

**Institution:** King's College London

**Unit of Assessment:** Department of Mathematics

**Title of case study:** New tools to study complex data sets

**1. Summary of the impact** (indicative maximum 100 words)

Research of Tiziana Di Matteo on network-based filtering techniques has led to powerful new tools for the characterization of dependencies in large complex data sets. This has generated impact on practitioners and professional services in the biotechnology industry and with financial regulators. The Swiss biotechnology firm THERAMetrics Holding AG has used Di Matteo's techniques for developing a quantitative methodology to validate their knowledge based research platform for drug repositioning research. Within a consultancy project awarded to her by the Financial Services Authority (FSA), the information filtering techniques were used to provide advice on methodological correctness of Econophysics techniques applied to a market cleanliness event study.

**2. Underpinning research** (indicative maximum 500 words)

Tiziana Di Matteo's (TDM) main research area is Complex Systems and Econophysics, including the application of methods from statistical physics and network theory to economic modelling, and the analysis of financial markets and social problems.

She was, in particular, the first scientist to propose analysing complex financial datasets (correlation and autocorrelation matrices of interest rates and stock market indices) from the perspective of geometrical and topological properties of metric graphs, embedded in spaces of appropriate dimensions and curvature.

This proposal addresses a problem one generally faces when observing the behaviour of large complex systems, namely that relevant features in such systems are typically *both* local and global, and that these different levels of organization emerge at different scales in a way that is intrinsically not reducible. It is therefore essential to detect clusters together with the different hierarchical patterns of dependencies both above and below the cluster levels. Graph embedding techniques have provided efficient tools to solve this task for a wide variety of complex systems, including financial and biological systems.

Specifically, in the last few years, TDM and collaborators have been focusing on the *dynamical* characterization of correlated financial data in terms of graphs [1,2], studies that, apart from their scientific interest, are very relevant to risk estimation and portfolio selection. One of the main research interests addressed in these studies concerns the *extraction* of meaningful information concerning the behaviour and interactions between variables describing a system under study, often from data sets containing a *high level of redundancy*, and to use such information to *model and forecast* their collective evolution. The main methodological outcome of these studies has been the discovery of a new method to filter information out of complex datasets [5]. Recent developments have shown that this approach can be used to extract clusters and hierarchies from high-dimensional complex data sets in an *unsupervised* and deterministic manner, *without the use of any prior information* [3,4]. This network-based approach to information filtering has opened new ways to study financial systems and also several other fields where a large number of interrelated variables are concerned, such as inference in biomedicine [6].

### Key Researchers

1. Dr Tiziana Di Matteo
  - King's College, since 01/09, Reader in Financial Mathematics
2. Ruggero Gramatica
  - King's College London, since 01/10, PhD Student (PT)
  - 04/10 – 09/13. CEO of MondoBiotech AG (now THERAMetrics Holding AG)

### 3. References to the research (indicative maximum of six references)

1. T. Aste, W. Shaw, T. Di Matteo, "Correlation structure and dynamics in volatile markets", *New J. Phys.* 12 (2010) 085009. DOI:10.1088/1367-2630/12/8/085009.
2. T. Di Matteo, F. Pozzi, T. Aste, "The use of dynamical networks to detect the hierarchical organization of financial market sectors", *The European Physical Journal B* 73 (2010) 3-11. DOI: 10.1140/epjb/e2009-00286-0
3. Won-Min Song, T. Di Matteo, T. Aste, "Nested hierarchies in planar graphs", *Discrete Applied Mathematics* 159 (2011) 2135-2146. DOI: 10.1016/j.dam.2011.07.018.
4. \* Won-Min Song, T. Di Matteo, T. Aste, "Hierarchical information clustering by means of topologically embedded graphs", *PLoS One* 7(3) (2012) e31929. DOI: 10.1371/journal.pone.0031929.
5. \* T. Aste, Ruggero Gramatica, T. Di Matteo, "Exploring complex networks via topological embedding on surfaces", *Physical Review E* 86 (2012) 036109. DOI: 10.1103/PhysRevE.86.036109.
6. \* R. Gramatica, D. Bevec, T. Di Matteo, M. Barbiani, S. Giorgetti and T. Aste, "Graph theory enables drug repurposing - How a mathematical model can drive the discovery of hidden Mechanisms of Action", submitted to *Plos One* (2013); available at <http://arxiv.org/abs/1306.0924>.

Articles marked with an asterisk best indicate the quality of the underpinning research.

### 4. Details of the impact (indicative maximum 750 words)

The research described in section 2 has generated two instances of impact. It has been used by a Swiss biotechnology company to validate principles underlying the construction of a knowledge based approach that allowed the discovery of patterns connecting a certain set of peptides with the occurrence of a set of rare diseases, and it has led to consultancy work done for the Financial Services Authority (FSA) in a drive by the FSA to improve their tool-kit used to carry out market cleanliness event studies.

TDM's consultancy work for the FSA can be understood as a consequence of the fact that she is acknowledged as one of the leading experts in Econophysics, the theory of complex systems, the analysis of financial markets using techniques of statistical physics, network theory and numerical methods, and in particular on the specific information filtering techniques she developed since joining KCL.

The recent financial crisis has led financial institutions to rethink the proper methodologies implemented at that time. In this context, TDM was approached by the Financial Services Authority (FSA) in 2010 to provide advice for a project to strengthen their toolbox used in so called market cleanliness event studies, specifically to test whether new techniques from Econophysics could help to improve the accuracy and diagnostic power of such studies. The cleanliness of markets is

**Impact case study (REF3b)**

important for London as a financial centre. Therefore the FSA undertakes and publishes market cleanliness studies annually, and provides a measure indicative of the level of suspicious trading activity (insider trading) in the London stock market by detecting anomalous trading and price-movement patterns which occur ahead of the release of important information, such as announcements of takeovers or regulatory changes.

TDM's work on network-based filtering has been used by the FSA to cross-validate the analysis of a set of financial events about its cleanliness. The work done by TDM within a consultancy project at the FSA was to provide advice on the methodological correctness and suitability of Econophysics techniques applied to such market cleanliness event studies; this included advice on coding and interpretation of results. Feedback received from the FSA suggests that TDM's contribution to the project was regarded as very valuable in enhancing the FSA's methodical awareness, and that TDM's network based filtering techniques in particular could be used to refine the FSA standard market cleanliness indicator.

A second instance of impact has been generated in the biotechnology industry. In 2009, Ruggero Gramatica (RG) contacted KCL's Financial Mathematics group, to study towards a PhD in Econophysics under the supervision of TDM. After joining KCL in 2010, RG was appointed CEO of the Swiss biotechnology company mondoBIOTECH AG, now THERAMetrics Holding AG, and he quickly realized that the network-related tools and techniques pioneered by TDM and co-workers could be generalized and fruitfully applied to the data-analysis problems of concern to his company, dealing with the discovery of drugs for rare diseases via a knowledge based process of *repurposing already existing drugs*.

Specifically, THERAMetrics was looking for an inferential methodology that could validate their line of research, which deals with automatically extracting bio-medical information on human physiology provided in published works of biochemists or physicians, which would allow the discovery of new Mechanisms of Action (MoA). While biochemists will refer to proteins, receptors, genes and biochemical processes, physicians and health practitioners will mention symptoms, clinical tests, diseases, body organs, tissues, and drugs available for treatment. The central task is to combine such unstructured and dispersed information in a manner that allows relating, for instance, information about biochemical processes with diseases, symptoms and treatments discussed by clinicians.

More than 10 man-years of research were invested at THERAMetrics, starting from the original model proposed by RG, where a knowledge based graph derived from more than 3 million scientific publications, and made of hundreds of thousand of nodes with a very dense set of correlated information is provided, with the aim to search for the non-obvious paths connecting certain molecules to certain diseases, stepping through a number of biological pathways (i.e. new MOAs). TDM's research and expertise, described in section 2, was instrumental to extract such emerging patterns and create new bio-mathematical tools. Using these tools, RG and the scientific team of THERAMetrics have been able to validate a number of the molecules-disease relations that had been present in their candidate pipeline, and in so doing were able to reinforce the scientific foundations of THERAMetrics' drug-discovery platform. Indeed, RG has meanwhile filed IP protection for the general semantic and mathematical model underlying this research.

THERAMetrics Holding AG has been loss-making in the past. However, amongst other factors related to the Company's restructuring plan, the successful results obtained by the above-mentioned

**Impact case study (REF3b)**

methodology helped the company to outline a proper research platform which became a valuable asset in the recent business combination realized by means of a reverse merger with Pierrel's Contract Research International with THERAMetrics Holding AG adding an innovative element in the drug rescuing and repurposing strategy. The takeover/merger was concluded in September 2013 and thanks to such business combination THERAMetrics Holding AG has significantly increased its market capitalization.

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

Financial/economic background concerning THERAMetrics including information about details of the corporate takeover/merger with Pierrel SpA in 2013 at

<http://www.therametrics.com/investor/investors/key-information>

[KCL mirror of THERAMetrics site](#)

Personal sources to corroborate impact at THERAMetrics Holding AG

- Former CEO of THERAMetrics Holding AG (testimonial received and available on request).
- Chief Scientific Officer at THERAMetrics Holding AG

Personal source to corroborate impact at the Financial Services Authority (FSA), now Financial Conduct Authority (FCA)

- Manager, Economics of Financial Regulation, FSA (testimonial received and available on request)