Modeling collective behavior using traces of individual activity

Luigi Lancieri, France Telecom R & D
Luigi.lancieri@orange-ftgroup.com

Abstract: *Our study focuses on the evaluation of consensual behavior phenomena in collective behavior. Our approach is based on the analysis of traces of individualactivity in two different contexts of group interactions. The first relates to individual mobility and regrouping in physical areas, whereas the second is connected to herd behavior in online financial transactions. We show that these two contexts, although they appear in different kinds of environments (physical vs. virtual) correspond, in fact, to comparable socio-cognitive mechanisms. In both cases, the individual's activity is conditioned by that of the group. We propose a simple, power-law-based model, which allows characterization of the collective influence capacity. We also put these discussions in prospective by a review of the literature on collective phenomena.*

Keywords: Datamining, collective behavior, emergence, mobility, decision-making

Introduction

Collective phenomena like herd behavior have long been studied by scientists from many disciplines, such as sociologists, ethologists and computer scientists. We intuitively understand that groups partially inherit and retro-propagate mutual influences from each individual combined to have a collective effect with a specific evolution. In this paper, we discuss this concept from the perspective of computer and telecommunications technologies. These media bring two important changes to collective behavior. In first place, the data-processing support enables us to recover a great quantity of traces of human interactions, which makes it possible to highlight and illustrate behavior which is not always visually observable. Indeed, having the statute of memory, traces make it possible to study a past phenomenon outside the constraints of its natural evolution. An event which lasts a split second can be expanded in time and replayed *ad infinitum*, allowing various forms of observation and interpretation. In the second place, the media has a particular influence on human interactions by modifying them. This influence intervenes, for example, in the lifting of space and time constraints, which can multiply the interactions.

The objective of the analysis of traces is to make quantitative information "speak" or give it a sense. As in the majority of the scientific disciplines which use traces (history, physics, astronomy, etc), the first most critical problem is, undoubtedly, to obtain information! The quantity of data, its format, its granularity, its reliability, the ethical and legal aspects, are all constraints which may call into question the use and interpretation of traces. The second problem, supposing that the traces are of good enough quality, is to analyze them so as to interpret and describe a particular phenomenon. This stage requires us to apply certain scientific steps, implying assumptions and rigorous validation of results, while remaining conscious that the exercise has its limits and that traces cannot tell us everything.

In this article, we show, through two different contexts, how quantitative data resulting from human activity can be useful to supplement our comprehension of collective phenomena. The 2 cases presented relate, firstly, to the formation of physical groups and secondly, to herd phenomena in online financial transactions. Having analyzed these two contexts, we highlight them through a review of the literature on collective phenomena.

Physical implication of collective behavior

Population movement and sedentation phenomena have been studied for a long time by sociologists. The regrouping of dwellings in cities or popular areas is a simple demonstration of this behavior. Indeed, it is observed that the majority of individuals tend to elect residence near other individuals and often use identical routes. The pattern of territory settlement is thus far from random but is rather distributed in clusters with large empty zones. In the same manner, traffic jams are also a demonstration of mass phenomena (Geroliminis and Deganzo, 2007).

In order to better understand these kinds of collective behavior, we carried out an experiment intended to measure individuals' movements in a professional environment. For this, we used a network of 17 WIFI access points, homogeneously distributed around a France Telecom R&D site with more than 300 users. First studies, carried out over a 125-day period, clearly showed that, as in the case of urban behavior, individuals' preferred routes are very few, but are traversed frequently and are shared to a great extent by many individuals (Benayoune and Lancieri, 2006).

This report led us to search for an underlying law with the distribution of the preferential courses and, consequently, a model of area occupation. In the following figures, the level (ord) and the row of frequentation (abs) of an area are bound by a power law. The left-hand curve shows, for each WIFI access point, the ranked level of path crossing (user connection and disconnection). For example, we can observe that the most frequented access point, ranked 17 in Figure 1a, totalized 2600 path crossings. The right-hand curve is given by a logarithm transformation of the left-hand curve on both y- and x-axes. The linearity of the right-hand curve is an important clue for identifying a power law (e.g. Pareto, Zipf-Mandelbrot Law, etc).

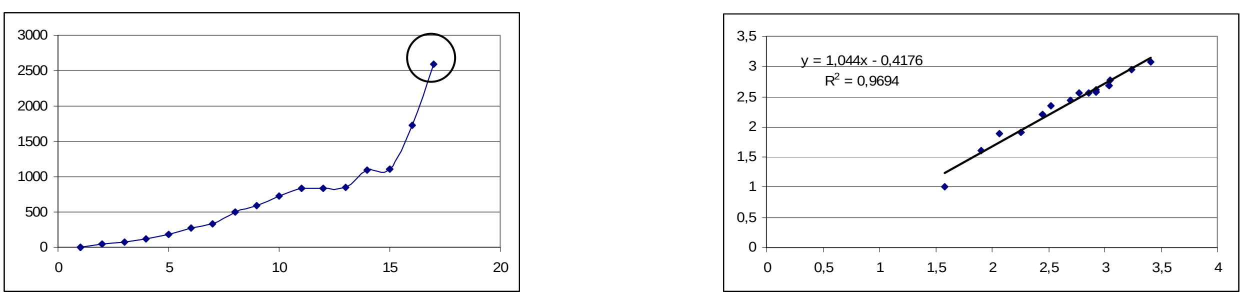

Figure 1: *Rank of area use for each WIFI access point. Real value (left), log transform (right)*

Furthermore, the slope of the right curve can be interpreted as describing the structure of area use. When the slope tends towards zero, all the area points tend to be similarly attended. On the contrary, when the slope tends towards infinity, there are zones of over- and under-frequentation. What is interesting here is to show that a single value (the slope of the right-hand curve, 1.044 here) can describe the structure of area occupation.

Herd behavior and financial transactions

Collective behavior influences individual decision-making, which results in action in both physical and virtual environments. Take the case of financial transactions that materialize a confrontation between purchasers and salesmen. The percentage of evolution of stock prices is the result of the search for equilibrium between convergence and divergence. (0 %: sales and purchases balance; +1% purchases greater than sales, etc). Many studies emphasize that financial transactions often escape the usual human psychological rationale. Popular opinion has always supported this theory. For example, French historian Jacques Bainville, said: "Stock exchange prices are matters of opinion; they reflect ideas, somber or amusing imaginations". The famous American businessman, Warren Buffet said: "The short terms fluctuations of the stock exchange and the value of the company are two completely different universes". Finally, Herbert Simon, Nobel Prize of Economy in 1978 and author of the theory of bounded rationality, said that stock prices are a "projection of anticipations". This idea is important because it shows that the lack of rationality comes from the fact that decision-making is only carried out on the basis of mental constructions inspired by past experience. Intuition and experience show that financial transactions do not escape the rules of collective behavior and, particularly, herd phenomena. Besides, the professionals have proverbs to illustrate this intuition. For instance: "One does not try to catch a knife which falls" or "never invest against the trend". These adages imply that the movement (rise or fall) tends to be a self-fuelling process and to develop and increase the level of risk.

To try to understand these phenomena, we studied the transaction traces on the CAC40 stock index (the 40 largest French companies) over a 4045-day period from October 1990 to March 2006. In order to check the group effect in transactions, we first evaluated in what conditions operators' actions could be influenced by previous days' common activity. The following graph (figure 2.a) indicates the percentage of time over which we can observe a number of consecutive days of identical movement (rise or fall). For example, two consecutives days of movement in the same direction appears 49 % of the time. This frequency decreases quickly and is almost divided by 2 for each additional day. The percentage of appearance of 5 consecutive days in the same direction is only 5%.

In the case of a trend continuing over several days, we can examine the probability of sudden tendency inversion. This is done after computing the number of cases of tendency breaks and continuation after a number of given days of identical tendency. In figure 2.b, we observe, after one day, nearly equal probability of both tendencies (50.6%). This probability then rises to a maximum (60.5 %) after 5 days of consecutive trend and then decreases. At the beginning, with the birth of the trend, operators are undecided as to its continuation. The longer the tendency lasts, the greater the majority of operators who think that the break is close. But, from 6 days of consecutive tendency collective opinion is reversed, creating a new and increasingly important consensus for the continuation of the initial tendency.

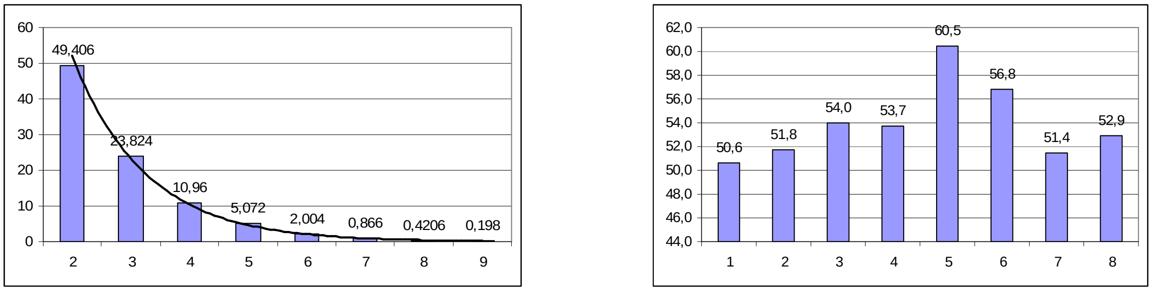

Figure 2: *(a) percentage of occurrence (ord) of consecutive days of identical movement (abs). (b) Probability (ord) of trend inversion after n days of identical movement (abs)*

Related work

The study of collective phenomena in relation to computer sciences could be divided into two main categories, according to whether the phenomena are observed or simulated. Collective behavior is often characterized by retro-propagated influence and consensus, involving emergent behavior. For many authors, these characteristics are closely related to that of complexity.

In the field of social sciences, where observation is the usual experimental tool, authors have pointed out that phenomena of emergence are marked by a paradox between differentiation and standardization. For example, in the emergence of fashion, the motivation to follow the tendency starts from a will to be different (e.g. differentiating factor for companies). However, the final result is standardization because the mass ends up following the tendency. Wanier *et al* speak about a phenomenon of “identical differentiation” (Warnier et al, 2004). This process is characteristic of relations between collectives and individualities. Indeed, when a fashion evolves to the point where it is shared by all, another fashion emerges, again making it possible for individuals to be different. These more or less slow and dependent mechanisms are difficult to forecast in the short term but clearly show long-term cycles.

In a comparable way, Kapur *et al* studied the mechanisms of emergence of collective convergences in newsgroups (Kapur et al, 2006). They developed a measurement of the group activity convergence by computing the number of interactions going in the same direction as *(Np)* and against (*Nn*) the objective laid down for the group *(Cr=Np-Nn)/(Np+Nn))*. In their experiment, they gave gauged problems (precise definition, results known by the observers) to groups of students and observed the interactions in the groups during the experiment. The convergence ratio (*Cr*) was computed at each stage of problem resolution. The authors observed that phenomena of emergence (convergence in the group) appeared very early, in the first 30 to 40% of the time of the group discussion. In addition, statistical tests showed that convergence was a very good guide to the level of performance of the group. The authors insist on the fact that a rule, or a simple quantitative indicator for the level of convergence, makes it possible to describe the behavior of a complex system like group performance very early (reference to the theory of dynamic minimalism).

Discussion

We showed that collective phenomena could be studied using easily-accessible traces of activity with a reasonably good spatial, temporal and semantic granularity. This approach offers the advantage of providing a good statistical representativeness of real phenomena, but requires careful interpretation, and needs to take into account ethical and legal aspects. We looked further into collective behavior of a complementary nature by highlighting phenomena of convergence-divergence in decision-making. We defend the idea that these phenomena take their root in two complementary socio-cognitive mechanisms: forgetting, and value estimate. The forgetting mechanism is involved in irrational behavior, since all of an individual's past experience does not have the same weight in individual's decision-making. Factors such as emotion, recentness or group pressure make the individual attribute non-logical priority to the value of past experiences (forgetting effect). One of the concrete results of this effect is the collective influence through the more or less conscious evidence of group efficiency. Outside the understanding of collective phenomena, we can imagine practical uses of activity traces in term of added value services (Lancieri et al, 2004, 2005, 2006). For example, the detection of emergent behavior in newsgroups, in online purchases services, or in relation with individuals' movements, can provide information about future topics or places of interest. In the context of pervasive activity, the knowledge of individuals' movements can be used to anticipate the optimal location for new services (i.e. information wide screen in the most frequented supermarket aisles). An identical philosophy can be applied to the analysis of online purchases or discussion forums that make it possible to identify emerging fashions. This approach can be useful for technical and economic watchdog strategies.

References

Benayoune, F. and L.. Lancieri (2006); Toward a modelisation of mobile learners behavior, IADIS Interntl. Journal on WWW/Internet.

Deguet J.,Demazeau Y.and L. Magnin (2005); Elements about the Emergence Issue, a survey of emergence definitions. In Proceedings of the European Conference on Complex Systems (ECCS),

Geroliminis, N. and C.F. Daganzo (2007); Macroscopic modeling of traffic in cities; 86th Annual Meeting Transportation Research Board,.

Kapur M., Voiklis J., Kinzer C.K. and J. Black (2006); Insights into the emergence of convergence in group discussions, Proceedings of the 7th Interntl. conference on Learning sciences 2006.

Lancieri L. (2004); Reusing Implicit Cooperation, A novel approach to knowledge management; In tripleC (Cognition, Cooperation, Communication) Interntl Journal, pp 28-46; ISSN 1726-670X.

Lancieri L., Lavallard, A. and P.Manson (2005); E-brainstorming: Optimization of collaborative learning thanks to online questionnaires. In IADIS Interntl Conference Cognition and Exploratory Learning in Digital Age (CELDA).

Lancieri L, and N. Durand (2006), Internet User Behavior: Compared Study of the Access Traces and Application to the Discovery of Communities, IEEE Interntl Journal, Transaction in Systems, Man and Cybernetics (T-SMCA).

Mendelbrot B (1997), Fractal and scaling in finances, Telos edition.

Warnier V,and X. Lecocq (2004) ; L'émergence de la mode dans un secteur: une stratégie collective. Le cas du Prêt-à-Porter. Actes de l'atelier Stratégies collectives.