Econophysics Colloquium 2024

20 Year Anniversary – Reassessing Impact and Defining the New Goals

When: June 3 - 5, 2024

Where: Complexity Science Hub, Josefstädterstraße 39, 1080, Vienna, Austria

Program and Abstracts

Talks

Monday, June 3, 2024

9:00 – 9:20 **Tiziana Di Matteo** (*Department of Mathematics, Kings College London*)
Stefan Thurner (*Complexity Science Hub, Medical University of Vienna, Austria*)

Opening

9:20 – 11:00 **Session 1:** Chair **Diego Garlaschelli** *(Theoretical Physics, Leiden University, Netherlands)*

9:20 – 9:40 **Luciano Pietronero** *(Enrico Fermi Research Center, Rome, Italy)*

Economic Fitness: Concepts, Methods and Applications

Economic Fitness and Complexity (EFC) is the recent economic discipline and methodology which makes use and develops the modern techniques of data analysis to build **Economic Models** based the science of Complex Systems to provide a sound **scientific framework**. It consists of a ***data based* and *bottom up*** approach that considers specific and concrete problems without economic ideologies and it acquires information from the previous growth data of all countries with methods of **Complex Networks, Algorithms, Machine Learning and AI**. Its main characteristics are the scientific rigor, the precision in the analysis and in the forecasting, transparency and adaptability. The new Fitness algorithm overcomes the conceptual and practical problems of the early attempts in this field and sets the basis for a testable and successful implementation of the field of Economic Complexity. According to **Bloomberg Views**: **"***New research has demonstrated that the "fitness" technique systematically outperforms standard methods, despite requiring much less data*" In addition EFC has provided a detailed understanding and forecasting of the fantastic growth of China in the past thirty years which has been a major mystery for most of the standard economic analysts. The Economic Fitness represents a synthetic measure of the **degree of competitivity** in terms of the capabilities to produce products and services. Mathematically the Fitness corresponds to the **diversification weigthed by the complexity of the products**. The diversification provides stability and resilience while the complexity of the products represents the exclusivity and the

relative wealth. **The European Commission (Joint Research Center)** has recently adopted these methods to evaluate the recovery fund projects (PNRR): https://publications.jrc.ec.europa.eu/repository/handle/JRC124939
Since a few years it has been used by **IFC-World Bank Group** and recently also by **EBRD (London)** to define specific economic actions tuned for specific countries, in particular for developing ones. One of the main targets is to identify the products or technologies which will **enable to open new markets**, considering the **specific situation of each country**. In this respect there is a complementarity with the **New Structural Economics developed by Justin Lin** and collaborators. An example for African countries can be found here: https://www.ifc.org/wps/wcm/connect/fb4761f5-809b-4685-8fd7-24bd23bad6d3/EMCompass-Note-88-West-African-Industrial-Development.pdf?MOD=AJPERES&CVID=ngxrg.e
Another concept which is important in this field is that of **Relatedness** which describes the dynamics in the **Product Space**. The usual approach is often based in the co-occurrency of products in the basket of countries. The problem with this approach is that the products are 5200 while the countries are only about 150. This leads to a serious problem of signal to noise that makes this approach not far from a random choice. In order to obtain concrete and testable results it is necessary to resort to much more sophisticated **Machine Learning** methods. The validity of these methods has been tested in detail with out of the box approaches. In this way it is possible to identify the possible trajectories for development which are characterized by the **Feasibility**, which describes how easy is to go in that direction and the **Complexity Gain** which is essentially the microscopic increase of Fitness of a given trajectory. This information leads to a scientifically based knowledge which defines the framework for the decisions of policy makers.
EFC for Companies. Also for companies the EFC analysis leads to a variety of original results. Companies show a **block-nested** pattern with respect to the matrix of the products which requires a different analysis with respect to the country matrix which is fully nested. The Fitness algorithm can be applied within each block to define the **Company Fitness**. Then from the patents one can obtain the technological network and introduce the concept of **coherency** for a groups of technologies related to a specific product. **The Product Progression** identifies the next product or technology that a company may be able to produce and its competitiveness in the various markets. Along these lines one can derive a number of results related to the opportunities to enter a certain market or to develop a new product. Also the analysis and optimization of the **Merging and Acquisition** process can be done with these methods.
Recently we have developed these methodologies to study **the impact of AI on the Job Market**. Introducing the concept of Job Fitness on average we observe an inverse proportionality between Job Fitness and AI impact. However, there are also important outliers which require additional considerations.

9:40 – 10:00 **Andrea Zaccaria** (*Institute for Complex Systems (CNR) - UoS Sapienza, Rome)*

Economic complexity algorithms for companies

The Economic Complexity (EC) approach consists in applying tools from statistical mechanics, complex networks, and machine learning to economics, with special attention to comparing theoretical results with empirical evidence. The starting point is usually a bipartite country-product network or, more in general, a network connecting economic actors with activities. EC provides two key concepts. The first is a measure of the competitiveness of a country in terms of the complexity of its capabilities, for instance, ECI or the fitness measure. The second one is relatedness, which measures the similarity between two economic sectors and the feasibility of entering into a new economic activity for an economic actor (e.g., a country starting to export a product). Both can provide powerful quantitative tools to investigate present economies and design future strategies of development, as proven by their adoption by both the World Bank and the European Commission. Usually, the data considered for the EC investigations are at the country level: for instance, international trade, reshaped as a bipartite country-product network. It is well-known that this network exhibits a nested structure, which is the foundation of different algorithms that

have been used to investigate countries' development and forecast national economic growth. In this talk, I critically review the use of this kind of data by comparing it to a unique database provided by the Italian Institute of Statistics (ISTAT). The ISTAT database connects more than 200000 Italian companies to the products they export. Changing the subject from countries to companies, a significantly different scenario emerges. While a globally nested structure is observed at the country level, a local, in-block nested structure emerges at the level of firms. Remarkably, this in-block nestedness is statistically significant with respect to suitable null models and the algorithmic partitions of products into blocks have a high correspondence with exogenous product classifications. These findings lay a solid foundation for developing a scientific approach to investigate companies and computing the “fitness” of a company. Traditionally, also relatedness is measured using complex networks approaches derived from country-level co-occurrences. I compare complex networks and machine learning algorithms trained on both country and firm-level data. In order to quantitatively compare the different measures of relatedness, I use them to predict future exports at the country and firm level, assuming that more related products have a higher likelihood to be exported in the near future. Our results show that relatedness is scale-dependent: the best assessments are obtained by using machine learning on the same typology of data one wants to predict. Moreover, while relatedness measures based on country data are not suitable for firms, firm-level data are quite informative also to predict the development of countries. In this sense, models built on firm data provide a better assessment of relatedness with respect to country-level data. In conclusion, I show the importance of considering data at the company level when applying complexity and network-based concepts to economics.

10:00 – 10:20 **Hiroyasu Inoue** *(Graduate School of Information Science, University of Hyogo and RIKEN Center for Computational Science)*

Establishment-level production network and simulations

Supply chain disruptions are threats for the real economy. The disruptions can be propagated to other firms through production networks. When a disruption reduces production of some firms due to exogenous shocks, such as disasters, pandemics, or national conflicts, the suppliers of the firms directly affected by the shock must reduce their production because of a lack of demand, and their customers must also shrink production because of a shortage of material, parts, or components (Carvalho et al., 2016; Barrot and Sauvagnat, 2016). As a result, even if the initial shock is small, it can lead to a substantial indirect effect, often more substantial than the direct effect of the shock itself (Tierney, 1997; Pelling et al., 2002) and hence large fluctuations across the economy (Bak et al., 1993; Delli Gatti et al., 2005; Acemoglu et al., 2012). Some earlier works rely on inter-industry analysis based on input–output (IO) tables to estimate propagations of shocks (Acemoglu et al., 2012; Haimes and Jiang, 2001; Okuyama and Hewings, 2004).
Since the IO tables aggregate networks between firms, they cannot be used to see correct substitutability between suppliers. This means that chained collapses do not happen and the estimations tend to be smaller than reality (Inoue and Todo, 2019) or erroneous (Diem et al., 2024). To deal with this issue, the estimations based on firm-level production network data have bloomed, especially accelerated by the recent availability of data (Poledna et al., 2018; Inoue and Todo, 2019; Diem et al., 2022; Pichler et al., 2022). Although these studies revealed complex behaviours of production networks, firm-level production networks still have drawbacks. When disasters or pandemics occur, shocks are given to firms and exact locations should be identified. However, if a firm does not involve locations of establishments, and generally has a headquarter’s location at best, it is obvious that the estimation is unreliable.
To fill this gap, we create establishment-level production network. Precisely, we focus on manufacturing sector in the presentation. To do so, we use the Company Information Database and Company Linkage Database collected by Tokyo Shoko Research that includes 1,520,605 firms and 5,860,726 supplier–client relationships in 2020, although we use manufacturing sector from them. In addition, we also use Economic Census for Business Activity conducted by Ministry of Economy, Trade, and Industry that includes 512,401 establishments in manufacturing sector in 2021.

Importantly, the latter data includes product information establishments produce, which enables us to decompose inter-firm linkages into inter-establishment linkages through the inference of recipe. Consequently, we obtain a manufacturing production network that covers 183,951 establishments in 157,537 firms and 919,982 inter-establishment linkages.
We apply a probabilistic model for supply chain disruptions to our network and assess various models to understand their responses to different propagation probabilities. Our models include: (1) a firm-level network model that ignores products and establishments; (2) a firm-level network model that considers products but not establishments; and (3) an establishment-level network model that accounts for both products and establishments.

Our findings are as follows: (1) The establishment-level model, which includes both establishments and products, exhibits a more robust response to disruptions compared to the firm-level model that only considers products. This robustness is due to the ability to distribute propagation across multiple establishments within a firm. (2) The traditional firm-level model, which does not consider either establishments or products, shows a lower overall impact from shocks when propagation probability is low. However, with high propagation probability, this model demonstrates a significantly higher impact compared to the others. This result, which resembles interaction effects, appears to stem from the network's core and peripheral structures. Nodes with high connections form the network's core and typically involve a diverse range of products. Models that include products offer substitutability of inputs. At low propagation probabilities, the core might not always be involved into a propagation. Conversely, at high propagation probabilities, the core is invariably involved, which suggests that models including products tend to show lower impacts.

10:20 – 10:40 **Emanuele Bazzichi** *(IMT - School for Advanced Studies, Lucca, Italy)*

Scale-Consistency and Coarse-Graining of Relatedness

Methods for complexity economics are increasingly used to predict economic performance in different areas and at different scales of analysis, from the micro to the macro level. Recently, alternative methods based on machine learning have been introduced to improve the predictive power of the relatedness principle. However, the scale consistency of this approach has not yet been investigated. In this paper, we introduce a method to measure the scale-consistency of different models across different levels of aggregation and validate our approach using trade data. We also test whether coarse-grained predictions, obtained from finer data and then aggregated to the corresponding macro-classes increase usefulness of the applied models. First, we find that among the machine learning methods considered, non-negative matrix factorization and relatedness are the most consistent. Second, finer information (6 digit data) leads to better results and third, the most consistent models are also the ones that benefit from coarse-grained predictions.

10:40 – 11:00 **Peter Klimek** (*Complexity Science Hub, Vienna, Austria)*

Mapping and Analyzing Supply Networks: From Firm-Level Dependencies to Global Food Supply Shocks

With recent advances in large-scale data collection and processing, it is now possible to map production networks at scale, providing a data-driven understanding of critical dependencies, their resilience, and how these properties evolve over time. In this talk, I will discuss recent work on firm-level supply networks, ranging from their reconstruction from large-scale web data to the assessment of regulatory risks in the context of the Corporate Supply Chain Due Diligence Directive. I will also present recent work on an adaptive multilayer shock propagation model for the global food supply, where shocks propagate through both production and trade links and adaptation strategies are inferred directly from the data. We show that due to this adaptation, the superposition of multiple shocks can lead to subadditive losses in the

shocked items, as opposed to superadditive losses across the entire spectrum of food products, revealing a substantial and non-trivial impact of the network's adaptive capacity on its resilience.

11:00 – 11:30 **Coffee break & Poster presentations**

11:30 – 13:30 **Session 2:** Chair **Peter Klimek** (*Complexity Science Hub, Vienna, Austria)*

11:30 – 11:50 **Matthias Raddant** (*Complexity Science Hub, Vienna, Austria)*

The dynamics of diversity on corporate boards

Diversity in leadership positions and corporate boards is an important aspect of equality. It is important because it is the key to better decision-making and innovation, and above all, it paves the way for future generations to participate and shape our society. Many studies emphasize the importance of the visibility of role models and the effect that connectivity has on the success of minorities in leadership. However, the connectivity of firms, the dynamics of the adoption of minorities, and the long-term effects have not been well understood. Here, we present a model that shows how these effects work together in a dynamic model that is calibrated with empirical data of firm and board networks. We show that homophily -- the appointment of minorities is influenced by the presence of minorities in a board and its neighboring entities -- is an important effect shaping the trajectory towards equality. We further show how perception biases and feedback related to the centrality of minority members influence the dynamic. We find that reaching equality can be sped up or slowed down depending on the distribution of minorities in central firms. These insights bear significant implications for policy-making geared towards fostering equality and diversity within corporate boards.

11:50 – 12:10 **Zlata Tabachova** (*Complexity Science Hub, Vienna, Austria)*

Supply chain contagion adjusted financial climate stress testing

EU Emission Trading System II (EU ETS II) will be introduced in 2027 with the price cap of 45 euro/t it will cause up to 945mn euro of additional costs to the commercial sector. It is unclear to what extend this climate transition policy will stress firms within country’s supply chain network and how the stability of the financial system, hence, will be affected. So far, these assessments had to rely on industry sector affiliations of firms to determine banks exposures to climate policies, e.g., the Climate Policy Relevant Sectors (CPRS) scheme.
Here we utilise a unique firm-level supply network data set that allows us to estimate the fossil fuel consumption of over 160 thousand companies that do not report their emissions yet, up from 140 firms included in EU ETS I. These new emitters have loans in banks worth of 10bn euro. Based on the emission estimates we test the resilience of firms to additional carbon costs by comparing them to their liquidity and equity buffers, and identify which firms cannot withstand the cost shock. Then we assess the effects on their suppliers and buyers with a supply chain network contagion model, and calculate additional losses to banks from firms’ indirect defaults.
Our results show that the modest CO2 price seems to have only very mild consequences on financial stability even when considering indirect effects of supply chain contagion. However, a sudden large jump in CO2 price could cause larger losses if the firms don’t transition in time. These findings suggest that countries need to make sure that systemically relevant firms manage the transition fast to avoid large supply chain induced losses from the green transition.

12:10 – 12:30 **James McNerney** *(Growth Lab, Harvard Kennedy School of Government, USA)*

A mathematical theory of technology scaling

Scale effects are ubiquitous in technologies but general theories of their scaling are largely absent. Here, we present a mathematical theory of technology scaling that captures flexibility in how a design can be scaled, and makes predictions about physical and economic aspects of this scaling. Optimal scaling minimizes the second derivative of the technology's cost curve. Consistent with many technologies, the theory predicts that (i) cost curves will tend to resemble power laws; (ii) deviations from power-law behavior will entail positive curvature; (iii) the structure of costs will tend to scale uniformly, even though (iv) the physical design of technology will likely scale non-uniformly. For technologies within a broad class we identify, the theory contains a mechanism to drive the cost curve toward power-law behavior. Under certain conditions, the Cobb-Douglas production function emerges from optimizing with respect to physical constraints. We present data on the cost curves of 49 technologies and other empirical examples to ground the theory and offer evidence.

12:30 – 12:50 **Alessandro Pluchino** *(Department of Physics and Astronomy "E.Majorana" of the University of Catania)*
Talent vs Luck (pluchino.it)

Exploring success in economics, science and sports: the TvL model legacy

The axiom at the root of the meritocratic paradigm is that individual success has to be rewarded since it is mainly due to talent, i.e. to personal qualities such as intelligence, smartness, efforts, hard work or risk taking. Even if we are willing to admit that a certain degree of luck could also play a role in reaching significant achievements, it is rather common to underestimate and often neglect the importance of external forces in individual successful stories. Starting from the observation that wealth distribution is typically heavy-tailed, while individual talent – as the majority of human features – is normally distributed, in 2018 we presented a computational model (called TvL, Talent vs Luck, model [1]) able to show that the combination of random external factors and multiplicative dynamics in capital or reward accumulation usually do not drive towards the highest peaks of success the most talented people, being the latter very often overtaken by averagely talented but sensibly luckier individuals. In this talk, with the help of several studies that have collected the legacy of the Tvl model [2-7], we will show quantitatively how this phenomenon also governs the achievement of success and the emergence of superstars in economics, in scientific production and in individual sports, thus underlining the limits and pitfalls of the meritocratic paradigm.

[1] A.Pluchino et al. "Talent vs luck: The role of randomness in success and failure", Adv. Complex Syst. 21(03–04) (2018) 1850014 - http://www.pluchino.it/talent-vs-luck.html
[2] D.Challet et al. "On the origin of the extreme wealth inequalities in the Talent vs Luck model", Advances in Complex Systems Vol.23 (2020) No.02 2050004.
[3] A.E.Biondo et al. "Simulating the Emergence of Superstar Firms: the role of Luck vs. Talent", Under Review
[4] A.Pluchino et al. "Exploring the Role of Interdisciplinarity in Physics: Success, Talent and Luck", PLOS ONE 14(6): e0218793 (2019)
[5] P.Sobkowicz et al. "Inequalities, chance and success in sport competitions: simulations vs empirical data", Physica A 557, 12489 (2020)
[6] C.Zappalà et al. "On The Role of Chance in Fencing Tournaments: an Agent-Based Approach", PLoS ONE, 17(5): e0267541 (2022)
[7] C.Zappalà et al. "The Paradox of Talent: how Chance affects Success in Tennis Tournaments", Chaos, Solitons and Fractals, Vol.176, 114088 (2023)

12:50 – 13:10 **Joseph Bradley** *(Department of Mathematics, Kings College London)*

Cluster-Driven Momentum & Reversal

Momentum in Finance is the phenomenon where stocks which have performed relatively well recently will on average continue to relatively outperform, and vice-versa [1]. Reversal is the opposite - stocks that have done relatively well recently will soon underperform [2]. However, authors still debate what drives momentum and reversal returns [3].
This work contributes to the debate by applying information networks filtering tools from Econophysics to explore momentum and reversal portfolios. In particular, we question whether industries are the best groupings in which to gather stocks when forming such portfolios. We use real financial data for 16,969 stocks traded worldwide across several markets from 1994 to 2023. Using a dynamic out-of-sample analysis, we construct dynamically evolving clusters that we can use as proxies for traditional industrial classifications.
We show that clustering methods including the Directed Bubble Hierarchical Tree (DBHT) [4] can explain key properties of momentum and reversal portfolios. We show that whilst using DBHT clusters to construct momentum portfolios is not profitable, they are useful for reversal portfolios, and outperform traditional industry adjusted reversal portfolios. We investigate why DBHT clusters perform well only for reversal portfolios, and show that the temporal structure of traditional industry classifications is too complicated to be captured by clustering stocks on their daily returns. For robustness, we compare the DBHT with traditional partitioning and hierarchical clustering procedures and find that the DBHT out-performs both.

[1] N. Jegadeesh and S. Titman, 'Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency', The Journal of Finance 48, 65–91 (1993).
[2] A. W. Lo and A. C. MacKinlay, 'When Are Contrarian Profits Due to Stock Market Overreaction?', Review of Financial Studies 3, 175–205 (1990).
[3] T.Wiest, 'Momentum: what do we know 30 years after Jegadeesh and Titman's seminal paper?', Financial Markets and Portfolio Management 37, 95–114 (2023).
[4[W.-M. Song, T. Di Matteo, and T. Aste, 'Hierarchical Information Clustering by Means of Topologically Embedded Graphs', PLoS ONE 7, edited by Y. Xu, e31929 (2012).

13:10 – 13:30 **Takayuki Mizuno** *(National Institute of Informatics, Tokyo, Japan)*

Generation of synthetic financial and economic data by Generative AI

This presentation explores the innovative use of generative AI to synthesize both financial time series and human mobility trajectories, addressing challenges in financial markets and societal phenomena. In the financial domain, we employ denoising diffusion probabilistic models (DDPMs) to generate synthetic financial time series. This method leverages wavelet transformation to transform time series into color images, facilitating the generation of multiple correlated series such as stock prices, volumes, and spreads, thereby reproducing key statistical properties like fat tails, volatility clustering, and intraday seasonality observed in financial markets. In the societal context, we adopt the framework developed by Mizuno, Fujimoto, and Ishikawa (Front. Phys. 2022) to convert geographical coordinates into a sequence of distinctive location and time interval tokens. Utilizing an autoregressive language model, GPT-2, we train these sequences to sequentially generate individual daily trajectories, incorporating unique tokens representing environmental factors and personal attributes such as meteorological conditions, gender, and age. This dual approach highlights the potential of generative AI to create realistic, multi-dimensional datasets that reflect complex real-world dynamics in both economic and social spheres.

13:30 – 14:30 **Lunch**

*

14:30 – 15:50 **Session 3:** Chair **James McNerney** *(Growth Lab, Harvard Kennedy School of Government, USA)*

14:30 – 14:50 **Mauro Gallegati** *(Università Politecnica delle Marche, Ancona, Italy)*

Economics is barking up the wrong tree. And econophysics?

From the middle of the Nineties, the heterodox approach that goes under the name of Econophysics has grown remarkably and made a powerful interdisciplinary synthesis between economic problems, statistical physics and network analysis methodologies. More extensively, such a synthesis concerns not only economics but also several fields in finance and sociology, always with the aim of understanding real systemic phenomena and of proposing empirical applications, beyond the unsatisfactory explanations of the mainstream approach.
Econophysics is fairly close to Complexity Theory, of which it might be understood as a sub-field. Econophysics evolved with a great deal of methodological instruments, mostly grounded on Statistical Physics. Moreover, differently from AB-modeling, that appears as a promising technique but still in search of a sound mathematical/computational theory, it reached a status of quantitative discipline. Econphysics has evolved due to the effort of physicist and heterodox economists and social scientists that had been motivated by not committing the same error of the mainstream. Either mainstream and heterodox scientists recognize that economies and societies are complex systems but while the formed did not accepted the need of a change of paradigm, as if one would describe a particles system by means of Newtonian ordinary differential equations, the latter accepted the fact that to understand many-particle systems from the bottom-up the tools of classical physics were not appropriated. This happened first in Physics, that still applies classical physics to problems for which it is appropriated, but to understand the nature of matter made of particles it developed statistical and, lately and more recently, quantum physics. This is because science accepts solutions if they are able to explain facts but never assumes a solution as definitive: a solution is always accepted until otherwise proved or until a better solution is found.
The main problem with Econophysics is, nevertheless, the same the formalist approach of mainstream economics faced, and still faces. We may even indulge in the luxury of dealing with analogies between natural and social systems but we must recognize that an analogy without a sound ontology is weak, and may lead to errors. Mainly, the more dangerous one is to focus the attentions too much on methods, techniques and mathematical apparatuses while forgiving the nature of phenomena we are studying. Shortly, as an elementary particle is the first brick of matter an individual is that of any sort of organized society, but the two are different things. Elementary particles behave without will and ambition, they do what Nature wants them to do. On the other hand, an individual behaves the way she can, with her own will and desires, ambitions, means and knowledge. While the latter is able to perform adaptive and evolutionary behavior, to face innovations and structural changes, the former do not.

14:50 – 15:10 **Massimiliano Fessina** *(IMT School for Advanced Studies, Lucca, Italy)*

Firm-level systemic risk on reconstructed production networks

Inter-firm networks have been shown to play a crucial role in the propagation and amplification of economic shocks at the national level [1]: as such, knowledge of these systems is of capital importance in order to properly address economic resilience to disruptive events. Comprehensive data about supply chains is, however, seldom available [2], triggering a flourishing literature about reconstruc- tion models, aimed at predicting the most probable configurations of these networks starting from the partial, available, information [3]. In this work we test different network reconstruction models rooted in the Maximum Entropy framework, namely the *stripe-corrected Gravity Model*

(scGM) [4], the *Input-Output scGM* (IO scGM), introduced in the paper, and the *density-corrected Gravity Model* (dcGM) [5], on the reconstruction the Ecuadorian production network, accessed through a unique dataset provided by the Ecuadorian Tax Office. Differently from previous reconstruction attempts, here we focus on the ability of the models to correctly identify firm-level systemic risk, as measured by ESRI [6]. The scGM displays the best overall reconstruction performance, proving to be able to generate reliable synthetic networks both from a structural and from an economic point of view (Fig. 1), despite requiring very limited information on the empirical network (namely the sector-specific in- and out-strength of firms): when this information is not available, it can be well approximated by the fluxes between industrial sectors, as in the IO scGM. Our findings shed light on the minimal amount of empirical information that is needed in order to generate realistic synthetic production networks, able to capture the systemic risk of an economy.

[1] Acemoglu et al. Econometrica 80(5), 1977-2016 (2012
[2] Bacilieri et al. INET Oxford Working Paper (2023).
[3] Mungo et al. arXiv:2310.00446 (2023).
[4] Ialongo et al. Scientific reports 12(1), 11847 (2022).
[5] Cimini et al. Scientific reports 5(1), 15758 (2015).
[6] Diem et al. Scientific reports 12(1), 7719 (2022).

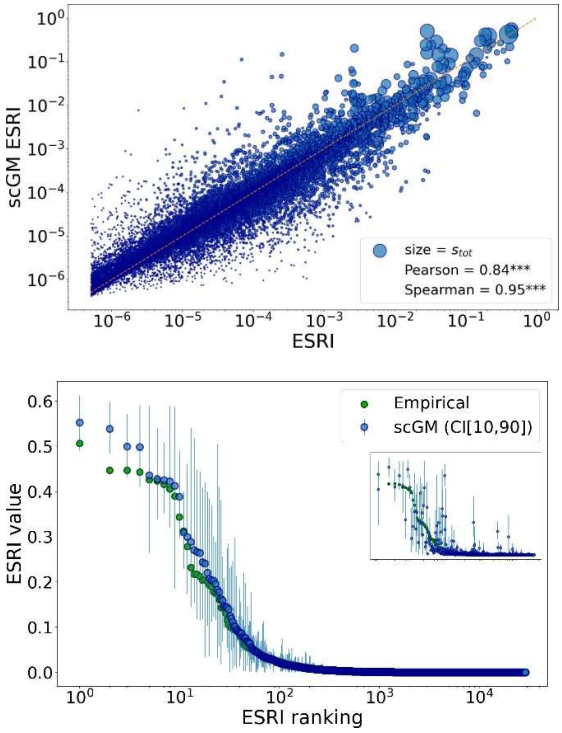


FIG. 1: Left: scGM ESRI values versus the empirical ones, for all the firms in the network. The two quantities are highly correlated, both in terms of values (Pearson = 0.84) and, particularly, in terms of rankings (Spearman = 0.95). Right: scGM and empirical ESRI values versus their ranking. The model correctly predicts the presence of a plateau composed by the most risky firms, which is very close in value to the empirical one. The two plateaus are not composed by the exact same set of firms, as evident when reordering the scGM ESRI values according to the empirical ranking (inset), but there is a great overlap between the two (7 out of the top 10 firms are the same).

15:10 – 15:30 **Arthur Matsuo Yamashita Rios de Sousa** (*Tokyo Institute of Technology, Japan*)

Extraction and clustering of trends in the USD/JPY foreign exchange market using the epsilon-tau procedure

We present the epsilon-tau procedure to segment time series in up- and down-trends [1]. It is particularly suitable for financial time series, making use of two parameters: the tolerance level epsilon, which reflects the risk tolerance of the observer, and the patience level tau, which controls the time scale of the trends (from microtrends to macrotrends). We apply the method to extract and analyze the trends of the U.S. dollar (USD)/Japanese yen (JPY) market time series from 2015 to 2018. After studying the distributions of trend lengths and amplitudes, we investigate their internal structure by clustering the trends by their shapes, i.e., the descriptions of the position of each point relative to the whole trend. We identify a specific down-trend shape that rarely occurs in the randomized data which displays a sharp appreciation of the JPY that is associated with exceptional events such as the Brexit Referendum in 2016 and the 2016 U.S. elections (see Fig. 1). This study suggests the use of the epsilon-tau procedure as a tool for historical analysis of trends in financial markets (and possibly in other fields) as well as hint at the direction of real-time market monitoring.

[1] A.M. Yamashita Rios de Sousa, H. Takayasu, and M. Takayasu, Segmentation of time series in up- and down-trends using the epsilon-tau procedure with application to USD/JPY foreign exchange market data, PLOS ONE 15, e0239494 (2020).

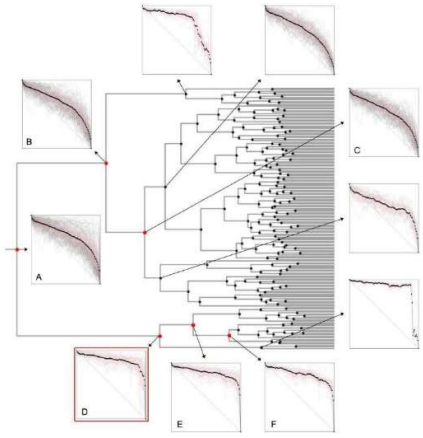


Fig. 1. Clusters showing down- trend shapes that rarely occur in randomized data (red dots). In particular, trend shape in cluster D (in red frame) is associated with exceptional events such as the Brexit Referendum and the U.S. Elections in 2016.

15:30 – 15:50 **Andrea Musso** *(Computational Social Science at ETH Zürich, Switzerland)*

City mobility across the urban hierarchy

Zipf's law for cities—the fact that the distribution of city populations follows a power law—has a long and rich history. First reported by (Auerbach, 1913), this stunning empirical regularity has intrigued a wide array of re- searchers over the years, leading to a plethora of proposed explanations. These include the principle of least effort (Zipf, 1949), random growth (Gabaix, 1999), the distribution of natural advantage (Krugman, 1996; Lee and Li, 2013), migration (Bettencourt and Zünd, 2020; Verbavatz and Barthelemy, 2020), and the nuances of urban structure (Rossi-Hansberg and Wright, 2007), among others.

Zipf's law emerges from city growth dynamics. As countries urbanize and rural populations migrate to urban centers, some cities attract more inhabitants than others. Over time, this results in significant differences in city populations, generating a highly skewed city size distribution. While many explanations for Zipf's law acknowledge its emergence from an underlying growth process, modeling and empirically efforts have

largely focused on replicating the outcome of this process—Zipf's law—rather than process itself.

This emphasis on outcome over process is largely due to a scarcity of relevant data. Numerous datasets capture city populations at a specific point in time, facilitating the observation of cross-sectional patterns like Zipf's law. However, comprehensive datasets covering extensive spatial and temporal ranges and maintain consistent definitions of cities throughout are lacking. Consequently, our understanding of the actual process of city growth—specifically, the trajectories of city population changes over time—remains limited.

In this project, we use the European Union's Global Human Settlement Layer dataset (Schiavina et al., 2023) and the Census Place Project (Berkes et al., 2023) to assemble the most comprehensive city size dataset to date (see Data section below for a description of the data). Utilizing this dataset, we aim to shed new light on the processes governing city population growth. Specifically, we seek to understand what factors that enable and drive the movement of cities up and down the city population ranking? What mechanisms allow some cities to climb these rankings while others decline? What structural features of the economy facilitate significant rank shifts? In pursuing these questions, we document new observations about city growth dynamics and the evolution of city population trajectories over time.

Data. We collected what we believe to be the most extensive database on city populations to date. This database comprises two datasets:

1. Global cities: This dataset features all urban areas worldwide from 1975 to 2020. It is constructed using the European Union's Global Human Settlement Layer initiative, which provides 1 km x 1 km grids detailing population and urbanization levels for each grid segment, largely based on building foot- print data from satellite imagery (Schiavina et al., 2023). We collected all available grids: from 1975 to 2020, at five-year intervals. We then aggregated these grids into clusters using an algorithm along the lines of Rozenfeld et al. (2008, 2011).

2. US cities: This dataset covers all US cities from 1850 to 1940. It is built by integrating the Census Place Project (Berkes et al., 2023) and the IPUMS full count census data (Ruggles et al., 2024). Essentially, we used location information from the Cen- sus Place Project to transform each year's census data into a 1km x 1km grid. We then clustered this grid to obtain consistently defined urban areas (i.e., clusters) from 1850 to 1950, covering the entire US. Since the data originates from the full-count census, we also extracted cluster-level industrial/occupational composition and inter-cluster migrations. Finally, we enriched this dataset by inte- grating the railroad and waterway network database from Donaldson and Hornbeck (2016).

For both datasets, we developed a new technique to match clusters over time, providing a comprehensive spatial and temporal perspective on the process of city growth.

15:50 – 16:30 **Coffee break & Poster presentations**

16:30 – 18:00 **Panel "Glorious Past":** Chair **János Kertész** *(Central European University, Vienna, Austria)*

18:00 – 21:00 **Art & Science event, Complexity Science Hub, Salon**

*

Tuesday, June 4, 2024

9:00 – 10:00 **Session 1:** Chair **Luciano Pietronero** *(Enrico Fermi Research Center, Rome, Italy)*

9:00 – 9:20 **Stefan Thurner** (*Complexity Science Hub, Medical University of Vienna, Austria*)

What can we learn from supply chains at the firm level?

We present an overview over several economic and societal topics that can be addressed with the availability of firm-level supply chain data. We show new possibilities from identifying and quantifying systemic risks in the real economy, how to estimate flows of systemic risk between countries, and applications in the monitoring of basic provisioning systems and policy testing.

9:20 – 9:40 **Christian Diem** (*Complexity Science Hub, Vienna, Austria)*

Estimating the loss of economic predictability from aggregating firm-level production networks

To estimate the reaction of economies to political interventions or external disturbances, input–output (IO) tables—constructed by aggregating data into industrial sectors—are extensively used. However, economic growth, robustness, and resilience crucially depend on the detailed structure of nonaggregated firm-level production networks (FPNs). Due to nonavailability of data, little is known about how much aggregated sector-based and detailed firm-level-based model predictions differ. Using a nearly complete nationwide FPN, containing 243,399 Hungarian firms with 1,104,141 supplier–buyer relations, we self-consistently compare production losses on the aggregated industry-level production network (IPN) and the granular FPN. For this, we model the propagation of shocks of the same size on both, the IPN and FPN, where the latter captures relevant heterogeneities within industries. In a COVID-19 inspired scenario, we model the shock based on detailed firm-level data during the early pandemic. We find that using IPNs instead of FPNs leads to an underestimation of economic losses of up to 37%, demonstrating a natural limitation of industry-level IO models in predicting economic outcomes. We ascribe the large discrepancy to the significant heterogeneity of firms within industries: we find that firms within one sector only sell 23.5% to and buy 19.3% from the same industries on average, emphasizing the strong limitations of industrial sectors for representing the firms they include. Similar error levels are expected when estimating economic growth, CO_2 emissions, and the impact of policy interventions with industry-level IO models. Granular data are key for reasonable predictions of dynamical economic systems.

9:40 – 10:00 **Tobias Reisch** (*Complexity Science Hub, Vienna, Austria)*

Towards a statistical understanding of the evolution of production networks on the firm level

Production networks form the structural core of our society's metabolism. The production, recycling and disposing of intermediate and final goods in any economy is typically spread across several hundreds of thousands of companies that are connected through millions of buyer-supplier dependencies. These production networks (PNWs) are subject to continuous change through the entry and exit of firms. Connections between existing firms are also subject to continuous restructuring, in the form of new buyer-supplier links forming and existing connections being terminated.

In this paper we statistically quantify the temporal evolution of an empirical PNW by examining the entry and exit of firms as well as the formation and termination of buyer-supplier connections. To this end we use monthly reported value added tax (VAT) data in Hungary in the period from 2014 to 2022 containing 711,248 companies and 38,644,400 connections.
We find that annually around 25% of firms exit and 28% new firms enter the Hungarian VAT network, so the number of firms effectively grows with 3.3%. On average, 55% of all links present in one year are not present the next year. However, relative to the previous year, also 61% of new link appear, resulting in an effective growth of 5.8%. The average link is found to have a half-life time of 13 months. New links attach more likely to firms that already have many connections, with slightly superlinear preferential attachment.
To understand how typical PNW meso- and macro-characteristics of production networks emerge from these restructuring processes, we calibrate a simple network generation model that reproduces the stylized characteristics of the Hungarian PNW, with on average 18,800 firms an 28,000 links. The model accurately reproduces the in- and out-degree distribution, as well as the assortativity and local clustering structure. It is realistic enough to reproduce the empirical pattern of the Economic Systemic Risk Index (ESRI).
Our work is useful to improve modelling of rewiring dynamics for shock propagation and can be used to generate time series of monthly PNW snapshots for researchers that don't have access to empirical data.

10:00 – 10:20 **Anatholy Schmidt** *(Finance and Risk Engineering, NYU Tandon School)*

An Impact of Greenhouse Gas Aversion on Optimal Portfolios

The notion of the greenhouse gas (GHG) aversion (GHGA) is introduced into the mean-variance portfolio (MVP) framework. GHGA is assumed to be a weighted sum of the portfolio holdings' GHG emission intensities. A new portfolio performance measure, the GHGA-tilted Sharpe ratio, is offered for GHG-averse investors. While the classical Sharpe ratio may monotonically decrease with growing GHGA, the GHGA-tilted Sharpe ratio has a maximum at intermediate values of GHGA, which defines an optimal GHGA-based MVP. The main holdings of such a portfolio represent promising investment leads for socially responsible investors who do not want to abandon the "brown" industries altogether. An example of a GHGA-based MVP formed with the major constituents of the energy sector is discussed.

10:20 – 10:40 **Fabio Caccioli** *(Dept. of Computer Science, University College London, Systemic Risk Centre, London School of Economics and Political Sciences)*

Upstreamness and downstreamness in input-output analysis from local and aggregate information

Ranking sectors and countries within global value chains is of paramount importance to estimate risks and forecast growth in large economies. However, this task is often non-trivial due to the lack of complete and accurate information on the flows of money and goods between sectors and countries, which are encoded in Input-Output (I-O) tables. In this work, we show that an accurate estimation of the role played by sectors and countries in supply chain networks can be achieved without full knowledge of the I-O tables, but only relying on local and aggregate information, e.g., the total intermediate demand per sector. Our method, based on a rank-1 approximation to the I-O table, shows consistently good performance in reconstructing rankings (i.e., upstreamness and downstreamness measures for countries and sectors) when tested on empirical data from the World Input-Output Database. Moreover, we connect the accuracy of our approximate framework with the spectral properties of the I-O tables, which ordinarily exhibit relatively large spectral gaps. Our approach provides a fast and analytical tractable framework to rank constituents of a complex economy without the need of matrix inversions and the knowledge of finer intersectorial details.

10:40 – 11:00 **Diego Garlaschelli** *(Theoretical Physics, Leiden University, Netherlands)*

Towards the reconstruction of higher-order cycles in economic and financial networks

Economic and financial networks are among the key propagators of shocks and risks across economies, but their empirical structure is typically not publicly available because of confidentiality. This limitation has triggered the development of methods of network reconstruction from partial, aggregate information. Unfortunately, even the best methods available fail in replicating the number of directed cycles, which on the other hand play a crucial role in determining graph spectra and hence the degree of network stability and systemic risk. Here, after reviewing some of the state-of-the art reconstruction methods built on the principles of maximum-entropy and scale-invariance, we explore the hypothesis that the statistics of higher-order cycles is strongly constrained by that of the shortest ones, i.e. by the amount of dyads with reciprocated links. First, we review various empirical results providing evidence that real-world economic networks are characterized by a significant level of link reciprocity. Then, we propose new network reconstruction methods capable of enforcing, only from the knowledge of publicly available node-specific constraints, both a desired sparsity and a desired link reciprocity. We confirm that the addition of reciprocity dramatically improves the prediction of several structural and spectral network properties, including the largest real eigenvalue and the eccentricity of the elliptical distribution of the other eigenvalues in the complex plane. These results illustrate the importance of correctly addressing the level of reciprocity in the reconstruction of financial networks.

11:00 – 11:30 **Coffee break & Poster presentations**

11:30 – 13:30 **Session 2:** Chair **Matthias Raddant** (*Complexity Science Hub, Vienna, Austria)*

11:30 – 11:50 **Rossana Mastrandrea** *(Laboratory for the Analysis of CompleX Economic Systems (AXES), IMT School for Advanced Studies, Lucca, Italy)*

Recurrent patterns in weighted networks: an application to the sectorial integration of energy systems

Network theory has been successfully applied to explore a variety of complex systems [1]. Particularly interesting are local patterns that can shed light on the emergence of global properties, such as motifs [2]. However, only few works have focused on sub-patterns in the weighted case [3, 4]. In this work we pro- pose a method [5] based on few ingredients: unbalanced weighted networks, a sink node, a random walker with a limited number of steps. The sink node compensates the excess of in-going flows balancing the network. It allows to highlight the role of weights heterogeneity in shaping the network structural organization. The approach gives the frequency of paths of different lengths observable within a fixed number of steps of the random walker. We fix a maximum of three steps to offer both analytical and computation results, but it can be generalised to any number of steps. The method can be applied to any kind of network and allows comparisons independently on their sizes. We found that networks belonging to the same field show very similar motifs profiles and focusing on specific motifs we can shed light on the underlying functioning mechanism. Specifically, we applied our approach to the energy system of different European countries for the last 30 years. They can be represented as a tripartite network of interconnected producing, transforming and consuming activities of carriers linked through energy flows for a total of 870 networks [6]. We use the aforementioned motif definition to explore the complexity of such networks performing over time and across

countries comparisons with focus on specific final sectors and renewable/not renewable energy carriers. We show the complexification of the energy systems over time and shed light on the under- lying evolutionary processes of given sectors. While diversification is found at different degrees and through diverse sub-patterns for different final sectors, the analyses confirm the persistence of silo structures for a few end-use sectors, the hard-to-decarbonize ones.

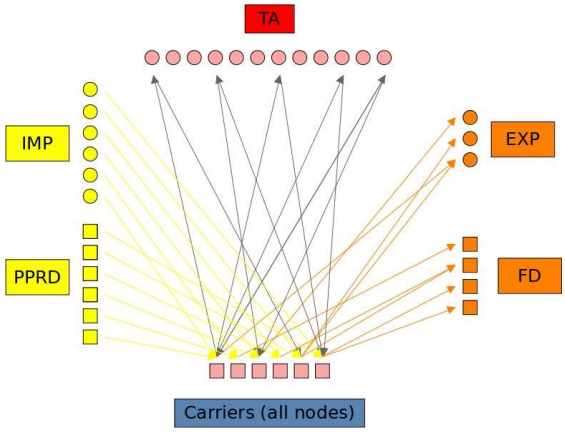


Figure 1: **Tripartite representation of the energy network.** We have on the left the source of carriers (on the bottom as pink squares): internal pro- duction (PPRD) or imports (IMP); on the top (pink circles) the transformation activities (TA) of carriers; on the right the sink group: final demand (FD) and exports (EXP).

[1] M. Newman, Networks. Oxford university press, 2018.
[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: simple building blocks of complex networks,” Science, vol. 298, no. 5594, pp. 824–827, 2002.
[3] J.-P. Onnela, J. Saramäki, J. Kertéz, and K. Kaski, “Intensity and coherence of motifs in weighted complex networks,” Physical Review E, vol. 71, no. 6, p. 065103, 2005.
[4] T. LaRock, I. Scholtes, and T. Eliassi-Rad, “Sequential motifs in observed walks,” Journal of Complex Networks, vol. 10, no. 5, p. cnac036, 2022.
[5] F. Picciolo, F. Ruzzenenti, P. Holme, and R. Mastrandrea, “Weighted net- work motifs as random walk patterns,” New Journal of Physics, vol. 24, no. 5, p. 053056, 2022.
[6] A. Diaz, S. Tchung-Ming, A. Diaz Vazquez, N. A. Matei, E. Moreno Cruz, and F. Fosse, “Understanding sectoral integration in energy systems through complex network analysis,” in International Conference on Complex Net- works and Their Applications, pp. 481–493, Springer, 2022.

11:50 – 12:10 **Yoshiharo Maeno** *(Meiji University, School of Interdisciplinary Mathematical Sciences, Tokyo Japan)*

Geographic profiling for socio-economic diffusion phenomena

Socio-economic diffusion phenomena [1], [2] include such examples as spreading of an infectious disease, transmission of financial distress in a market, and dissemination of false information on social media. These phenomena are described by a reaction-diffusion process mathematically, that is a complex spatial-temporal stochastic process. Discovering the source event of diffusion in a geo- space, or solving an inverse problem for the initial state, is called geographic profiling. The problem is often too complex to be tractable with conventional statistical inference.
This work studies machine-learning-based inverse problem solvers for such a reaction diffusion process [3]. The focus is on the performance of a state-of-the-art convolutional neural network (CNN). This solver is compared with three solvers: native Bayes classifier (NBC), random forest classifier (RFC), and multinomial logistic regressor (MLR). The performance is investigated with synthetic datasets for standard SIR

epidemiological compartments on a square grid 2D geo-space (N = 32 × 32 sub-populations on the grids). The convolutional neural network works effectively in discovering a source sub-population as an index patient. It achieves the largest time average of accuracy (A ≥ 0.85) for growing infection (reproductive ratio: R = 1.2) in a heterogeneous geo-space. The hit score remains near the lower bound over time (H ≤ 1.5 × 10−3~N−1). It is also suggested that discovering multiple sources is feasible by learning the dataset for a source.
The performance of the solver is demonstrated with a historical dataset for a cholera outbreak in London in 1854. The applicability of the solver to analyzing similar public health issues, financial crises, and circulating misinformation is discussed as a work-in-progress research topic**.**

[1] Y. Maeno, Transient fluctuation of the prosperity of firms in a network economy, Physica A: Statistical Mechanics and its Applications 392:3351 (2013).
[2] Y. Maeno, Detecting a trend change in cross-border epidemic transmission, Physica A: Statistical Mechanics and its Applications 457:73 (2016).
[3] Y. Maeno, Inverse problem solver for epidemiological geographic profiling, Evolutionary and Institutional Economics Review https://doi.org/10.1007/s40844-024-00281-3 (2024)

12:10 – 12:30 **Isaak Mengesha** *(University of Amsterdam, Netherlands)*

Spatial Analysis of Economic Development

Remote sensing, such as observing night-time lights (NTL), has unlocked the analysis of global economic development on high spatial resolutions and increased independence from nation- ally reported data [1, 2]. Despite ample work on distributive properties and local dynamics of eco- nomic development, their global geo-temporal dynamics remains poorly understood. The un- equal distribution of economic development has been noted consistently across regions and time [3, 4]. As the development of India shows, urbanization is an important process for development and regional distribution of the merits of growth do not percolate evenly through the different sub-regions [5]. The driving mechanisms for this observation are being studied under the framework of evolutionary economic geography and complexity economics [6, 7]. Contrary to mainstream neoclassical economists they emphasize the role of clustering, formation of networks, technological evolution and spatial embedding. We expand on the approach developed by Kushwaha et al. to detect chains of activities in economic development by adding a more robust Monte Carlo approach to the spatial gridding and sensitivity analysis to threshold values of event categorization [8]. Our objective is to quantify the relevant scales at which these dynamics unfold by identifying chains of economic development and influencing conditions for them to emerge. We draw from methods to under- stand self-organized criticality as well as analyse the correlation matrix of the cells, mutual in- formation between cells, and extract chains of dependency [9, 10]. Our preliminary analysis showed that below the time scale of a year, fluctuations are too large and aggregation effects too small to meaningfully detect signals of subnational GDP changes via NTL. Therefore our sub- sequent analysis utilizes the yearly spatially gridded harmonized GDP and electricity consumption estimates by Chen et al. [11]. Similar to Kushwaha et al. we identify the relevant scales for chains of economic growth to emerge and relate to established spatial distributions of economic activity such as Zipf's law [12]. We recognize the modifiable areal unit problem as central to the validity of our spatial analysis [13]. Consequently, we design extensive robustness checks for our inference of growth chains via Monte Carlo approaches to griding [14]. We then localize such interactive chains and analyse their properties. Of particular interest is the spatial embedding of networks of economic dependency. Furthermore, we investigate the necessary conditions of these interactive chains to form by leveraging dataset on the existence of critical infrastructure and the prevalence of deprivation [15, 16]. Our work provides the basis to quantify the potential influence of violent conflict (or disasters) on economic development [17, 18, 19]. For this we utilize the ACLED dataset [20]

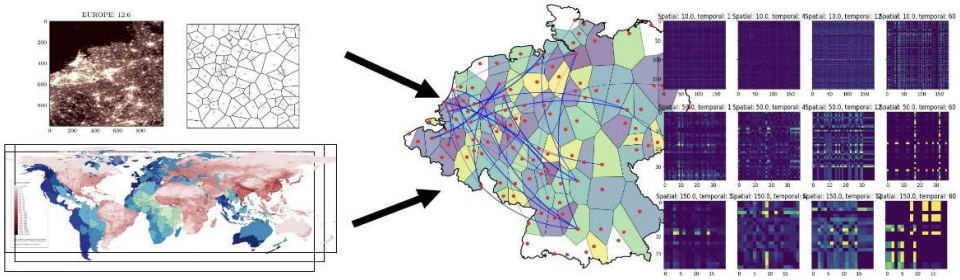


Figure 1: Aggregating of various data sources (GDP, NTL, Deprivation, Critical Infrastructure, etc) in spatial grid via voronoi-tesselation. The result are geo-spatially located time series from which we can extract network information (source: [21, 22])

[1] Stephen J Redding, "The empirics of new economic geography," Journal of regional science, vol. 50, no. 1, pp. 297–311, 2010.

[2] J Vernon Henderson, Adam Storeygard, and David N Weil, "Measuring economic growth from outer space," American economic review, vol. 102, no. 2, pp. 994–1028, 2012.

[3] Mohamed Ali Labidi, "Spatial distribution of economic activities, growth policies and externalities in developing countries," Journal of economics and sustainable development, vol. 9, pp. 10–16, 2018.

[4] Yi Jiang, "Asian cities: spatial dynamics and driving forces," *The Annals of Regional Science*, vol. 66, no. 3, pp. 609–654, 2021.

[5] Jonathan Colmer, "Urbanisation, growth, and development: Evidence from india," 2015.

[6] Ron Boschma and Koen Frenken, "Evolutionary economic geography," *The new Oxford handbook of economic geography*, pp. 213–229, 2018.

[7] W Brian Arthur, "Foundations of complexity economics," *Nature Reviews Physics*, vol. 3, no. 2, pp. 136–145, 2021.

[8] Niraj Kushwaha and Edward D Lee, "Discovering the mesoscale for chains of conflict," *PNAS nexus*, vol. 2, no. 7, pp. pgad228, 2023.

[9] Henrik Jeldtoft Jensen, *Self-organized criticality: emergent complex behavior in physical and biological systems*, vol. 10, Cambridge university press, 1998.

[10] Claudio Castellano, Santo Fortunato, and Vittorio Loreto, "Statistical physics of social dynamics," *Reviews of modern physics*, vol. 81, no. 2, pp. 591, 2009.

[11] Jiandong Chen, Ming Gao, Shulei Cheng, Wenxuan Hou, Malin Song, Xin Liu, and Yu Liu, "Global 1 km× 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data," Scientific Data, vol. 9, no. 1, pp. 202, 2022.

[12] Kwok Tong Soo, "Zipf's law for cities: a cross-country investigation," Regional science and urban Economics, vol. 35, no. 3, pp. 239–263, 2005.

[13] A Stewart Fotheringham and David WS Wong, "The modifiable areal unit problem in multivariate statistical analysis," Environment and planning A, vol. 23, no. 7, pp. 1025–1044, 1991.

[14] Sebastian Schutte and Claire Kelling, "A monte carlo analysis of false inference in spatial conflict event studies," PLoS one, vol. 17, no. 4, pp. e0266010, 2022.

[15] Center for International Earth Science Information Network - CIESIN - Columbia University, "Global gridded relative deprivation index (grdi), version 1," Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC), 2022, Accessed DAY MONTH YEAR.

[16] Sadhana Nirandjan, Elco E Koks, Philip J Ward, and Jeroen CJH Aerts, "A spatially-explicit harmonized global dataset of critical infrastructure," Scientific Data, vol. 9, no. 1, pp. 150, 2022.

[17] Michelle R Garfinkel and Stergios Skaperdas, "Economics of conflict: An overview," Handbook of defense economics, vol. 2, pp. 649–709, 2007.

[18] Christopher Blattman and Edward Miguel, "Civil war," Journal of Economic literature, vol. 48, no. 1, pp. 3–57, 2010.

[19] H˚vard Hegre, Gudrun Østby, and Clionadh Raleigh, “Poverty and civil war events: A disaggregated study of liberia,” Journal of Conflict Resolution, vol. 53, no. 4, pp. 598–623, 2009.
[20] Clionadh Raleigh, Rew Linke, H˚vard Hegre, and Joakim Karlsen, “Introducing acled: An armed conflict location and event dataset,” Journal of peace research, vol. 47, no. 5, pp. 651–660, 2010.
[21] United Nations, World Population Prospects, “World population prospects,” https:// ourworldindata.org/life-expectancy, 2022, Accessed: 2024-04-26. Human Mortality Database, “Human mortality database,” https://ourworldindata.org/ life-expectancy, 2023, Accessed: 2024-04-26.

12:30 – 12:50 **Kiyoshi Kanazawa** *(Department of Physics at Kyoto University, Japan)*

Is the square-root law universal for price impacts?: empirical validation based on microscopic datasets

The price impact is the average price movement due to market-order submissions. For small amounts of market orders, the price impact is empirically proportional to Q, where Q is the total volume of the market orders. On the other hand, for large amounts of market orders, the price impact is empirically proportional to Q^{δ} with $\delta \approx 0.5$. This empirical law is called the square-root law and is believed to be robust for various markets.

Here we have a natural question: Does this law belong to the universal power law? In statistical physics, the universal power law is defined as the power law whose characteristic exponent is independent of the system's microscopic details; conversely, the non-universal power law is defined as the power law whose exponent is dependent on the system's microscopic parameter. There has been a debate about this question in econophysics: some theories support the universality of the square-root law, but others do not. It has not been easy to settle this debate because microscopic datasets – to track trading behaviour at the level of individual traders – are necessary to support/reject either hypothesis.

In this talk, we will provide the current best evidence about the universality of the square-root price-impact law. We have access to the microscopic datasets of the Tokyo Stock Exchange market – the biggest stock market in Japan – and studied the price impact law for large metaorder at the level of traders. We extract the empirical distribution of the power-law exponent – δ – and find that the exponent distribution has a sharp peak around 1/2. In addition, we will report thorough tests of several theoretical models claiming the non-universality of the power-law exponents.

12:50 – 13:10 **Riccardo Milocco** *(IMT - School for Advanced Studies, Lucca, Italy)*

Multi-Scale Node Embeddings for Networks

Complex networks represents several societally-relevant processes, ranging from economic flows to the response of ecosystems to climate change. More importantly, these interacting phenomena are coexisting at multiple scales where the network properties may strongly vary. Consequently, the outcomes of any model defined at a single scale become unreliable if derived at a extrapolated level. This limitation highlights the need for a “scale-invariant” approach to abstract from the chosen level.

A node embedding is the assignment of a vector to each node to ultimately encode its role in the network. Here, we will focus on the problem of node embedding in a multi-scale setting where the finest observable network is binary and undirected. As sketched in Figure 1a, one can attach a vector to each node of both the microscopic (observed) network and the “coarse-grained” one. The latter is obtained by means of a pre-assigned node partition of the observed network, e.g. merging nodes into a desired set of macro nodes. This procedure could be iterated up to obtaining a fully-connected network. That said, the research question is how the microscopic embedding vectors relate with the coarse-grained ones.

Typically, node-embedding methods, e.g. the LogisticPCA [1] or the Configuration Model [2], are designed to fit the network at the given resolution level, thus, assuming that the “accessible scale” is the fundamental one on which all the

outcomes should depend. We will refer to this type of model as "single-scale" models (SSMs).
To overcome this limitation, we will use the "scale-invariant" model [3] to tackle the node-embedding problem. Moreover, we will enhance it with node-vectors (SIM) similarly to the SSMs. The crucial difference with respect to SSMs is that the SIM relies neither on the fixed partition nor on the constitutive scale: it could be either coarse-grained, mimicking the "forward" unfolding of the observed network, but also fine-grained in the "reversed" direction, to disaggregate each macro node into its microscopic constituents. For the "forward" direction, the SIM prescribes to *sum* the microscopic node-vectors in order to obtain the macro-node embedding (see Figure 1a). This rule leads to a solid interpretation of the "sum" of node-vectors, whereas this operation is barely addressed by the node-embedding literature.

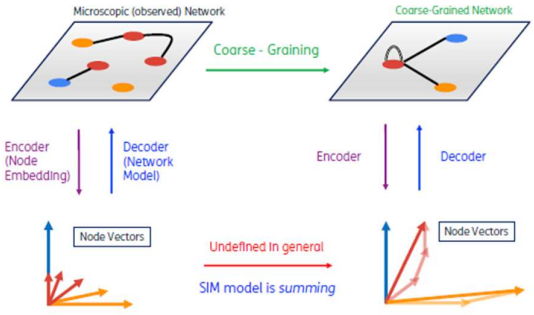
In addition, it has been recently claimed that a low-dimensional embedding is not capable of producing the observed amount of triangles in real networks. Contrarily we will show that this is possible both at a single-scale and at higher resolution scales, by analyzing the local binary clustering coefficient (BCC) as shown in Figure 1b. More precisely, the local BCC is defined as the ratio between the number of triangles and the number of wedges for a given node.

Our evaluation will proceed by iteratively coarse-graining a real-world undirected transaction network to obtain a multi-scale setting wherein to compare the SSMs and the SIM. In Figure 1b, we have selected level 0 to make a fixed-scale comparison between models, and level 2 to compare models across scales. In the main panels we show the expected BCC (by the LPCA or the SIM) against the observed BCC, while in the insets we show the probability matrix of the "coarse-grained" model against the "re-fitted" model, i.e. the one fitted again at the new scale. The SSMs outperforms the SIM at the fixed-scale level 0, while it is the other way around when the models are compared at the coarse-grained level 2. Parsimoniously, the SIM is found to exhibit small discrepancies both at single and multiple scales. This behaviour is reproduced also for the degree and the average nearest-neighbor degree (not shown). Finally, we compared the AUC-ROC and Precision-Recall curves across different levels by treating the coarser levels as the test sets. This comparison highlighted the inconsistency among the SSMs vectors at different scales and the improvement achieved by the SIM due to its multi-scale nature.

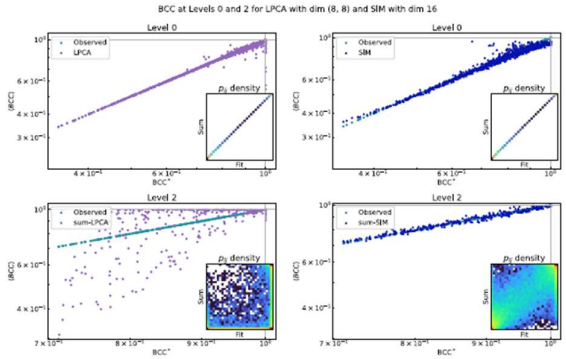
[1] S. Chanpuriya et Al., An Interpretable Graph Generative Model with Heterophily, ArXiv, 2111.03030 [PrePrint] (2021)

[2] J. Park et Al., Origin of degree correlations in the Internet and other networks, Phys. Rev. E, Vol. 68, 026112 (2004)

[3] E. Garuccio et Al., Multiscale network renormalization: Scale-invariance without geometry, Phys. Rev. Res., Vol. 5, 043101 (2023)



(a) Multi-Scale Node-Embedding Sketch



(b) Multi-scale Comparison of LPCA and maxlSIM

13:10 – 13:30 **Xavier Brouty** *(ESILV, La Défense, France)*

A new Econophysics approach for Bid-Ask Spreads Estimation in the presence of serial dependence

Serial dependence biases traditional bid-ask spread estimators, with a stronger effect for negative autocorrelation. By making new assumptions about the underlying dynamics of financial prices and bid-ask spreads in presence of serial dependence, we derive four new estimators of the bid-ask spread using only high frequency prices. We also derive an estimation of the Hurst exponent in the presence of microstructure noise. We show the consistency of the estimators and apply the methods on simulated and real high frequency data.
The simulations are made with fractional Brownian motion processes with varying Hurst exponent and Ornstein-Uhlenbeck processes, allowing us to create several time series with different dynamics.
By simulating time series with infrequent trading orders, we show that our estimators are unbiased when trading is infrequent. We also deal with a new estimation for the negative estimates allowing us to reduce the variance of our estimators.

13:30 – 14:30 **Lunch**

14:30 – 16:30 **Session 3:** Chair **Siew Ann Cheong** *(Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore)*

14:30 – 14:50 **Irena Vodenska** *(Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Republic of North Macedonia; Administrative Sciences Department, Financial Management, Metropolitan College, Boston University, USA)*

Forecasting stock price movements using ESG risk ratings and sentiment analysis

In the modern financial market landscape, volatility has become a defining characteristic, shaping investment strategies and risk management practices. This volatility is not merely a consequence of economic fluctuations but is deeply intertwined with the dynamic interactions between various stakeholders within the financial ecosystem. As we embark on a new era characterized by unprecedented technological advancements and interconnected global markets, the dynamics of volatility are further compounded by emerging trends, such as the Environmental, Social, and Governance (ESG) movements. Moreover, the pervasive influence of news media, with its instantaneous dissemination of information and market sentiment, exerts a palpable

impact on stock prices and market behavior, underscoring the intricate interplay between external factors and market volatility.

In this study, we delve into the intricate landscape of stock markets, scrutinizing the underlying dynamics amidst the backdrop of the evolving global financial ecosystem. Our study expands the scope of conventional analyses as we delve into the interplay between stock market fluctuations and Environmental, Social, and Governance (ESG) risk ratings alongside news and media sentiments. We utilize structured and unstructured data, consisting of daily stock market closing prices, ESG risk ratings, and news items from GDELT, Google News, Reddit, and Twitter. For our analysis, we select the timeframe from March 1, 2019, to March 1, 2023.
We focus our analysis on a subset of company stocks representing diverse sectors to capture a comprehensive understanding of market trends, more specifically, we look at the following companies: Amazon, Apple, Exxon, Nike, Pfizer, and Tesla. Our methodology encompasses a comprehensive approach to forecasting stock price movements. We integrate signals derived from logarithmic returns and sentiments extracted from selected news and media sources, as well as utilize SMA and WMA measures alongside the Relative Strength Index of stock prices. We feed these inputs into an XGBoost model, enabling us to predict the direction of stock price movements. Following our predictive modeling, we employ explainability, i.e., the SHAP (SHapley Additive exPlanations) model, to gain deeper insights into the relationship between stock price movement direction, sentiments, and ESG (Environmental, Social, and Governance) risk ratings.
Our analysis reveals that the logarithmic return from the previous day exerts the most significant influence on predictions, followed closely by sentiment analysis. Interestingly, according to the SHAP analysis, ESG risk ratings have a negligible impact on the stock price forecasting model. These intriguing results prompt further investigation, driving our commitment to expanding our research scope with larger datasets to validate or challenge our findings. Moreover, our methodology serves as a foundational framework for future researchers, emphasizing the significance of continued exploration into financial markets to enhance our understanding and predictive capabilities.

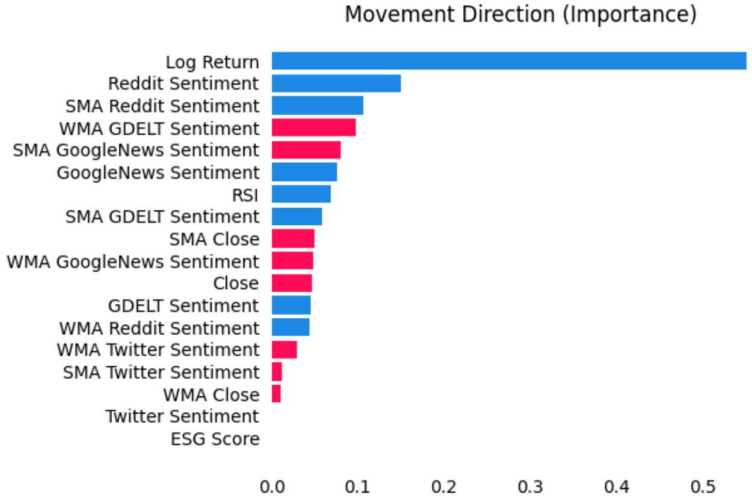


Figure 1: The importance of each feature when forecasting the movement direction of the stock price of the Exxon company (blue color denotes positive correlation, whereas red color denotes negative correlation)

14:50 – 15:10 **Peter Yen** (*School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*)

Scale-Dependent Inverse Temperature Features Associated with Market Crashes

For extreme events like market crashes we identified the scale-dependent inverse temperature β features. We collected daily returns of component stocks of S&P 500,

and Nikkei 225 from 1 Jan 2007 to 30 April 2023. Dynamically, we discovered prominent features in β peaks. These can be esoteric, or with a universal identifiable exponent. We examine five esoteric peaks and seven characteristic peaks, and identify a universal exponent $b^{\pm}$narrowly distributed between 0.3−0.4 in both markets. Following this, we constructed heatmaps of β to classify different types of crashes. For example, the 2007−2009 Global Financial Crisis is an endogenous crash in the US market, which caused an exogenous crash in the Japan stock market. We then delve deeper into the dynamics of β by examining heatmaps of ΔI (change of number of links), and ΔQ (change of number of triangles). We found oscillations that became very intense across all scales when market-level events occur.

15:10 – 15:30 **Luigi Riso** *(Department of Economic Policy, Università Cattolica del Sacro Cuore, Milan, Italy)*

A novel Sparse Graphical Model for optimizing portfolio assets allocation

Establishing the most pertinent dependencies among variables within a given dataset is a pivotal aspect of data science. This is crucial not only for extracting a greater wealth of information from the variables but also for constructing effective models capable of predicting and explaining the phenomena of interest. In this context, Graphical Models (GM) theory, integrating probability with graph theory, provides suitable tools for representing the dependencies among a set of variables through a graph. In broad terms, GMs constitute a family of probability distributions defined within the framework of a directed or undirected graph. The nodes within the graph correspond to random variables, and joint probability distributions are established by computing products over functions defined on connected subsets of nodes [Lauritzen and Wermuth, 1989]. As stated by Jordan et al. [2004], the GM formalism remains agnostic to the distinction between frequentist and Bayesian statistics. Two primary branches of graphical representations for distributions are commonly used within GM, namely undirected graphical models and Bayesian networks. Even if they both embody the principles of factorization and independence, they differ in terms of graphical representation, the specific independences they can encode, and the factorization of the distribution they induce [Koller and Friedman, 2009]. In essence, GMs aim to represent a multivariate probability distribution that effectively characterizes a given dataset through either an undirected graph or a directed acyclic graph (DAG) [Lauritzen and Wermuth, 1989]. The challenge arises significantly when dealing with a large number, say p, of variables. In such multivariate scenarios, the model must account for a minimum of $O(p^2)$ interrelations among the variables, resulting in a quadratic scaling of model parameters with the number of variables. Adopting a parsimonious approach necessitates identifying a model that accurately captures the statistical properties of the observations, while minimizing the number of parameters [Barfuss et al., 2016]. In this context, a solution to tackle this challenge is found in a specific undirected GM known as the graphical lasso or GLASSO [Friedman et al., 2008]. To introduce the newly proposed sparse graph, termed the Best Path Algorithm for Sparse Graphical Models (BPASGM), let us briefly outline the rationale behind the BPA as suggested by Riso et al. [2023]. Consider a dataset $\boldsymbol{X}$ comprising n observations and p^1 variables $X_1, X_2, \ldots, X_{p-1}, Y$. The BPA initially constructs a graphical model [see, e.g., Edwards et al., 2010, Riso and Guerzoni, 2022] representing the relationships among the variables in the dataset. Subsequently, for a specific variable of interest X_i, the BPA identifies the most appropriate subset of variables. Given the best optimal set $\boldsymbol{X}wx_i$ of predictors for Xi detected via the BPA, the following relationship holds:

$$X_i \approx \phi(\boldsymbol{X}_{wx_i})$$

where $\phi(\cdot)$ is a generic function belonging to the family of generalized linear models (GLMs) [Eshima and Tabata, 2007]. The BPA algorithm can be iteratively applied to each variable of the dataset to construct a sparse graphical model (SGM). As shown below, this approach proves highly effective as it addresses three critical issues associated with the GLASSO method: the problem of the correlation matrix's invertibility, the selection of an appropriate regularization parameter λ, and ultimately, its ability to detect only linear dependencies

between variables. where $\phi(\cdot)$ is a generic function belonging to the family of generalized linear models (GLMs) [Eshima and Tabata, 2007]. The BPA algorithm can be iteratively applied to each variable of the dataset to construct a sparse graphical model (SGM). As shown below, this approach proves highly effective as it addresses three critical issues associated with the GLASSO method: the problem of the correlation matrix's invertibility, the selection of an appropriate regularization parameter λ, and ultimately, its ability to detect only linear dependencies between variables.

W. Barfuss, G. P. Massara, T. Di Matteo, and T. Aste. Parsimonious modeling with information filtering networks. Physical Review E, 94(6):062306, 2016.

D. Edwards, G. C. De Abreu, and R. Labouriau. Selecting high-dimensional mixed graphical models using minimal aic or bic forests. BMC bioinformatics, 11(1):18, 2010.

N. Eshima and M. Tabata. Entropy correlation coefficient for measuring predictive power of generalized linear models. Statistics & probability letters, 77(6):588–593, 2007.

J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3):432–441, 2008.

M. I. Jordan et al. Graphical models. Statistical science, 19(1):140–155, 2004.

D. Koller and N. Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.

S. L. Lauritzen and N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. The annals of Statistics, pages 31–57, 1989.

L. Riso and M. Guerzoni. Concept drift estimation with graphical models. Information Sciences, 606:786–804, 2022.

L. Riso, M. G. Zoia, and C. R. Nava. Feature selection based on the best-path algorithm in high dimensional graphical models. Information Sciences, 649:119601, 2023.

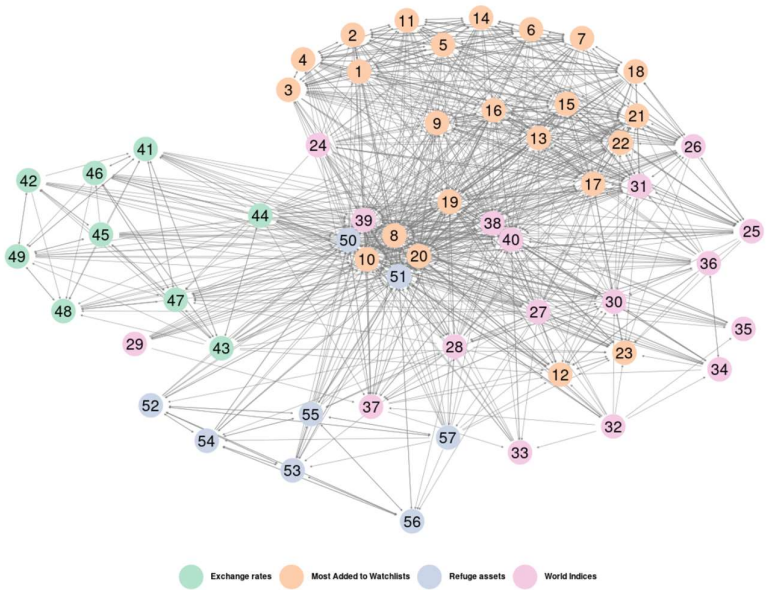


Figure 1: application of BPASGM at 57 assets, sampled perdiod 01-01-2007,31-21-2023

15:30 – 15:50 **Axel Araneda** *(Institute of Financial Complex Systems, Department of Finance, Masaryk University, Brno, Czechia)*

Price modelling under scaled Brownian motion

Standard diffusion processes have become the standard tool for price modeling in a continuous time setting, where the Black-Scholes model appears as the flagship of the modern finance theory. In this talk, we study the potential applications of alternative stochastic processes featured by anomalous diffusion, in particular the use of the so-called Scaled Brownian motion (SBM), a centered Gaussian process featured by the Markov and martingale property, subordinated to a standard Brownian motion via the non-linear deterministic time change $t \to t_* = t^\gamma$, with $0 < \gamma < 2$.
On one hand, we analyze the use of the SBM as the driving process for the log-return process (i.e., Black-Scholes-like type), replacing, in practice, the constant volatility of the Geometric Brownian motion by a time-varying power-law structure, leading to a better description of the implied volatility term structure, validating this fact through its ability to reproduce actual European ATM options data.
On the other hand, we consider the volatility of the price fluctuations to be a power law dependent on both price and time. This is achieved by replacing the standard Brownian with an SBM in the well-known constant elasticity of variance (CEV) model, which is suitable for default derivatives uses because the origin is an attainable and absorbing boundary via diffusion to zero. We derive both the related transition density and the first-passage time over zero-state (default) probability, linking them to the results for the square Bessel processes, computing both the option pricing and Credit Default Swaps (CDS) spreads.
As a final approach, we consider an immediate generalization for the SBM through a time-dependent scaling parameter, that allows the models to price differently under short and long maturities, leading to a better fit with actual option market quotes. Moreover, in the case of the CEV structure, it solves one of the reported weaknesses of the standard CEV, producing in a consistent way, higher default probabilities and coupon rates without appealing to a jump-killing process (jump-to-default).

15:50 – 16:30 **Coffee break & Poster presentations**

16:30 – 18:00 **Panel "Glorious Future":** Chair **Stefan Thurner** (*Complexity Science Hub, Medical University of Vienna, Austria*)

19:00 **Dinner at Heuriger Schübl-Auer**

Wednesday, June 5 2024

9:00 – 10:00 **Session 1:** Chair **Alessandro Pluchino** *(Department of Physics and Astronomy "E.Majorana" of the University of Catania)*

9:00 – 9:20 **Kimmo Kaski** *(Aalto University School of Science, Finland)*

Social Physics: Data-Driven Studies of Human Social Networks

Social Physics of today focuses on studying large-scale socially and societally relevant datasets using Network Theory based data-analytics and computational modelling, to get insight into structures of human social networks and behavioral processes therein. This has - among a few other studies -been demonstrated by our analysis of mobile phone datasets, where we find the social network having modular structure of communities with strong internal and weak external ties. As the data includes the phone users' demographics, i.e. gender and age, we have been able to investigate the nature of social interactions from the egocentric viewpoint to get insight into the gender and age-related social behaviour patterns and dynamics of human relationships, across their lifespan. Moreover, fusing openly accessible data from country statistics and geophysical records with the mobile phone data, we have been able to decipher the daily and seasonal activity patterns of individuals and the society. In these studies, computational modelling plays a key role in finding plausible mechanisms for the formation of and dynamics in social networks.

9:20 – 9:40 **Siew Ann Cheong** *(Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore)*

Timing Topological and Geometrical Changes in Stock Market Cross Correlations

Econophysicists investigating structural changes in a stock market frequently start from the cross correlations of its stocks. Signatures of such changes can be discerned from information filtering (e.g., minimal spanning tree [1] and planar maximally filtered graph [2]) of the cross correlations, or by clustering stocks (using K-means [3] or hierarchical clustering [1]) based on their cross correlations. By comparing clusters obtained from sliding time windows, we realized that these are not stable, but frequently merge with each other, or break up into smaller clusters [4]. Such an analysis is challenging, because clusters within one time window become unrecognizable in the next, making it difficult to work out the rules governing their recombinations. This led us to try out tools from persistent homology in topological data analysis (TDA), which should in principle allow us to identify not only clusters, but also other persistent hierarchical structures [5]. Our hope is that by following the time evolution of the Ricci curvatures of 'necks' connecting these structures, we can figure out how they change with time [6]. Unfortunately, some important changes are still too rapid to follow. More recently, we found that the spectra of the Laplacians of correlation networks at different scales become persistently gapped during market crashes [7]. Using this highly sensitive spectral gap indicator, we developed a method to cross filter information from the left and right ends of a sliding time window to very accurately time turbulent periods in stock markets. Eventually, this inspired us to test information filtering using a pair of adjoint time windows, which has the structure of a derivative. However, unlike the usual differentiation, the method allows us to determine the timing of a sudden change very precisely, even when the time series data is very noisy. After applying this method to the mean and variance of a financial time series, we illustrate

how the method can be generalized to obtain point estimates of cross correlations and entropies.

[1] Mantegna, R. N. (1999). Hierarchical structure in financial markets. The European Physical Journal B-Condensed Matter and Complex Systems, 11, 193-197.
[2] Tumminello, M., Aste, T., Di Matteo, T., & Mantegna, R. N. (2005). A tool for filtering information in complex systems. Proceedings of the National Academy of Sciences, 102(30), 10421-10426.
[3] Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. Expert Systems with Applications, 37(12), 8793- 8798.
[4] Teh, B. K., & Cheong, S. A. (2015). Cluster fusion-fission dynamics in the Singapore stock exchange. The European Physical Journal B-Condensed Matter and Complex Systems, 88, 1-14.
[5] Yen, P. T. W., & Cheong, S. A. (2021). Using topological data analysis (TDA) and persistent homology to analyze the stock markets in Singapore and Taiwan. Frontiers in Physics, 9, 572216.
[6] Yen, P. T. W., Xia, K., & Cheong, S. A. (2021). Understanding changes in the topology and geometry of financial market correlations during a market crash. Entropy, 23(9), 1211.
[7] Yen, P. T. W., Xia, K., & Cheong, S. A. (2023). Laplacian spectra of persistent structures in Taiwan, Singapore, and US stock markets. Entropy, 25(6), 846.

9:40 – 10:00 **Victor Le Coz** (*Quant AI lab, Paris; Ecole polytechnique, Paris, France)*

*Revisiting Elastic String Models of Forward Interest Rates (*https://arxiv.org/abs/2403.18126)

Twenty five years ago, several authors proposed to model the Forward interest Rate Curve (FRC) as an elastic string along which idiosyncratic shocks propagate, accounting for the peculiar structure of the return correlation across different maturities. In this paper, we revisit the specific “stiff” elastic string field theory of Baaquie and Bouchaud (2004) in a way that makes its micro-foundation more transparent. Our model can be interpreted as capturing the effect of market forces that set the rates of nearby tenors in a self-referential fashion.
The model is parsimonious and accurately reproduces the whole correlation structure of the FRC over the time period 1994 − 2023, with an error below 2%.We need only two parameters, the values of which being very stable except perhaps during the Quantitative Easing period 2009 − 2014. The dependence of correlation on time resolution (also called the Epps effect) is also faithfully reproduced within the model and leads to a cross-tenor information propagation time of ≈ 10 minutes. Finally, we confirm that the perceived time in interest rate markets is a strongly sub-linear function of real time, as surmised by Baaquie and Bouchaud (2004). In fact, our results are fully compatible with hyperbolic discounting, in line with the recent behavioural Finance literature (Farmer and Geanakoplos, 2009).

10:00 – 10:20 **Giulio Cimini** *(Physics Department and INFN, University ‘Tor Vergata’, Rome, Italy; ‘Enrico Fermi’ Research Center (CREF), Rome, Italy)*

The causal role of the Reddit collective action on the GameStop short-squeeze

In early 2021, the stock price of GameStop (GME) experienced a dramatic increase, triggered by a short squeeze operation that has been largely attributed to Reddit’s retail investors. There is ample evidence of the connection between financial markets and social media platforms. User activity on platforms such as Twitter or Google Trends not only mirrors the current state of financial markets but can at times anticipate the evolution of stock prices. This predictive power is primarily attributed to users sharing and discussing financial information online. However, the use of online social networks extends

beyond information consumption: they serve as platforms for users to coordinate and catalyze global phenomena, akin to the Arab Spring or on-the-ground protests. The GME short squeeze showcased, for the first time, the potential of online social networks to catalyze financial collective action. How, when and to what extent Reddit users played a key role in driving up these prices, however, remains unclear.

In this preprint we characterize the discourse surrounding the GME event using the volume of GME-related activity on Reddit and on Twitter, the former reflecting the conversation within the community of users and the latter capturing the broader public attention to the stock markets. We employ causal inference techniques with high temporal resolution to elucidate the interplay between social media activity and stock markets during the period surrounding the GME squeeze. We find that Reddit discussions foreshadowed trading volume before the GME short squeeze, with their predictive power being particularly strong on hourly time scales (See Figure 1 **a**). This effect emerged abruptly and became prominent a few weeks before the event, but waned once the community of retail investors gained widespread visibility through Twitter. Discussing an asset does not equate to ownership or acquisition. Reddit users provide evidence of their financial investments by posting screenshots of their accounts taken from trading platforms, which attest their positions and movements. We estimate the financial positions of Reddit users on GME using computer vision techniques and by summing the positions across all users, we obtain the collective investment. As the causal link unfolded, the collective investment of the Reddit community closely mirrored the market capitalization of the stock, representing 1% of the total market capitalization of GME (See Figure 1 **b**). Overall, our results shed light on the dynamics behind the first large-scale financial collective action driven by social media users.

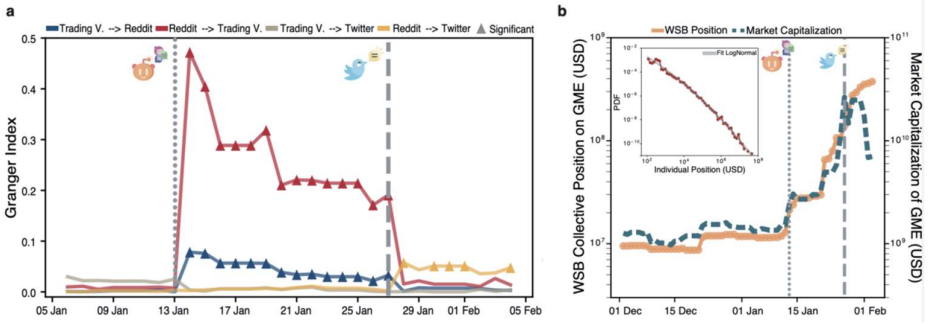


FIG. 1: a) Granger Index capturing the predictive power of a signal on another (with a lag of 1 hour) for the following pairs: Trading Volume-to-Reddit (blue), Reddit-to-Trading Volume (red), Trading Volume-to-Twitter (gray) and Reddit-to-Twitter (yellow). Each point is computed considering time series spanning the 15 preceding days. Triangles correspond to statistical significant values (p-values < 0.1).
The Granger Index assesses whether the inclusion of past values of a variable improves the prediction of another variable in a vector autoregressive model of lag l, defined as the ratio of the error sum of squares in the restricted model to that in the full model. The vertical gray dotted line indicates the beginning of the Granger causality link (13 January 2021), whereas the vertical gray dashed line corresponds to the Tweet by Elon Musk (27 January 2021) that brought the squeeze to the public attention. b) Daily value of the WSB collective position on GME (orange, y-right axis) and GME Market Capitalization (blue, y-left axis). Inset: Distribution of values of users' position on GME.

10:20 – 10:40 **Leonardo Ialongo** (*Complexity Science Hub, Vienna, Austria)*

Economic shocks in reconstructed firm-to-firm networks from input-output data

The recent global pandemic and geopolitical tensions have increased the attention on supply chains and their vulnerability to disruption. As more data on national and international firm-to-firm networks has become available, our understanding of their

structure and economic importance has improved, but remains far from complete. Unfortunately, since such data are generally proprietary and difficult to access, input-output analysis are still mainly conducted at an aggregate level, which can lead to severe biases and underestimation of risks [1, 2]. Inspired by this need, we devise a methodology to generate realistic firm-to-firm networks which are consistent with available public information. More specifically, we constrain the size distribution of firms and the density of the graph while maintaining consistency with input-output tables which report the economic exchanges between the different sectors a country's economy. In order to do so we use a novel methodology based on a model which is invariant to arbitrary node aggregations [3] to incorporate the information available on sector linkages and define a fitness-model that embodies the knowledge of the sector- wise in-strength per firm and the empirical link density. To the generative network model we couple an economic one which has found wide applicability over the last years to study the impact of economic shocks such as the one due to COVID-19 and natural disasters [1, 4]. Our contribution is thus twofold: on the one hand, we provide a new methodology which allows to generate realistic firm-to-firm network, consistent to the aggregate known data; on the other hand, we investigate how the micro-property of such networks impact on the aggregate effects of economic shocks. Depending on the type of shock applied to the system, the simulations performed on the aggregate or firm- level data can diverge significantly. With our methodology we show that it is possible to predict the aggregate response of the system without knowledge of the network itself. We find that the simulations performed on the samples drawn from our ensemble give a probability distribution of aggregate impact that is consistent with what we obtain performing the simulations on a Dutch dataset of firm-to-firm payments. This motivates the use of this methodology as a more reliable alternative to simulating shocks on aggregate data.

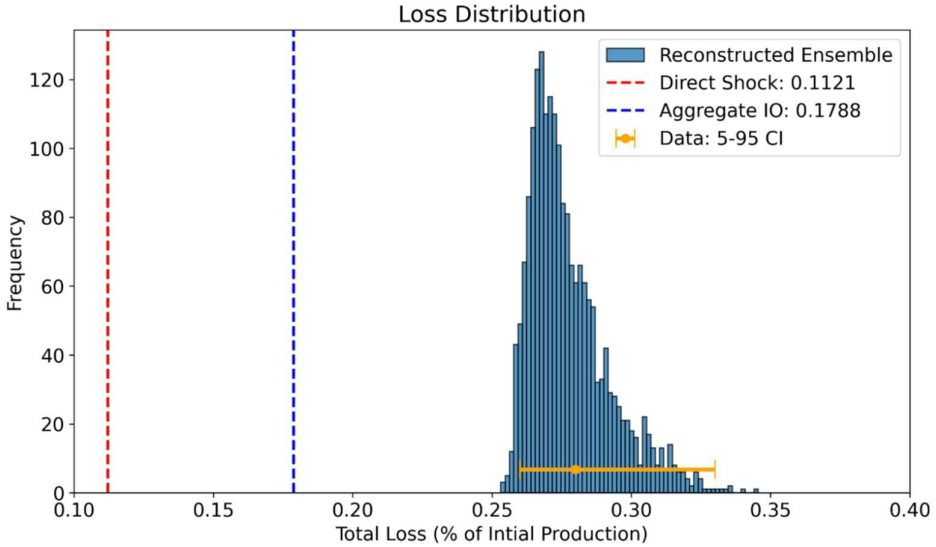


FIG. 1: Comparison between percentage decrease in total production for shock simulations performed on the aggregate IO network (red dotted line), on the empirical firm-to-firm network (yellow dot with confidence interval) and the reconstructed ensemble (histogram). We simulated 50 firm-level shocks compatible with a single aggregate one, both for the empirical and reconstructed networks.

1. Henriet et al. Journal of Economic Dynamics and Control, 36(1), (2012)
2. Diem et al. PNAS Nexus 3(3) (2024)
3. Garuccio et al. Physical Review Research 5(4) (2023)
4. Pichler et al. Journal of Economic Dynamics and Control, 144, (2022)

10:40 – 11:00 **Pavel Chvykov** (*Unaffiliated at the moment*)

Firms as inverse insurance companies

Why do firms self-organize in free markets? Why do laborers give up their freedoms for a smaller but stable income? Ergodicity Economics (EE) has pointed out that the

expected utility framework breaks down when wealth dynamics are non-ergodic (such as for exponential growth processes). This led to a new understanding of cooperation in general, and insurance in particular, as win-win scenarios, rather than as zero-sum games. In this work, we propose addressing the "firm puzzle" in economics in parallel to how EE addresses "the insurance puzzle." While we might expect independent freelance laborers to be better off in expected utility (short-term), we show that they can lose in time-average growth rate (long-term) compared to stably employed workers. In this sense, we conceptualize the firm as an "inverse insurance company" - where clients pay a "negative premium" (receive a salary), and offload "positive risk" (the fluctuating gains from their labor) to the company. This simple framework allows us to formalize some questions about the firm, and provide a new modeling and optimization paradigm for firms - which turns out to closely parallel portfolio construction in finance.

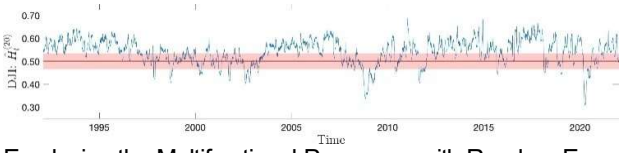
11:00 – 11:30 **Coffee break**

11:30 – 13:30 **Session 2: Kimmo Kaski** *(Aalto University School of Science, Finland)*

11:30 – 11:50 **Sergio Bianchi** *(Sapienza' University of Rome, Italy; Department of Finance and Risk Engineering, Tandon School of Engineering, New York University, USA)*

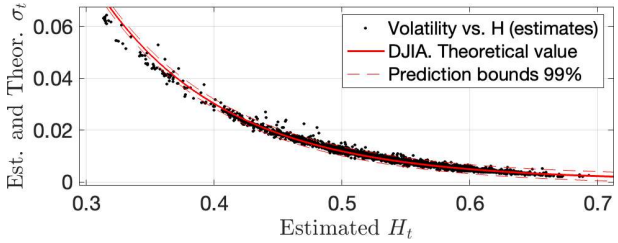
Financial risk, fair volatility and Hurst-Holder regularity

Motivated by the need of more effective and realistic models of financial markets dynamics, we disentangle financial risk from volatility, acknowledging that the latter does not comprehensively capture the risk inherent in financial markets—a perspective widely recognized among practitioners. The general consensus of the volatility as a pivotal indicator of risk is in fact implicitly based on the assumption that the Efficient Market Hypothesis (EMH) is true; only in this case the discounted prices are a (Brownian) martingale, and their quadratic variation $[X]t$ is linked to the realized volatility by $\sigma = \frac{[X]t}{}$. Nonetheless, this relationship may not adequately represent the complexity and dynamics of market behavior once the assumption of efficiency is relaxed. In this case, an alternative approach based on the Hurst-Hölder pointwise regularity can overcome the overreliance on volatility and explain more accurately the multifaceted nature of financial risk. In fact, unlike volatility —which solely measures data dispersion— the Hurst-Hölder exponent encapsulates memory and autocorrelation within the data, thus offering deeper insights into the departures from market efficiency and the potential predictability of returns.



Employing the Multifractional Processes with Random Exponent (MPRE), we discuss how the Hurst-Hölder pointwise regularity can be more informative than volatility, because it captures the evolving complexity and dependency structures observed in market data. As an example, the Figure shows the dynamics of the Hurst- Hölder exponent for the Dow Jones Industrial Average. The market is efficient (i.e. acts like a martingale) when the exponent lies within the colored area, and inefficient when it is out- side. Of course, volatility in itself does not provide this information nor does it reveal the magnitude of the pull exerted by efficiency on the dynamics. Since, under mild conditions, at each point *t* along its trajectory the MPRE can be viewed as tangent to a fractional Brownian motion (fBm) with parameter equal to the random exponent at *t* of the MPRE, we exploit this Local Asymptotical Self-Similarity (LASS) property to deduce a log-linear inverse relationship with volatility. An example is given in the Figure below, which plots the estimated volatility versus the estimated Hurst-Holder exponent along with the theoretical value provided by the MPRE (the example refers to the DJIA). Furthermore,

when the volatility process follows a fractional Ornstein-Uhlenbeck (fOU) process, the relationship indicates that the random exponent of the MPRE follows a fOU process as well, with properly scaled parameters.
An extensive empirical analysis is then performed on nine stock indexes worldwide (DJI, SPX and IXIC, from USA; SX5E, from Europe; UKX, from UK; HSI, from Hong Kong; SHCOMP, from China; STI, from Singapore; NI225, from Japan). Our findings confirm that:



- the Hurst-Holder regularity effectively captures significant deviations from market efficiency, indicating that markets undergo alternating phases of efficiency and inefficiency (both positive and negative);
- there exists a pervasive relationship between Hurst-Hölder regularity and realized volatility, consistent with theoretical expectations.

11:50 – 12:10 **Johannes Stangl** (*Complexity Science Hub, Vienna, Austria)*

Firm-level supply chains to minimize unemployment and economic losses in rapid decarbonization scenarios

Urgently needed carbon emissions reductions might lead to strict command-and-control decarbonization strategies with potentially negative economic consequences. Analysing the entire firm-level production network of a European economy, we have explored how the worst outcomes of such approaches can be avoided. We compared the systemic relevance of every firm in Hungary with its annual CO2 emissions to identify optimal emission-reducing strategies with a minimum of additional unemployment and economic losses. Setting specific reduction targets, we studied various decarbonization scenarios and quantified their economic consequences. We determined that for an emissions reduction of 20%, the most effective strategy leads to losses of about 2% of jobs and 2% of economic output. In contrast, a naive scenario targeting the largest emitters first results in 28% job losses and 33% output reduction for the same target. This demonstrates that it is possible to use firm-level production networks to design highly effective decarbonization strategies that practically preserve employment and economic output.

12:10 – 12:30 **Deb Panja** *(Computer Science and Vice-Director Centre for Complex Systems Studies (CCSS), Utrecht University, Netherlands)*

Timeliness criticality: A(n endogenous) mechanism for the "excess volatility puzzle"?

A large economy should not fluctuate too much: large economies contain a large number of firms, so the shocks they endure should average out for macroeconomic variables – at least that is how the standard argument goes. However, the United States, for example, has had an average real GDP growth of 3.1% per year since 1947, but with a standard deviation of the same order of magnitude: 2.1%. This illustrates the "excess volatility puzzle": it is a major fundamental unsolved problem in economics. Another facet of the excess volatility puzzle is that relatively small events cause unexpectedly large bubbles and crashes in prices and supply. A classic example of such an event is the singular tanker incident in the Suez Canal in March 2021, whose macroeconomic effects have been felt for over two years.
A significant amount of effort in economics and in the physics of complex systems has been devoted to understanding the excess volatility puzzle, leading to the creation of agent-based models where firms – constituting the backbone of economic production

– buy inputs from each other to create their own output, which they then sell to other firms that need them in turn. However, the timeliness aspect of production systems remains underappreciated, even though timeliness has been ubiquitously and integrally adopted as a quality standard for production systems. The basic idea is simple: for a firm to be able to produce something, it not only needs to be able to order inputs from its supplier, but it also needs to receive these inputs in time. Large delays in supply chains can cause a loss of coordination in production systems that provide the backbone for economic performance, exemplified by the ongoing backlog in the automotive supply chain as a consequence of the collapse of the Baltimore Key Bridge in March 2024.
We have recently developed a novel stylised model that has brought timeliness and operational delays in the provision of goods and services into the picture. Reinforced by competitive pressures, operators often myopically optimize for cost- and time-efficiencies, running the risk of inadvertently pushing production systems towards the proverbial “edge of a cliff” in the sense of timeliness. We have shown that this cliff edge is a true critical point – identified as timeliness criticality – implying that system efficiency and robustness to perturbation are in tension with each other [1]. Specifically for production systems, we suggest that the proximity to timeliness criticality is a measure for their fragility, resulting in large swings in being available on time can indeed cascade through supply chain, and is therefore a possible mechanistic route for unravelling the excess volatility puzzle [2].

[1] J. Moran, M. Romeijnders, P. Le Doussal, F. P. Pijpers, U. Weitzel, D. Panja, J.-P. Bouchaud. https://arxiv.org/abs/2309.15070 (to appear in Nature Physics).
[2] J. Moran, F. P. Pijpers, U. Weitzel, J.-P. Bouchaud, D. Panja. https://arxiv.org/abs/2307.03546.

12:30 – 12:50 **Henrik Olsson** (*Complexity Science Hub, Vienna, Austria)*

Networks of Beliefs: An Integrative Theory of Individual- and Social-Level Belief Dynamics

We introduce the Networks of Belief theory, which explains the interaction between personal beliefs and social beliefs. This theory improves upon previous theories in three key ways: it links individual belief networks with belief dynamics on social networks, recognizes the imperfect alignment between perceived and actual beliefs, and explains diverse belief phenomena through differences in attention to dissonances. Our computational model, based on a statistical physics framework, is supported by findings from two surveys (N1=973, N2=669). We offer insights into belief dynamics, including group consensus, polarization, radicalization, minority influence, and observed belief distributions, and discuss directions for future research.

12:50 – 13:10 **Pasquale Casaburi** *(Department of Mathematics, King’s College London)*

Source of multiscaling in rough volatility models

The multiscaling behavior of financial time-series has become one of the acknowledged stylized facts in the literature [1]. The origin of multiscaling and the best way to quantify it have been widely discussed in the literature [2,3]. In this talk, I will show results of the application of Generalized Hurst Exponent method to different financial time series and I will discuss the interplay between multiscaling and volatility roughness, defined as the (low) Hurst exponent of the volatility process [4]. Our study delves into the statistical roots of multiscaling through a robust analytical approach by also utilizing surrogate time series. By replicating distinct elements of the original data, these surrogate time series facilitate the disentanglement of various components within the multiscaling spectrum and help advancing the knowledge on the real source of multiscaling in real financial time series. Unraveling the sources of multiscaling in time series is essential to develop accurate models that can describe financial time series behaviours and our results contribute towards a better understanding of these models.

*

[1] T. Di Matteo, Quantitative Finance 7(1) (2007) 21.
[2] Jozef Barunik, Tomaso Aste, T. Di Matteo, Ruipeng Liu, Physica A 391 (2012) 4234–4251.
[3] Brandi G., Di Matteo T. (2022). On the statistics of scaling exponents and the multiscaling value at risk. The European Journal of Finance, 28(13–15), 1361–1382. https://doi.org/10.1080/1351847X.2021.1908391
[4] Giuseppe Brandi, T. Di Matteo, Multiscaling and rough volatility: an empirical investigation, International Review of Financial Analysis 84 (2022) 102324.

13:10 – 13:30 **Tiziana Di Matteo** (*Department of Mathematics, Kings College London*)
Stefan Thurner (*Complexity Science Hub, Medical University of Vienna*)

Closing

13:30 – 14:30 **Lunch**

*

Poster Presentations

Atushi Ishikawa *(Kanazawa Gakuin University, Japan)*

Correlation between the area enclosed by the movement trajectory and its perimeter, individual attributes, and external factors

Using data on the movement trajectories of approximately 15,000 people per day passing through Urayasu City, Chiba Prefecture, Japan, in August 2022, we confirmed that there is a strong correlation between the area and perimeter of the polygons they depict. The correlation parameters showed that the movement trajectories of people traveling long distances of 5 (km) or more were linear, and that this tendency was stronger on weekends and holidays, while those traveling short distances of less than 5 (km) were circular or square, enclosing a large area, and the tendency did not change significantly from weekdays to weekends or holidays.
We further analyzed the individual attributes of gender and age, which are added to the movement trajectories, and confirmed that the linear trend for long-distance travel is stronger for women than for men, and for age groups, it is stronger for those under 30 years of age. These results indicate that long-distance travelers have a specific destination and travel to and from that destination in a straight line, and that this tendency is stronger among women under 30 years of age.

To confirm whether the above characteristics are affected by external factors, similar analyses were conducted for Shimogyo-ku, Kyoto City in November 2021 and January 2022 (movement trajectory data for approximately 2,500 persons per day), and for Ishikawa Prefecture in August and September 2023 (approximately 5,000 persons per day). As a result, we confirmed that there is still a strong correlation between the area and perimeter of polygons drawn by movement trajectories, and that the movement trajectories of people traveling long distances of 5 (km) or more are linear, while those traveling short distances of less than 5 (km) are like circles or squares that enclose a large area. However, unlike in Urayasu City, Chiba Prefecture, the linear trend was observed to weaken on weekends and holidays. This may be related to the presence of Disneyland, Japan's largest theme park, in Urayasu City, Chiba Prefecture, whereas Kyoto and Ishikawa Prefecture have several historical tourist attractions.

Rintaro Karashima *(University of Hyogo, Graduate School of Information Science, Kobe, Japan)*

Complexity of Technologies in Japanese Corporations

In the rapidly evolving landscape of global technology, identifying the crucial technologies by their quality is essential for understanding competitiveness of corporations. Here, we apply the Technological Complexity Index (TCI), a metric to qualify the technologies, and struggle to reveal the core technologies in Japanese corporations. The TCI, which was proposed by Balland et al.[1] following the foundational methodology of Hidalgo and Hausmann[2], assigns higher values to technologies that are ”standard” as they are specialized in by many corporations which typically also specialize in other ”standard” technologies. The data for this study is sourced from the Japan Patent Office (JPO), utilizing 2,875,696 registered patents filed by 1929 corporations from fiscal year 1981 to 2010. We aggregate the data by both corporate patent holders and patent classifications proposed by Schmoch[3], which maps the International Patent Classification (IPC) codes onto 35 technological fields. Using the aggregated data, we construct a bipartite graph that signifies corporations’ specialization in certain technological fields, and determine the TCI as the eigenvector of the technological field nodes. Through a 30-year longitudinal analysis and decadespecific reviews, we find that the consistently high TCI rankings of Pharmaceuticals and Biotechnology, and the progressively increasing TCI ranking of Surface Technology and Coating technologies. These findings could represent the historical context such as the nascent stages of growth in the pharmaceuticals sectors, and the crucial development of the surface technology in response to rapid industrialization in the 1980s and mass production of electronic devices thereafter. Also, as evidenced by the fact that neither the number of patents nor the patent-holding corporations significantly

correlates with the TCI, it can be seen that technologies with high TCI rankings do not necessarily possess a vast number of patents or corporate patent holders. However, the verification of the relationship between technological quantity and the TCI remains insufficient, and thus the future study should further explore this area's utility and impact. Our methodology contributes to revealing technology trends aligned with the social context, providing scholars and policymakers with insights into the dynamics of technological advancement that cannot be captured by quantitative metrics alone.

[1] Balland, P. A., and Rigby, D. (2017). The geography of complex knowledge. Economic geography, 93(1), 1-23.
[2] Hidalgo, C. A., and Hausmann, R. (2009). The building blocks of economic complexity. Proceedings of the national academy of sciences, 106(26), 10570-10575.
[3] Schmoch, U. (2008). Concept of a technology classification for country comparisons. Final report to the world intellectual property organisation (wipo), WIPO.

Andrea Monaco *(University College Dublin, School of Mathematics and Statistics, Ireland)*

Wealth dynamics in a multi-aggregate closed monetary system

In this work we examine the statistical properties of a closed monetary economy with multi-aggregates interactions. Building upon Yakovenko's single agent monetary model (Dragulescu and Yakovenko, 2000), we investigate the joint equilibrium distribution of aggregate size and wealth. By comparing theoretical and simulated data, we validate our findings and investigate the influence of both micro dynamics and macro characteristics of the system on the distribution. Additionally, we analyse the system's convergence towards equilibrium under various conditions. Our laboratory model may offer valuable insights into macroeconomic phenomena. Despite relying on simplified assumptions regarding wealth dynamics, it enables to replicate typical wealth distribution patterns observed in the real economy.

Irena Vodenska *(Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Republic of North Macedonia; Administrative Sciences Department, Financial Management, Metropolitan College, Boston University, USA)*

ESG Factor Effects on Corporate Risk and Performance: Governance

This study investigates the relationship between corporate governance factors and firm performance and risk, focusing on financial considerations, board characteristics, and adherence to governance principles. It aims to explore the impact of governance practices on business resilience and financial performance. The investigation seeks to understand whether effective corporate governance practices lead to better financial performance and lower risk, enhancing firm's value.
The paper applies panel regression analysis on 601 public companies from the US (437) and the EU (164) between October 2019 and February 2023. The companies belong to various sectors, including Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Healthcare, Industrials, Information Technology, Materials, Real Estate, and Utilities. The EU sample covers Developed European countries, including Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.
Preliminary results reveal significant relationships between governance factors and firm performance and risk. Higher accruals (a typical measure of low earnings quality) are associated with lower net income. Independent audit committees may negatively impact financial performance but are not significantly related to corporate risk. Having a risk management expert on the Board is associated with higher volatility and lower performance while having a financial expert is positively related to net income. The involvement of individuals or entities closely connected to the companies in various transactions, including asset transfers, loans, and loan guarantees, is negatively related to financial performance measured by gross margin.
The study highlights the different perspectives on governance practices between US and EU companies. Higher accruals are negatively related to financial performance in the US, while they are positively related to performance in the EU. Also, while having a financial expert on the Board is

positively associated with performance in the US, it is negatively related to gross margin in the EU. The transactions by related individuals and entities and audit committee independence significantly negatively affect firm performance only in the US. On the other hand, the negative relationship between having a risk management expert on the Board and financial performance is only significant in the EU and not significant in the US.