execution enables fair comparison of various methods for the same random outcomes for nodes activating neighbours.
Results for greedy based ranking for 50,000 simulations represented as percentage of maximum coverage C_{Max} averaged over individual configurations and ordered by the ratio C_{SN}/C_{Max} are presented in Fig. 1 (A).
Additionally, the upper bound $C_{GreedySN} * e(e-1)$ [3] is for many configurations up to 50% greater than real maximum value observed, Fig. 1 (A1). These results demonstrate that upper limit derived from greedy approach is not tight. Sequential methods always outperform single stage ones, if we consider averaged coverage, for any strategy as well as for every configuration. Moreover, degree SQ is able to cover more nodes than greedy SN, especially for configurations with coverage significantly lower than C_{Max}, Fig. 1 (A1).
There were 8,100,000 individual simulation cases. They are ordered by coverage obtained in the single stage method. The resulting plots demonstrate that C_{SQ} (sequential) performs better than C_{SN} (single stage) in almost every case. The greatest increase is observed for $C_{SN} \in [30\%;85\%]$; simply the space for improvement is larger in such cases.
The global results for all networks, strategies, parameters and random binary choices for each edge to propagate information or not across this edge made at the simulation initialization yield better values of C_{SQ} than of C_{SN} in 96.7% of cases. The increase over 5% was achieved in 20.2% of cases. The results were dependent on node ranking strategy with 96.0% better results for random rankings, 100% for degree based rankings and 93.9% for greedy approach based ranking. The improvement over 5% was observed for 11.2% cases for random strategy

and as much as 38.7% cases for degree-based selection and 10.9% for greedy method. It should be 1 also noted that, surprisingly, the random selection performs slightly better than the expensive greedy selection.

Results for individual cases with respect to maximum coverage are presented in Fig. 1 (B).
The sequential results are localized above the single stage border, filling the space towards maximum (100%). The lowest dispersion of results is observed for random based node selection, Fig. 1 (B1), while it is the highest for degree based rankings Fig. 1 (B2). Moreover, if the single stage coverage is at least at the level of 90%, all strategies are able to provide sequential cases very close to maximum. For lower coverages, only degree-based rankings can improve results so much.
The results revealed an important phenomenon: sequential seeding based on degree selection in 92.2% of cases outperforms single stage greedy approach. A similar improvement is also observed for the averaged values. It means that sequential approach almost always is able to boost performance of the simple degree-based ranking over computationally expensive greedy heuristic. Moreover, $C_{degreeSQ}$ is greater than $C_{greedySQ}$ in 62.6% of cases but for single stage methods such superiority can be observed in only 0.01% of cases.
Hence, the main finding is that the computationally ineffective greedy strategy is suitable for single stage approach, while degree based selection for sequential seeding is significantly better than other selection methods.
In general, the obtained results were dependent on the network profile and parameters of the diffusion. Space between the maximum coverage C_{Max} and the single stage seeding C_{SN} is an area in which sequential approaches deliver improvement. This area on average is $(C_{Max}-C_{SN})=C_{SN}$ so only 25%, of the all simulation cases.
To evaluate the improvement, a gain measure G was defined; it is based on average coverage values: $G = ((C_{SQ}-C_{SN})=(C_{Max}-C_{SN}))*100\%$. It shows what part of the improvement area is reachable by sequential approach. Depending on process parameters and network, the gain varies from 30% to 83%. In general, the greater propagation probability value, the greater coverage and gain, Fig. 1 (C). The same phenomenon arises for seeding percentages, Fig. 1(D). Regarding node selection strategies, the highest average gain 74% was observed for degree based selection, while for greedy and random strategies it is much lower: 27% and 33% respectively, Fig. 1(E). The gain strongly depends on network structure, see Fig. 1 (F). The highest gain (83%) was achieved for network N6, whereas the least gain (30%) was achieved for network N3.

References

[1] Jankowski J., et al., Balancing Speed and Coverage by Sequential Seeding in Complex Networks. Scientific Reports, 7, Article number: 891 (2017), DOI: 10.1038/s41598-017-00937-8.
[2] Jankowski J., et al., Probing Limits of Information Spread with Sequential Seeding, in reviews.
[3] Kempe D., Kleinberg J., Tardos E. (2015) Maximizing the spread of influence through a social network.Theory of Computing11(4):105(147).

WA6	Katarzyna Sznajd-Weron	Wrocław University of Science and Technology
Title	*Phase transitions in models of opinion dynamics*	

During this talk I will present an idea of modeling opinion dynamics and show what could be the source of phase transitions in such a models from the perspective of social psychology. Then I will present one of the most interesting models of binary opinions, the q-voter model, and discuss phase transitions that are observed within this model. Finally, I will show what kind of questions, inspired by social theories and experiments, can be asked within the model.

Tuesday 24th Morning Session: Physical and Mathematical models of Complex Systems

TM1	Yamir Moreno,	BIFI-Universidad de Zaragoza
Title	*Multilayer systems: from diffusion processes to transportation networks*	

Multilayer networks are a new paradigm in the study of complex interconnected systems, as they allow to cast into a single framework the complex topological and dynamical interdependencies between different networked systems. In this talk, we first provide a brief overview of what is the state-of-the-art in network science, paying special attention to multilayer networks. Then, in order to discuss some specific case studies, we present results concerning how to characterize diffusion processes that take place on top of these systems as well as on what new insights can be gained using multilayer approaches to study interdependent transportation systems.

TM2	Jörg Lücke,	Universtat Oldenburg
Title	*The Complexity of Learning with Bayesian Networks and Systems of Statistical Physics*	

Learning is the process of extracting information from data in order to understand and to solve tasks. Systems that are able to learn have complex architectures and dynamics in order to accomplish the advanced computations required. Biological examples are the neural networks of animal brains. Artificial examples are a diversity, very actively investigated networks based on different architectures and probabilistic or deterministic interactions. The research communities of Machine Learning and Statistical Physics share many mathematical models and analytical tools, and use as central object stochastic networks that can change their interaction rules based on the data they process. Bayesian networks and statistic networks such as Boltzmann or Helmholtz Machines are prominent examples. The major challenge for any such systems is represented by their computational complexity, i.e., by the computational demand required for inference and learning - be it in real physical realizations of such networks or by their realizations in digital computers. After discussing the state-of-the-art of such networks and their challenges in the first part of my talk, I will work out what tools are currently deployed in the community to tackle computational complexity problems. I will discuss statistical approximations such as mean field approaches and their use in Machine Learning and Statistical Physics, and will point to recent alternatives.
I conclude by discussing hybrids of Monte Carlo and variational optimization approaches, recent successes of `black box' learning, and by pointing to limitations of current deep learning approaches.

TM3	Christophe Oguey,	LPTM, CNRS-UCP
Title	*Physical support of molecular recognition in biology*	

Molecular recognition is key in many biological processes. I'll focus on some protein-DNA interactions where the DNA intrinsic, sequence dependent, mechanics can be shown to play a role.
Transcription factors (for ex. Jun-Fos or the papillomavirus) target very specific recognition sites, but the affinity is modulated by an intrinsic dynamical signature of the DNA molecule over a domain extending well outside the contact loci. This signature can be coded in a sequence dependent flexibility scale, gathering structural and dynamical characteristics of the B-DNA double helix. Complemented by an interface analysis, this signature also sheds light to the stability and positioning of the nucleosome along the DNA. The nucleosome is the fundamental building block of DNA compaction in the chromatin of eukaryotic cells, a non-specific interaction in a multi-scale organization.

TM4	Alexander Hartmann,	Institut für Physik, Oldenburg University
Title	*Large deviation properties of random graphs and other random models*	

When describing random objects or statistical processes, within analytical and numerical simulations one often just studies the behavior of means or variances. Nevertheless, as for any random process, a complete description is only given if the full distribution of the measurable quantity is available. In this talk sophisticated large-deviation algorithms are described which allow to obtain the distributions of properties of equilibrium and non-equilibrium processes. Probabilities as small as 10^-180 can be accessed using an artificial finite temperature (Boltzmann) ensemble.
Results from this approach are first shown for various properties of random graphs, in particular Erdoes-Renyi random graphs. First, distributions of the size of the largest component, in particular the large-deviation tail, are studied. The distributions for the Erdoes-Renyi ensemble agree well with previously obtained analytical results. Next, the distribution of the size of the 2-core is shown. Third, the distributions of the diameter are presented. Here, partial analytic results are available from previous studies for Erdoes-Renyi random graphs in the small connectivity region. The numerical results follow a Gumbel distribution and agree well with the analytics. For higher connectivities, where no analytic results are available, the simulation results show that the distributions are qualitatively different from the low-connectivity region.
Finally, it is outlined how the approach can be used in general for various problems. Examples which are given here include the distribution of endpoints of Fractional Brownian motion, the distribution of convex hulls for random walks, the distribution of heights of the Kardar-Parisi-Zhang equation, and the distribution of work for an Ising model in non-equilibrium.

TM5	Clàudia Payrató Borràs	LPTM, CNRS-UCP
Title	*Characterization of ecological networks*	

Mutualist ecosystems, like plant-pollinators or plant-seed dispersers, are defined by the fact that the interaction between two agents is naturally beneficial for both. Their network representation corresponds to a bipartite matrix where the two types of nodes represent the plant and the pollinator (or seed dispersers) species and the interactions only take place between nodes of different kind. Observations of such systems in nature have revealed that there is a widespread common feature characterizing the topology of their interaction networks, which far from being random, shows a very peculiar organization, called nestedness. This organization is such that the contacts of a species of degree k are a subset of those of any species of degree k'>k. Interestingly, the same type or organization has also been observed in economic systems, like the World Trade network or the auction or bilateral market networks (eg. Boulogne sur Mer Fish Market).
In this talk we will briefly discuss how the network representation of mutualistic ecosystems provides a useful tool to study both the statistical properties of interactions in the system as well as their dynamical consequences, in particular concerning the evolution of biodiversity under phenological variations induced by climate change.

TM6	Alessandro Torcini,	LPTM, CNRS-UCP
Title	*Ubiquity of collective irregular dynamics in balanced networks of spiking neurons*	

We revisit the dynamics of a prototypical model [1] of balanced activity in networks of spiking neurons. A detailed investigation of the thermodynamic limit for fixed density of connections (massive coupling) shows that, when inhibition prevails, the asymptotic regime is not asynchronous but rather characterized by a self-sustained irregular, macroscopic (collective) dynamics. So long as the connectivity (in-degree) grows proportionally to the size, this regime is found in many different setups: leaky as well as quadratic integrate-and-fire neurons; large and small coupling strength; weak and strong external currents.

References:

[1] S. Ostojic, Nat Neurosci 17, 594 (2014).
[2] E. Ullner, A. Politi, and A. Torcini, ArXiv e-prints (2017), 1711.01096.

TM7	Elisabeth Logak,	AGM, CNRS-UCP
Title	*A nonlocal system modelling the spread of epidemics on networks*	

After a short introduction on epidemic networks, we consider the example of a SIS-type integro-differential system giving the time evolution of Susceptibles and Infected, which is derived from a discrete model. The topology of the network is given by the degree distribution of its nodes. We obtain analytical results on the system such as global existence and existence of an endemic equilibrium above some optimal threshold value. We investigate the stability of this endemic equilibrium and study the asymptotic behaviour of the solution for large times.

Tuesday 24th Afternoon Session: Complex Dynamics II. Evolutionary Game Theory, Population Dynamics applied to Economic and Biological Systems

TA1	H. Peyton Young,	Department of Economics Oxford University
Title	*The Speed of Innovation Diffusion in Complex Social Networks*	

New technologies typically gain a foothold through the actions of a few innovators, and then diffuse more rapidly as more and more people come into contact with prior adopters.
Much of the prior literature focuses on the rate of diffusion as a function of the topology of a given network. Here we derive "topology-free" bounds on the expected waiting time until a given fraction of the population has adopted the innovation.
The bounds depend on the perceived benefits from using the innovation, and on spillover effects from neighbors' adoption decisions, but they do not depend on the network structure per se. In particular, the bounds hold for directed and undirected networks of arbitrary size whose structure may be evolving over time.

TA2	Peter Vida	THEMA, UCP
Title	*Strategic Stability of Equilibria in Multi-Sender Signaling Games*	

We show that in multi-sender signaling games strategically stable equilibria can be maintained by unprejudiced beliefs. Roughly speaking, in the case of two senders, a belief is unprejudiced if whenever the receiver is surprised he believes in the content of the signal sent by the non-deviant. Putting it differently, a belief is prejudiced at an unexpected event if the receiver thinks that both senders were deviating at the same time, even when the event can be explained by a single deviation.

TA3	Maël Le Treust,	ETIS, ENSEA
Title	*Persuasion with limited communication capacity*	

We consider a Bayesian persuasion problem where the persuader and the decision maker communicate through an imperfect channel which has a fixed and limited number of messages and is subject to exogenous noise. Imperfect communication entails a loss of payoff for the persuader. We establish an upper bound on the payoffs the persuader can secure by communicating through the channel. We also show that the bound is tight: if the persuasion problem consists of a large number of independent copies of the same base problem, then the persuader can achieve this bound arbitrarily closely by using strategies which tie all the problems together. We characterize this optimal payoff as a function of the information-theoretic capacity of the communication channel.

TA4	Elena Verónica Belmega	ETIS, ENSEA
Title	*Online Mirror Descent: An Application to Wireless Communications*	

Reducing energy and power consumption at the device level is one of the major challenges in IoT networks, which are envisioned to connect a large number of autonomous and heterogeneous devices. This problem is exacerbated by two main factors: a) the fact that these devices operate in a highly dynamic environment where traditional power control algorithms no longer apply; and b) the lack of sufficient information at the device end. We

propose a regret-based framework to address the distributed power minimization problem that accounts for arbitrary network dynamics. We then derive online power control algorithms based on online mirror descent which are provably capable of adapting to such changes in the network. An important tradeoff emerges and is further discussed between the amount and quality of the feedback available at the device end and its performance in terms of regret minimization rate.

TA5	Valerio Volpati	CEA, Saclay
Title	*The spatial organization of the population density in cities*	

Although the average population density of a city is an extremely simple indicator, it is often used as a determinant factor for describing various aspects of urban phenomena. On the other hand, a plethora of different measures that aim at characterizing the urban form have been introduced in the literature, often with the risk of redundancy. Here, we argue that two measures are enough to capture a wealth of different forms of the population density. First, fluctuations of the local density can be very important and we should distinguish almost homogeneous cities from highly heterogeneous ones. This is easily characterized by an indicator such as the Gini coefficient G, or equivalently by the relative standard deviation or the entropy. The second important dimension is the spatial organization of the heterogeneities in population density and we propose a dispersion index η that characterizes the degree of localization of highly populated areas. We argue that these two dimensions are enough to characterize the spatial organization of cities, and we discuss this approach using a dataset of about 4,500 cities belonging to the 10 largest urban areas in France, for which we have high resolution data. Representing cities in the plane (G,η) allows us to construct families of cities. On average, compactness increases with heterogeneity, and we find four large categories of cities (with population >10,000 inhabitants): (i) first, homogeneous and dispersed cities with small density fluctuations, (ii) very heterogeneous cities with a compact organization of large densities areas. The last two groups comprise heterogeneous cities with (iii) a monocentric organization or (iv) a more delocalized, polycentric structure. Integrating these two parameters in econometric analysis could improve our understanding of the impact of urban form on various socio-economical aspects