AGjDPG 2: Focus Session: Big Data

Time: Monday 15:00–17:00 Location: HE 101

Invited Talk AGjDPG 2.1 Mon 15:00 HE 101
Cities and Complexity — •Michael Batty — CASA, UCL, Gower Street, London WC1E 6BT

Cities, like many human systems, evolve as the product of a multitude of individual decisions concerning location and movement, generating order that emerges from the bottom up. In the last decade, they have been used as exemplars par excellence of many features that now define the complexity sciences: interacting dynamic systems, far-from-equilibrium, with strong path dependence, and surprising and novel behaviours. Cities are thus the crucibles of innovation in the economy and society and have become ever more central to the way we articulate our understanding of human systems. In parallel to these concerns, cities appear to becoming even more complex. New forms of behaviour are being generated largely through the development of new information technologies which enable individuals to communicate in countless novel ways a for example in the development of social media, while new forms of city-wide data are emerging as ICT is being fashioned into new systems underpinning the wired city. Transport of all kinds if being revolutionized by the import of ICT and in the near future, it is likely that the development of new forms of urban econophysics dealing with urban markets for land, housing as well as specific markets involving the production and consumption of goods at the spatial level will become the subject of the city focus. In this talk, I will summarise three of these developments: cities and the complexity sciences the rise of big data and the city, and smart cities.

Invited Talk AGjDPG 2.2 Mon 15:30 HE 101
Experimental Computational Finance & Big Data Environment — •Philip Treleaven — University College London, UK

High-frequency algorithmic trading is growing rapidly accounting for 70% of US equity volumes in 2010 (according to Reuters and Bloomberg), but is also of major concern due to potential catastrophic "Flash Crashes". Likewise, the behavior and risk of individual trading algorithms is poorly understood.

For the past seven years UCL has worked with the major investment banks and funds developing algorithmic trading systems, and more recently with the regulators investigating high-frequency trading risk and systemic risk. To support this work we have developed our Algorithmic Trading & Risk Analytics Development Environment (ATRADE) platform which can be used both for virtual and real trading and has access to terabytes of "big" data. It has been designed to: a) speed the development of trading algorithms, b) evaluate algorithm risk, and c) assess algo programmers. As an evaluation of the performance of ATRADE, in 2011 it was used to support a global algorithmic trading competition which attracted over 300 traders in 100 teams scattered across Europe, North America and Australia. Moving forward, UCL is now extending ATRADE with a comprehensive social media engine that supports scraping and analyzing of a wide range of social media data (called Social media Streaming, Repository and Analytics Manager (SocialSTREAM) platform). This presentation will present ATRADE and SocialSTREAM.

Invited Talk AGjDPG 2.3 Mon 16:00 HE 101
Embedding high dimensional data on networks — •Tiziana DiMatteo — King's College London

In this talk I will introduce a graph-theoretic approach to extract clusters and hierarchies in complex data- sets in an unsupervised and deterministic manner, without the use of any prior information [1,2]. This is achieved by building topologically embedded networks containing the subset of most significant links and analyzing the network structure. For a planar embedding [3] this method provides both the intra-cluster hierarchy, which describes the way clusters are composed, and the inter-cluster hierarchy which describes how clusters gather together. I will discuss performance, robustness and reliability of this method by investigating several synthetic data-sets finding that it can outperform significantly other established approaches. Applications to financial data-sets show that industrial sectors and specific activities can be extracted and meaningfully identified from the analysis of the collective fluctuations of prices in an equity market.

[1] Won-Min Song, T. Di Matteo, T. Aste, Discrete Applied Mathematics 159 (2011) 2135. [2] Won-Min Song, T. Di Matteo, T. Aste, "Hierarchical information clustering by means of topologically embedded graphs", PLoS ONE (2011). [3] M. Tumminello, T. Aste, T. Di Matteo, R. N. Mantegna, PNAS 102, n. 30 (2005) 10421.

Invited Talk AGjDPG 2.4 Mon 16:30 HE 101
Econophysics and social research with large sets of data — •Rosario N. Mantegna — Dipartimento di Fisica, Università di Palermo, Viale delle Scienze Ed. 18, 90128, Palermo, Italy

The interaction between physics, economics and social sciences, that econophysicists have pursued over the past twenty years, is based on the idea of using statistical physics methods and paradigms to describe economic and social systems. This interaction is not new, for example, Daniel Bernoulli introduced one of the key concepts of Economics, i.e. the utility function while he was also contributing to hydrodynamics. The interaction was not only in the old past but it has been present in other periods of the history of science. For example, it was proposed again, in the last century, by Ettore Majorana in its tenth article published posthumously in 1942. However, the recent contribution of econophysicists in the analysis and modeling of economic and social systems presents specific aspects because nowadays economic and social sciences have switched from disciplines characterized by a low rate of data production to disciplines with a high rate of data production. Today the society produces an enormous amount of data from scientific, social and business activities. The latter type of data is a gold mine for the study of economic and social phenomena when searching for stylized facts (a technical term used to indicate statistical regularities). The data mining of these data allows the building of an empirical base that is used in the modeling of economic and social phenomena taking into account statistical regularities observed empirically.